

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



Presentado por

JAIME NAVARRO CRUZ

DISEÑO DE UN SISTEMA AUTOMÁTICO DE RECONOCIMIENTO
DE PATRONES PARA LA CLASIFICACIÓN DE PALTA EN LA
COMUNIDAD DE MOLLEBAMBA- CHINCHEROS

Asesor

ING. ROBERTO QUISPE QUISPE

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE

INGENIERO DE SISTEMAS

ANDAHUAYLAS – APURÍMAC – PERÚ

2023

Aprobación del Asesor

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS RESOLUCIÓN N°360-2021-CFI-UNAJMA RESOLUCIÓN DE COORDINACIÓN DE FACULTAD DE INGENIERÍA

Andahuaylas, 25 de noviembre de 2021

VISTO: La Carta N°014-2021-DUI-JJOC-FI-UNAJMA de fecha 25 de noviembre de 2021, el Ing. Juan José Oré Cerrón director de la Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional José María Arguedas, solicita la aprobación de designación de Asesor del Proyecto e Informe Final de Tesis de **JAIME NAVARRO CRUZ** de la escuela profesional de Ingeniería de Sistemas y;

CONSIDERANDO:

Que, por Ley N° 28372 del 29 de octubre del 2004, se crea la Universidad Nacional José María Arguedas, con sede en la provincia de Andahuaylas, Región Apurímac; y que por Resolución N° 035-2017-SUNEDU/CD de 02 de octubre del 2017, el Consejo Directivo de la Superintendencia Nacional de Educación Superior Universitaria, otorga la Licencia Institucional a la Universidad Nacional José María Arguedas para ofrecer el Servicio Educativo Superior Universitario;

Que, la Ley Universitaria 30220 en su Artículo Octavo respecto a la autonomía universitaria, establece que: "El estado reconoce la autonomía universitaria". La autonomía inherente a las universidades se ejerce de conformidad a la Constitución, las leyes y demás normativa aplicable, esta Normativa se manifiesta en los siguientes regímenes: Normativo, De gobierno, Académico, Administrativo y Económico;

Que, mediante Carta Múltiple N° 020-2014-SG-UNAJMA, de fecha 30 de julio del 2014; la Secretaría General de la UNAJMA comunica que mediante Acuerdo N° 03 de Sesión Ordinaria de la Comisión de Gobierno se **AUTORIZA** la emisión de **RESOLUCIONES DE COORDINACIÓN DE LA FACULTAD** estrictamente para asuntos académicos y deberán remitirse un original a la Secretaría General;

Que, mediante carta N° 236-2016-SG-UNAJMA de fecha 05 de agosto de 2016 el Ing. Enrique Edgardo Córdor Tinoco, Secretario General de la UNAJMA, comunica que el Dr. Oswaldo Luizar Obregón, Presidente de la Comisión Organizadora de la UNAJMA ha dispuesto que las resoluciones emitidas por la Facultad se deriven a la Vicepresidencia Académica;

Que, el **art. 39 incisos a y d del TITULO II, CAPITULO II del Reglamento General de la UNAJMA**, aprobado mediante Resolución N° 0130-2016-CO-UNAJMA, establece que "Son funciones de las Facultades: a) dirigir el desarrollo académico y administrativo de las Escuelas Profesionales y Departamentos Académicos adscritos a esta, dentro de la normatividad legal, d) administrar el sistema de matrícula en coordinación y apoyo con la oficina respectiva";

Que, el **art. 59 del capítulo III "DEL ASESOR DEL PROYECTO DE TESIS Y DEL INFORME FINAL" del Reglamento General de Grados y Títulos de la UNAJMA**, aprobado con Resolución N°255-2021-CO-UNAJMA, de fecha 10 de setiembre de 2021, establece "la tesis requiere de un Asesor y además un Co-asesor, si así lo amerita el caso, a propuesta del tesista, [...]";

Que, mediante **Resolución N°121-2021-CFI-UNAJMA** de fecha 20 de mayo de 2021 se designa al director de la Unidad de Investigación de la Facultad de Ingeniería;

Que, con Acta N°014-2021-DUI-JJOC-FI-UNAJMA, de Designación de asesor, de fecha 25 de noviembre del 2021, la dirección de la Unidad de Investigación de la Facultad de Ingeniería dirigida por el MSc. Juan José Oré Cerrón aprueba la designación del Ing. Roberto Quispe Quispe como asesor del proyecto e Informe Final de Tesis del bachiller de la Escuela Profesional de Ingeniería de Sistemas **JAIME NAVARRO CRUZ**;

Que, con Carta N°014-2021-DUI-JJOC-FI-UNAJMA, de fecha 25 de noviembre del 2021, el Msc. Juan José Oré Cerrón en su condición de director de la Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional José María Arguedas, solicita la aprobación mediante acto resolutivo la designación del Ing. Roberto Quispe Quispe del Proyecto e Informe Final de Tesis con fines de titulación de **JAIME NAVARRO CRUZ** de la Escuela Profesional de Ingeniería de Sistemas;

Que, en atención a la Carta N°014-2021-DUI-FI-UNAJMA, el Dr. Yalmar Temístocles Ponce Atencio, Coordinador de la Facultad de Ingeniería de la Universidad Nacional José María Arguedas, dispone a la Secretaría

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
RESOLUCIÓN N°360-2021-CFI-UNAJMA
RESOLUCIÓN DE COORDINACIÓN DE FACULTAD DE INGENIERÍA

Académica de la Facultad de Ingeniería proyectar la Resolución correspondiente, la que se aprueba con cargo a dar cuenta a la Vicepresidencia Académica;

Por estos considerandos y en uso de las atribuciones conferidas como Coordinador de la Facultad de Ingeniería, designado mediante Resolución N° 0298-2019-CO-UNAJMA, de fecha 15 de octubre de 2019;

SE RESUELVE:

ARTÍCULO PRIMERO: APROBAR la designación del Ing. Roberto Quispe Quispe como Asesor del Proyecto e Informe Final de Tesis con fines de titulación de **JAIME NAVARRO CRUZ** de la Escuela Profesional de Ingeniería de Sistemas.

ARTÍCULO SEGUNDO: ENCARGAR a la Escuela Profesional de Ingeniería de Sistemas, ejecute y adopte las acciones académicas y administrativas que correspondan, para el cabal cumplimiento de la presente Resolución.

ARTÍCULO TERCERO: REMITIR la presente Resolución a la Vicepresidencia Académica, Escuela Profesional de Ingeniería de Sistemas, Docente Asesor, y al interesado para su conocimiento y fines pertinentes.

REGÍSTRESE, COMUNÍQUESE Y ARCHÍVESE.


UNIVERSIDAD NACIONAL
JOSÉ MARÍA ARGUEDAS
Dr. Yaimar Ponce Alencastro
COORDINADOR DE LA FACULTAD DE INGENIERÍA


UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
Ing. Richard A. Flores Condori
SECRETARIO ACADÉMICO

Copia del acta de sustentación



Universidad Nacional José María Arguedas

Identidad y Excelencia para el Trabajo Productivo y el Desarrollo



FACULTAD DE INGENIERIA

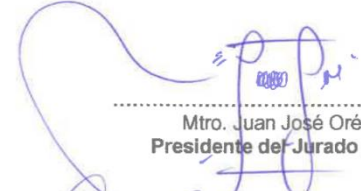
ACTA DE SUSTENTACION DE TESIS


En la Av. José María Arguedas del Local Académico SL01 (Ccoyahuacho) en el auditorio de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional José María Arguedas ubicado en el distrito de San Jerónimo de la Provincia de Andahuaylas, siendo las 09:00 horas del día 01 de marzo del año 2023, se reunieron los docentes: Mtro. Juan José Oré Cerrón, Mag. Rubén Apaza Apaza, Mag. Humberto Silvera Reynaga, en condición de integrantes del Jurado Evaluador del Informe Final de Tesis intitulado: "DISEÑO DE UN SISTEMA AUTOMÁTICO DE RECONOCIMIENTO DE PATRONES PARA LA CLASIFICACIÓN DE PALTA EN LA COMUNIDAD DE MOLLEBAMBA – CHINCHEROS", cuyo autor es el Bachiller en Ingeniería de Sistemas **JAIME NAVARRO CRUZ** y el asesor Ing. Roberto Quispe Quispe y, con el propósito de proceder a la sustentación y defensa de dicha tesis.


Luego de la sustentación y defensa de la tesis, el Jurado Evaluador **ACORDÓ: Aprobar** por **Unanimitad** al Bachiller en Ingeniería de Sistemas **JAIME NAVARRO CRUZ**, obteniendo la siguiente calificación y mención:

Nota escala vigesimal		Mención
Números	Letras	
15	Quince	Bueno

En señal de conformidad, se procedió a la firma de la presente acta en 03 ejemplares.


.....
Mtro. Juan José Oré Cerrón
Presidente del Jurado Evaluador


.....
Mag. Rubén Apaza Apaza
Primer Miembro del Jurado Evaluador


.....
Mag. Humberto Silvera Reynaga
Segundo Miembro del Jurado Evaluador

Aprobación del Jurado Evaluador



UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS RESOLUCIÓN N°089-2023-CFI-UNAJMA RESOLUCIÓN DE COORDINACIÓN DE FACULTAD DE INGENIERÍA

Andahuaylas, 30 de enero de 2023

VISTO: La Carta N°067-2023-DUI-JJOC-FI-UNAJMA de fecha 25 de enero de 2023, el Mtro. Juan José Oré Cerrón presidente de la Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional José María Arguedas, solicita la aprobación de la reconfiguración del jurado evaluador del bachiller de Ingeniería de Sistemas **JAIME NAVARRO CRUZ** y;

CONSIDERANDO:

Que, por Ley N° 28372 del 29 de octubre del 2004, se crea la Universidad Nacional José María Arguedas, con sede en la provincia de Andahuaylas, Región Apurímac; y que por Resolución N° 035-2017-SUNEDU/CD de 02 de octubre del 2017, el Consejo Directivo de la Superintendencia Nacional de Educación Superior Universitaria, otorga la Licencia Institucional a la Universidad Nacional José María Arguedas para ofrecer el Servicio Educativo Superior Universitario;

Que, la Ley Universitaria 30220 en su Artículo Octavo respecto a la autonomía universitaria, establece que: "El estado reconoce la autonomía universitaria". La autonomía inherente a las universidades se ejerce de conformidad a la Constitución, las leyes y demás normativa aplicable, esta Normativa se manifiesta en los siguientes regímenes: Normativo, De gobierno, Académico, Administrativo y Económico;

Que, mediante Carta Múltiple N° 020-2014-SG-UNAJMA, de fecha 30 de julio del 2014; la Secretaría General de la UNAJMA comunica que mediante Acuerdo N° 03 de Sesión Ordinaria de la Comisión de Gobierno se **AUTORIZA** la emisión de **RESOLUCIONES DE COORDINACIÓN DE LA FACULTAD** estrictamente para asuntos académicos y deberán remitirse un original a la Secretaría General;

Que, mediante carta N° 236-2016-SG-UNAJMA de fecha 05 de agosto de 2016 el Secretario General de la UNAJMA, comunica que el Dr. Oswaldo Luizar Obregón, Presidente de la Comisión Organizadora de la UNAJMA ha dispuesto que las resoluciones emitidas por la Facultad se deriven a la Vicepresidencia Académica;

Que, el **art. 39 incisos a y d del TÍTULO II, CAPÍTULO II del Reglamento General de la UNAJMA**, aprobado mediante Resolución N° 0130-2016-CO-UNAJMA, establece que "Son funciones de las Facultades: a) *dirigir el desarrollo académico y administrativo de las Escuelas Profesionales y Departamentos Académicos adscritos a esta, dentro de la normatividad legal, d) administrar el sistema de matrícula en coordinación y apoyo con la oficina respectiva*";

Que, en el **Art. 69° del CAPÍTULO V (DE LA TESIS PARA LA OBTENCIÓN DEL TÍTULO PROFESIONAL) del Reglamento General de Grados y Títulos en la UNAJMA**, aprobado con Resolución N° 012-2023-CO-UNAJMA, de fecha 16 de enero de 2023 establece que: "La aprobación del proyecto de tesis se inicia con la presentación de dos (02 juegos) impresos firmados por el tesista o tesistas y el asesor y/o coasesor y la Declaración Jurada de Autenticidad de la Tesis (de acuerdo al formato del anexo N°32: Declaración Jurada de Autenticidad) ante la Primera Autoridad de la Facultad, la misma que corre traslado a la Unidad de Investigación de la Facultad, para que en un plazo no mayor a (05) días hábiles, designe a dos docentes revisores. Los docentes revisores pueden ser ordinarios o contratados y deben ser de la especialidad o afin al tema de la tesis.

Que, con Resolución N° 360-2021-CFI-UNAJMA de fecha 25 de noviembre del 2021, se aprueba la designación del Ing. Roberto Quispe Quispe como Asesor del Proyecto e Informe Final de Tesis con fines de titulación del Bachiller en Ingeniería de Sistemas **JAIME NAVARRO CRUZ**;

Que, con Resolución N° 001-2022-CFI-UNAJMA, de fecha 05 de enero de 2022, se aprueba la designación del Jurado Evaluador del Proyecto e Informe Final de Tesis del Bachiller en Ingeniería de Sistemas **JAIME NAVARRO CRUZ**;

Que, con Resolución N° 338-2022-CFI-UNAJMA, de fecha 09 de noviembre de 2022, se aprueba la reconfiguración del Jurado Evaluador del Proyecto e Informe Final de Tesis del Bachiller en Ingeniería de Sistemas **JAIME NAVARRO CRUZ**;

Que, con Resolución N° 037-2022-CFI-UNAJMA, de fecha 22 de febrero de aprueba el Proyecto de Tesis del Bachiller en Ingeniería de Sistemas **JAIME NAVARRO CRUZ** para optar el Título Profesional de Ingeniero de Sistemas por la Modalidad de Sustentación de Tesis, el mismo que ha sido **APROBADO SIN OBSERVACIONES**;





UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
RESOLUCIÓN N°089-2023-CFI-UNAJMA
RESOLUCIÓN DE COORDINACIÓN DE FACULTAD DE INGENIERÍA

Que, con solicitud S/n de fecha 24 de enero del 2023, el tesista JAIME NAVARRO CRUZ; solicita la re conformación del primer miembro del jurado evaluador para su proyecto de tesis;

Que, con Carta N°067-2023-DUI-JJOC-FI-UNAJMA de fecha 25 de enero de 2023, el Mtro. Juan José Oré Cerrón presidente de la Unidad de Investigación de la Facultad de Ingeniería de la Universidad Nacional José María Arguedas, solicita la aprobación de la re conformación de los miembros del jurado evaluador del Proyecto e Informe Final de Tesis de **JAIME NAVARRO CRUZ**;

Que, con Acta N° 067-2023-DUI-JJOC-FI-UNAJMA, de Designación de jurado evaluador de fecha 25 de enero del 2023, la Comisión de la Unidad de Investigación de la Facultad de Ingeniería presidida por el Mtro. Juan José Oré Cerrón aprueba la re conformación de los miembros del jurado evaluador del proyecto e Informe Final de Tesis del bachiller de Ingeniería de Sistemas **JAIME NAVARRO CRUZ**;

Que, en atención a la Carta N°067-2023-DUI-JJOC-FI-UNAJMA, el Dr. David Juan Ramos Huallpartupa, Coordinador de la Facultad de Ingeniería de la Universidad Nacional José María Arguedas, dispone al Secretario Académico de la Facultad de Ingeniería proyectar la Resolución correspondiente, la que se aprueba con cargo a dar cuenta a la Vicepresidencia Académica;

Por estos considerandos y en uso de las atribuciones conferidas como Coordinador de la Facultad de Ingeniería, designado mediante Resolución N° 275-2022-CO-UNAJMA de fecha 08 de agosto del 2022

SE RESUELVE:

ARTÍCULO PRIMERO: APROBAR la re conformación de los miembros del jurado evaluador del Informe Final de Tesis, del bachiller de Ingeniería de Sistemas **JAIME NAVARRO CRUZ**, de acuerdo al siguiente detalle:

Proyecto de Tesis titulado	"DISEÑO DE UN SISTEMA AUTOMÁTICO DE RECONOCIMIENTO DE PATRONES PARA LA CLASIFICACIÓN DE PALTA EN LA COMUNIDAD DE MOLLEBAMBA – CHINCHEROS"	
Tesista	Bachiller en Ingeniería de Sistemas JAIME NAVARRO CRUZ	
Asesor	Ing. Roberto Quispe Quispe	
Jurado Evaluador	Presidente:	Mtro. Juan José Oré Cerrón
	Primer Miembro:	Mag. Rubén Apaza Apaza
	Segundo Miembro:	Mag. Humberto Silvera Reynaga

ARTÍCULO SEGUNDO: ENCARGAR a la Escuela profesional de Ingeniería de Sistemas de la Universidad Nacional José María Arguedas, adopte las acciones correspondientes para el cabal cumplimiento de la presente resolución.

ARTÍCULO TERCERO: REMITIR la presente Resolución a la Vicepresidencia Académica, Escuela Profesional de Ingeniería de Sistemas, miembros del Jurado Evaluador y al interesado para su conocimiento y fines pertinentes.

REGÍSTRESE, COMUNÍQUESE Y ARCHÍVESE.

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
Dr. David Juan Ramos Huallpartupa
COORDINADOR DE LA FACULTAD DE INGENIERÍA

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
FACULTAD DE INGENIERÍA
MSc. Jesús Farfán Inca Roca
SECRETARIO ACADÉMICO

Resulta del informe de similitud de Tesis

Reporte de similitud

NOMBRE DEL TRABAJO

Jaime Navarro - Diseño de un sistema automático de reconocimiento de patrones (1).docx

AUTOR

JAIME NAVARRO CRUZ

RECUENTO DE PALABRAS

12339 Words

RECUENTO DE CARACTERES

71246 Characters

RECUENTO DE PÁGINAS

99 Pages

TAMAÑO DEL ARCHIVO

6.0MB

FECHA DE ENTREGA

Jun 29, 2023 11:20 PM GMT-5


FECHA DEL INFORME


Jun 29, 2023 11:21 PM GMT-5

● 27% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos

- 27% Base de datos de Internet
- Base de datos de Crossref
- 5% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref


Ing. Roberto Cisneros Cisneros
ASESOR.


Bach. Jaime Navarro Cruz
TESISTA

Resumen

DEDICATORIA

Dedico el trabajo de investigación a mi familia. Principalmente a mis padres que son motor y motivo de mi día a día. Gracias por enseñarme de cómo enfrentar la vida sin perder nunca la cabeza en el intento.

Me han guiado en mi proceso de formación profesional y personal para llegar a ser la persona que soy hoy. Con enorme dedicación y amor y sin pedir nada a cambio.

También dedico este trabajo a mis hermanos y amigos. Por su apoyo constante y comprensión, que tal vez sufrieron impacto directo de consecuencias de este trabajo. Realmente ellos me ayudaron a alcanzar equilibrio que permite dar mi potencial.

AGRADECIMIENTO

A mis padres por brindarme el soporte que fue necesario para culminar la presente investigación, a mis hermanos por su infinita paciencia y apoyo constante en proceso de formación profesional.

Y finalmente a mi alma mater universidad Nacional José María Arguedas por haberme albergado y brindado los cimientos necesarios para encaminar esta travesía llamada "la formación de una persona del bien".

ÍNDICE

DEDICATORIA.....	VII
AGRADECIMIENTO.....	VIII
LISTA DE TABLAS.....	XI
LISTA DE FIGURAS	XII
ÍNDICE DE ECUACIONES.....	XIV
RESUMEN.....	XV
ABSTRACT	XVI
CHUMASQA	XVII
INTRODUCCIÓN	18
CAPITULO I.....	20
PROBLEMA DE INVESTIGACIÓN.....	20
1.1 Realidad problemática.....	20
1.2 Formulación de problema.....	21
1.3 Objetivos	22
1.4 Justificación.....	23
1.5 Viabilidad	23
1.6 Limitaciones	24
CAPITULO II	26
ANTECEDENTES.....	26
2.1 Antecedentes Locales	26
2.2 Antecedentes Nacionales.....	27
2.3 Antecedentes Internacionales.....	28
CAPITULO III	32
MARCO TEÓRICO	32
3.1 Lenguaje de programación Python.....	48
3.2 SQLite: La Base de Datos Embebida.....	51
CAPITULO IV	56
METODOLOGÍA DE LA INVESTIGACIÓN.....	56
4.1 Tipo de investigación	56
4.2 Enfoque de investigación	56

4.3	Diseño de investigación	56
4.4	Nivel de investigación.....	56
4.5	Población.....	56
4.6	Muestra.....	57
4.7	Técnicas y análisis de datos	57
4.8	Hipótesis de investigación.....	58
4.9	Nivel de significancia y nivel de confianza	58
4.10	Prueba de normalidad.....	59
4.11	Prueba estadística	59
CAPITULO V DESARROLLO DEL OBJETO		61
5.1	Diagramas de bloques de la propuesta de solución.....	61
5.2	Recopilación de imágenes.....	62
5.3	Prototipo para captura de imágenes	62
5.4	Iluminación:	64
CAPITULO VI RESULTADOS		81
6.1	Entrenamiento del modelo de regresión lineal para estimación de peso.....	81
6.2	Objetivos	87
CONCLUSIONES		89
RECOMENDACIONES		90
REFERENCIAS BIBLIOGRÁFICAS		91
ANEXO.....		95

LISTA DE TABLAS

Tabla 1	Valor Nutricional de la palta	32
Tabla 2	Clasificación según calibre.....	35
Tabla 3	Clasificación según calibre indiferente	36
Tabla 4	Operacionalización de variables.....	60
Tabla 5	Muestras	81
Tabla 6	Prueba de normalidad a los datos procesados	83

LISTA DE FIGURAS

Figura 1	cuadro de pixeles	37
Figura 2	pixel.....	38
Figura 3	Imagen blanco y negro	39
Figura 4	Tono de gris entre blanco y negro.....	39
Figura 5	Modo RGB.....	40
Figura 6	Modo HSV	41
Figura 7	Extraído de Google.....	42
Figura 8	Métodos y conceptos de la visión artificial.....	43
Figura 9	Métodos y conceptos de visión Artificial.....	43
Figura 10	Apertura.....	44
Figura 11	Extraída del libro métodos y conceptos de visión Artificial	45
Figura 12	Metodología de Desarrollo del Software.	46
Figura 13	Extraído de platzi	47
Figura 14	Diagrama de bloques extraccion de características y analisis de color.....	61
Figura 15	Diagrama de bloques de entrenamiento del modelo de regresión lineal.....	61
Figura 16	Cámara para capturar imágenes	62
Figura 17	Diseño del Prototipo.....	63
Figura 18	Cámara de captura.....	63
Figura 19	Iluminación led.....	64
Figura 20	Flujograma de la instalación del hardware de sistema.....	65
Figura 21	Adquisición de imágenes	65
Figura 22	Apartado de imágenes	66
Figura 23	Apartado	66
Figura 24	Lectura de fotogramas	67
Figura 25	Conversión de fotogramas.....	67
Figura 26	Interfaz de la captura de imagen	68
Figura 27	Imágenes de las paltas sanas	68
Figura 28	Imágenes de las paltas defectuosas	69
Figura 29	Mejoramiento del contraste de las imágenes.....	69
Figura 30	Resultado:.....	69

Figura 31	Imagen escala de grises	70
Figura 32	binarizar.....	70
Figura 33	Binarizado	70
Figura 34	Extracción de contornos	71
Figura 35	Conversión de las dimensiones a cm	72
Figura 36	Análisis de color grafica.....	¡Error! Marcador no definido.
Figura 37	Análisis de color código.....	73
Figura 38	Representación del área de la zona verde.....	74
Figura 39	menú principal.....	75
Figura 40	Configuración.....	76
Figura 41	Sub menú.....	76
Figura 42	Categorización.....	77
Figura 43	procesamiento.....	77
Figura 44	datos estadísticos	78
Figura 45	Nuevo usuario.	79
Figura 46	Procesamiento	80
Figura 47	Carga de datos (pesos de palta).....	82
Figura 48	Grafica de la distribución de datos y su densidad de probabilidad	83
Figura 49	prueba de D' de agostino.....	84
Figura 50	Prueba de correlación de	85
Figura 51	Realización del test.....	85
Figura 52	Entrenamiento del modelo	86
Figura 53	Capacidad predictiva	86
Figura 54	Flujo de procesamiento de palta.....	87
Figura 55	Diagrama caso de uso.....	88
Figura 56	Diagrama Entidad – Relación de la base de datos	88

ÍNDICE DE ECUACIONES

<i>Ecuación 1</i> <i>Equivalencia de Componente de Calor</i>	39
<i>Ecuación 2</i> <i>Cierre</i>	44
<i>Ecuación 3</i> <i>Umbralización</i>	44
<i>Ecuación 4</i> <i>Umbral de la Imagen</i>	45
<i>Ecuación 5</i> <i>Segmentación</i>	45
<i>Ecuación 6</i> <i>Regresión Lineal</i>	52
<i>Ecuación 7</i> <i>Regresión Lineal Poblacional</i>	53
<i>Ecuación 8</i> <i>Varianza de la Homocedasticidad</i>	55

RESUMEN

Este trabajo propone desarrollar e implementar sistema automático de selección de paltas Hass, según calidad (rozamiento, quemaduras, manchas negras y peso). Sistema integra hardware y software que mejora calidad de exportación y productividad en industria agrícola.

El trabajo propone solución al problema actual empresas medianas y pequeñas empresas que comercializan palta. La palta, aguacate o avocado, tipo Hass, es el más conocido, comercializado y consumido, siendo necesario llevar control óptimo de calidad, proceso que es ineficiente por ser artesanal.

Como mejora de proceso, propone automatización de diversas etapas mecánicas, creadas para mover paltas y algoritmo de proceso de imagen, para detectar desorden de paltas. Al final se ven pruebas de validación y resultados de análisis usados en evaluación de solución propuesta.

Palabras clave: palta, automatización, control de calidad, etapa de selección, procesamiento de imágenes.

ABSTRACT

This research project proposes the development and implementation of an automatic sorting system for Hass avocados, using their different quality indexes (rubbing, burns and black spots). The system integrates hardware and software to improve export quality and productivity in the agricultural industry.

This project is the proposed solution to the current problem of avocado exporting companies in the selection stage. Avocado, also known as avocado or avocado, in its Hass variety, is the best known and is grown exclusively for export, for this reason it is necessary an optimal quality control; however, currently this process is inefficient as it is handmade.

In order to improve this process, its automation is proposed, which consists of several mechanical stages, designed to perform the movement of the avocados and an image processing algorithm, in charge of detecting disorders in the fruits. Finally, the validation tests and the results of the analysis used in the evaluation of the proposed solution are shown.

Keywords: Avocado; Automation; Quality control; Sorting stage; Image processing.

CHUMASQA

Kay proyecto de investigación nisqa, huk sistema de selección automática nisqa ruwayta, ruwayta ima yuyaychakun, chay aguacates tipo Hass nisqapaq, imaymana índices de calidad nisqawan (fricción, quemaduras y manchas yanas). Chay sistemaqa hardware nisqawan software nisqawanmi hukllawakun, hawa llaqtakunaman apachiypa allin kayninta, ruruchiypas chakra llamkay industriapi allinta ruwananpaq.

Kay proyectoqa, kunan tiempoqa aguacate hawa llaqtakunaman apachiq empresakunaq sasachakuyninkupaq, akllay etapapi. Chay aguacate, aguacate utaq aguacate nisqapas, Hass nisqa variedadninpi, aswan riqsisqa, hawa llaqtakunaman apachinapaqlla wiñasqa, chayraykum control óptimo de calidad nisqa necesario; ichaqa kay ruwayqa kunan pachaqa mana allinchi makiwan ruwasqa kasqanrayku.

Kay ruwayta allinchanapaqmi, automatizacinninta yuyaychakun, chaymi imaymana etapas mecánicas nisqamanta ruwasqa, chaymi ruwasqa kachkan aguacatekunata kuyuchinapaq hinallataq algoritmo de procesamiento de imágenes nisqa, chaymi encargakun frutakunapi desórdenes nisqakunata riqsinanpaq. Tukuyninpiqa, pruebas de validación nisqa, chaymanta chay analisis nisqapa ruwakusqanmi qawarichikun chay solución propuesta nisqa chaninchaypi.

INTRODUCCIÓN

Palta, aguacate o avocado, fruto originario de Centroamérica, usado en medicina y alimentación, pudiendo ser de varios tipos. En Perú, el tipo Hass es cultivado para exportación y el tipo Fuerte, suple demanda interna del mercado (Armando, 2019)

Según Ministerio de Agricultura y Riego (MINAGRI), exportar palta Hass es un rubro relevante y con gran crecimiento económico en Perú, siendo segundo exportador más grande a nivel mundial. De toda producción de palta Hass en Perú, 95% se exporta y 5% para consumo doméstico, siendo consumo per cápita el menor en el país comparado con otros lugares como EE.UU, México o Chile, por ello la necesidad de realizar una buena selección del producto. (Agencia Peruana de Noticias, 2022)

En selección manual puede haber errores, por ser trabajo repetitivo, cansado para la vista y manual. Para dar solución esta problemática, este estudio creo e implementó software automático que reconoce patrones basado en procesamiento digital de imagen. Implementar software determina cantidad total de Paltas y los discrimina según tipo y tamaño automáticamente. Así es posible prescindir de operario calificado y subjetividades. Estudio se organizó en 7 capítulos:

Capítulo I Problema de investigación, describe y fundamenta realidad problemática de, identifica objetivos, justificación, viabilidad y limitaciones.

Capítulo II Antecedentes, trabajos previos del tema, siendo locales, nacionales e internacionales.

Capítulo III Marco teórico, se ven teorías y términos empleados.

Capítulo IV Metodología de investigación, se ve tipo de estudio, enfoque, diseño, nivel, población, muestra, técnicas y análisis de datos, hipótesis de investigación y pruebas estadísticas.

Capítulo V Desarrollo de objeto IoT, señala etapas para desarrollo de prototipo de

sistema ACUATEC.

Capítulo VI resultados producto de adiestramiento de sistema ACUATEC.

Capítulo VI resultados del trabajo se relacionan con otros estudios. Continuando con conclusiones, recomendaciones y referencias bibliográficas.

CAPITULO I

PROBLEMA DE INVESTIGACIÓN

1.1 Realidad problemática

En el mundo, banana, piña, palta y mango son frutas tropicales que se comercian a gran escala. Argentina, Bolivia, Chile, Perú y Uruguay, lo conocen como palta mientras que Centroamérica y México como aguacate.

A nivel de Latinoamérica en Perú existen factores que afectan desempeño de agricultores en exportación. Hay problemas que se asocian con características del producto, poco conocimiento en comercio internacional, mala praxis agrícolas, y escasos programas de agricultura que ayuden a automatizar la selección para reducir el tiempo de trabajo (Portafolio, 2018)

Problemas para exportar Palta, derivan de alteración en cosecha y poscosecha de agricultores, que genera problemas de calidad, no cumplir demanda de palta para distribución internacional. Aunque producción tiene tasa de crecimiento anual 32%, y total 95% es para exportación. (Armando, 2019)

La selección del fruto es manual, con 6 a 8 operarios en fechas de alta producción que laboran 15 horas al día. Por ser manual, subjetivo y error humano, dato estadístico de empresa, 5 a 8% paltas mal clasificadas (Bermeo, 2019).

Los productores de la palta en la actualidad optan por realizar la clasificación y selección de manera rudimentaria apelando a la intuición humana, esto muchas veces conlleva la falta de precisión y confiabilidad.

Todo ello genera imprecisiones y errores en la clasificación de la palta y en consecuencia generando pérdidas de ingreso para el productor que muchas veces

realizan grandes esfuerzos con el objetivo de obtener ganancias que permita brindar buena calidad de vida al seno familiar.

El índice nacional hasta septiembre del 2020, afirma que creció 32,4% comparada con mes del año pasado, sustentando en incremento en superficie cosechada y buena condición que favorezca fase de maduración. (INEI, 2020)

Si bien la producción a nivel del departamento de Apurímac se redujo, no obstante, se pronostica que la tendencia a nivel nacional va en aumento. Cabe resaltar que actualmente existen máquinas industriales que permiten la clasificación de la palta por su tamaño, no obstante, para el pequeño productor no le es posible adquirir estas grandes máquinas al mismo tiempo estas grandes soluciones no permiten generar un análisis del estado de la palta. El sistema de clasificación propuesto, es un método de selección y/o clasificación de los frutos (paltas) que permite extraer y seleccionar los frutos aptos para el consumo humano.

1.2 Formulación de problema

Problema general

¿Cuáles es el diseño de sistema automático en reconocimiento de patrones para la clasificación de la palta en la comunidad de Mollebamba- Chincheros?

Problemas específicos

- ¿Cuál es el algoritmo aplicando la librería OPENCV en lenguaje de programación Python para estimar dimensiones de la palta?
- ¿Cuál el algoritmo aplicando librería OPENCV y SCIKIT-LEARN en lenguaje de programación Python para estimar el peso de la palta?
- ¿Cuál el algoritmo aplicando librería OPENCV en lenguaje de programación

Python para estimar el tamaño de la palta?

- ¿Cuál el algoritmo aplicando librería OPENCV en lenguaje de programación Python para estimar el porcentaje de enfermedad o daño de la palta?
- ¿De qué manera sistema automático de reconocimiento de patrones para clasificación de la palta estima el precio?

1.3 Objetivos

Objetivo general

Diseñar sistema automático en reconocimiento de patrones para optimización de la clasificación de la palta en la comunidad de Mollebamba- Chincheros.

Objetivos específicos

- Diseñar el algoritmo aplicando la librería OPENCV en lenguaje de programación Python para optimización de la estimación de las dimensiones de la palta
- Diseñar el algoritmo aplicando la librería OPENCV y SCIKIT-LEARN en lenguaje de programación Python para optimización de estimación el peso de palta
- Diseñar el algoritmo aplicando la librería OPENCV en el lenguaje de programación Python para optimización de la estimación el tamaño de la palta
- Diseñar el algoritmo aplicando librería OPENCV en lenguaje de programación Python para optimización de estimación en el porcentaje de enfermedad o daño de la palta
- Estimar el precio mediante sistema automático de reconocimiento de patrones para clasificación de la palta

1.4 Justificación

Brinda alternativa de solución viable que cumple con los requerimientos esenciales para la clasificación y selección de la palta para pequeños y medianos productores agrícolas por alta demanda ascendente a nacional e internacionalmente, el diseño y la implementación del sistema automático de reconocimiento de patrones para generar una categorización adecuada de la palta garantizando una producción óptima y una precisión alta al clasificar, con el fin de brindar mayores herramientas y facilidades al productor.

El estudio se apoya para la categorización y selección de la palta en una norma de calidad nacional, que nos indica un criterio de estandarización para el consumo nacional, basándose en lo argumentado esta aplicación de reconocimiento de patrones, nos permitirá profundizar y enriquecer los conocimientos sobre las diferentes tecnologías del procesamiento digital de imágenes.

Asimismo, con este desarrollo del sistema de clasificación de la palta se pretende mejorar la producción de pequeñas y medianas empresas agrícolas a nivel nacional y a la vez ser una fuente de investigación para estudiantes o para investigadores que quieran inmiscuirse en la rama del procesamiento digital de imágenes.

1.5 Viabilidad

Viabilidad Técnica

Viable porque se tiene conocimiento para realizarlo, ya que hay herramientas de libre acceso y además se cuenta con información técnica de apoyo en libros y sitios web.

El diseño y la implementación del presente proyecto se basan en la Norma Técnica Peruana 2014 que indica los parámetros de la palta para determinar la calidad.

Viabilidad económica.

En el presente proyecto las herramientas de desarrollo a utilizarse como el hardware son de fácil acceso a nivel nacional y de un costo muy módico, y en cuanto a los algoritmos a aplicar en este proyecto son de libre acceso ya que son pertenecientes a las herramientas catalogadas OpenSource, y por lo tanto el proyecto a presentarse es económicamente viable.

Viabilidad Social.

El presente proyecto tendrá como resultado un software automatizado para la clasificación y selección de la palta, que permitirá a pequeñas y medianas productoras de palta realizar una selección y clasificación de la palta según su tamaño y color de manera que se emplee menos recursos y gastos.

El software automatizado tendría muchas aplicaciones como en:

- Clasificación: clasificación de la palta según la Norma técnica de la palta.
- Generalización: se podría ampliar las capacidades y adecuar el software para otros productos.
- Comercio: la información obtenida podría utilizarse con fines comerciales.

Descrito lo anterior concluye que estudio es socialmente viable.

1.6 Limitaciones

La obtención de imágenes digitales se realizará con la ayuda de una cámara ubicada en una determinada zona en condiciones controladas y estas posteriormente se enviarán a un centro de procesamiento en donde se realizará la extracción de las características con la ayuda de un ordenador y el software automático instalado. Cabe mencionar que el análisis con el software se limita exclusivamente a la variedad de HASS (PERSEA

AMERICANA, MILL), las anomalías que se pueden presentar en este proceso y que puedan impedir el normal desarrollo del análisis de la imagen son la mala calidad de las imágenes a consecuencia por factores como ruidos, falla en el sistema eléctrico, etc.

CAPITULO II

ANTECEDENTES

2.1 Antecedentes Locales

Ataucusi E. (2021), En su tesis titulado “Sistema automatizado en gestión de asistencia y control de temperatura con reconocimiento facial de trabajadores en empresa eléctrica Ambato EEASA” De universidad José María Arguedas con objetivo mejorar control de acceso, usando Raspberry Pi en automatización de registro de asistencia y evaluación de temperatura corporal. La metodología aplicada, cuantitativa, pre experimental y explicativo en 48 participantes. Resultando con Shapiro Wilk, distribución normal, contraste de hipótesis empleo t Student en muestras relacionadas, con significancia menor a 0.05, concluyen que sistema con reconocimiento facial y medición de temperatura corporal (RFMTC) afecta significativamente control de acceso de participantes en Trabaja Perú, Talavera

Cáceres, E (2018) En su estudio titulado “Aplicación Móvil de reconocimiento facial en sujetos con antecedentes de abuso sexual, Andahuaylas, Apurímac- 2018” En la Universidad José María Arguedas, con objetivo usar aplicación móvil de reconocimiento facial en sujetos con antecedente de abuso sexual. En metodología empleo Fase API que detecta, verifica, identifica rostros así como buscar rostros similares y su agrupación. Como resultado la aplicación Microsoft Azure, aplicada con xvi herramienta de programación Visual Studio y Android Studio, formato JSON, va comunicar aplicación móvil de reconocimiento facial con plataforma en nube alojara algoritmo de reconocimiento facial y rostro de personas entrenadas para reconocerlas.

2.2 Antecedentes Nacionales

Santa J. et al (2021) En su estudio “Reconocer patrones de imágenes con sistema de visión artificial MATLAB” De la Universidad San Martín, tuvo como objetivo mejorar reconocimiento con visión artificial, medir proceso reconocimiento, implementar visión artificial y medir relación entre reconocimiento de patrones y sistema de visión artificial. La aplicada, cuasi experimental, transversal, con 8 patrones. Resultando tomando 1 muestra con software de visión artificial promedio de conteo para muestras de personajes en grupo experimental es 10,75 segundos. Concluyen que tiempo de procesamiento para reconocer 8 patrones de grupo experimental 10,75 segundos y control 67,75 segundos con grado de relación entre reconocimiento de patrones y sistema de visión artificial 72 %.

Rivas, L (2020) Presento su estudio titulado “Diseño conceptual de sistema automático para clasificar paltas Hass” De Pontificia Universidad Católica del Perú, la cual tuvo como objetivo diseñar sistema automático compacto, competitivo y rentable para productores. Funcionamiento sinérgico clasificara producto en 5 calibres regulables. Resultado se tienen cálculos teóricos que prueban efectividad de proceso y viabilidad económica que demuestra poder diseñar sistema que clasifica paltas con métodos no invasivos. Concluyen que definió sistema de clasificación según visión por computadora y aire que disgregue producto final según categorías preestablecidas con funcionamiento reduce pérdidas e incrementa productividad de proceso

Bermeo & Zorilla (2019) En su estudio titulado “Desarrollo de sistema automático para seleccionar paltas Hass según índice calidad para empresa Agroindustrias Verde flor S.A.C.” De Universidad Peruana de ciencias Aplicadas. Con el objetivo de proponer desarrollo e implementar de sistema automático para seleccionar paltas Hass, según

calidad (rozamiento, quemaduras y manchas negras). Para ello la metodología es integra hardware y software para automatizar y mejora selección de paltas. Resultando disminución de probabilidad de error humano 5-8% a 3.3%, menos a 5% establecido con empresa, validación de software tiene concordancia muy buena, con Kappa de Cohen. A más número de pruebas, más concordancia entre analista y máquina. Concluyen que diseño prototipo de sistema automático para seleccionar paltas Hass según calidad de rozamiento, quemaduras y manchas negras o lenticelas.

2.3 Antecedentes Internacionales

Sánchez, H (2018) En su estudio titulado “Diseño e implementación de controlador con reconocimiento de patrón para mejorar abastecimiento de insumos alimentarios de galpones avícolas” de Politécnica Chimborazo, con objetivo implementar controlador con reconocimiento de patrones, que mejore abastecimiento de insumos alimentarios de galpones avícolas. Estudio cuasi experimental en ambiente definido, descriptiva y aplicativa, comparando hipótesis chi-cuadrado resultado 1,35 indica que diferencias no son significativas, pero beneficios en tiempo y presencia humana de galpón es notable. En reconocimiento de patrones evaluó ambiente sin seccionar con precisión 38,9% y en ambiente seccionado según AGROCALIDAD versión 2016, con precisión en reconocimiento 91,25%, Concluyen según fundamentos; espacio, altura y disposición de cámara para tomar imágenes necesita constancia porque al variar posición o altura reconocimiento no es óptimo.

Basantes N (2018) En su tesis “Desarrollo de plataforma tecnológica para gestión de seguridad en institución educativa, inicial, usando sistemas móviles, reconocimiento facial y sistemas de alertas” De Universidad Guayaquil , con el objetivo de hacer Sistema Integral de Gestión de Seguridad, para ingresar información de personas en

base de datos, identifique y asocie mediante reconocimiento facial y notifique estudiante cuando se retiró de institución educativa. Metodología SCRUM, que ofrece posibilidad de entrega parcial de componentes de aplicación en todo momento. Resultando que sistema envía notificaciones automáticamente por SMS, correo electrónico y WhatsApp, cuando alumno es entregado, usando OpenSource para envío de SMS y correos electrónicos, vía WhatsApp uso API WhatsMate. Concluyen que Aplicación para control de entrega de alumno hecha en Android, para ser portable y usando herramientas OpenSource (PHP, MySql y OpenCV) es conforme.

Aguilar & Castaño (2016) En su estudio “Diseño e implementación de sistema para reconocer patrones usando videocámaras” de Universidad Tecnológica de Pereira, tuvo como objetivo Implementar metodología para para reconocer automáticamente patrones de botellas plásticas usando vídeo cámaras. En metodología analizo modelos según histogramas para detectar patrones como rayones o golpes y crear aplicación para control de calidad usando descriptores según histogramas. Resultando en tablas de rendimiento promedio métodos con buen rendimiento: HOG $88,52 \pm 8,35 \%$, RGB $83,68 \pm 8,72 \%$, CMYK $84,8 \pm 8,83 \%$ y TSL $84,84 \pm 8,93 \%$. Concluyen que metodología usada para fusión de datos tuvo desempeño alto cuando se extrajo características de objetos, ya que combinar métodos mejora resultados de clasificación para mejorar calidad; mejor combinación entre métodos TSL&HOG tuvo eficiencia $89,70 \pm 7,8 \%$.

Montoya C et al (2014) En su estudio titulado “Sistema automático de reconocimiento de frutas según visión por computadora” Tuvo como objetivo presentar sistema de reconocimiento que identifique fruta tropical latinoamericana con visión por computador. Estudio comparo clasificadores KNN y bayesiano y modelos color RGB

y HSV, junto con características de tamaño y forma usados en estudios pasados de esta área, en Malasia, Brasil y EEUU. Establecieron características que se describieron mejor son valores medios de canales RGB y longitud ejes mayor y menor con clasificador bayesiano, teniendo exactitud 90% en pruebas, donde seleccionar no tiene mayor cantidad de variables y generar vector descriptor permite que clasificadores sean más acertada, por eso considere en variables debe haber valor bajo de dependencia. Concluyen probabilidad de acierto en mejor caso es alta, 90%; reconociendo 9 tipos de frutas.

Escobar (Rodriguez, 2016) En trabajo titulado “Diseño e implementación de herramienta de computación en análisis asistido espermática en bovinos con Técnicas Visión Artificial” En la Universidad de Cundinamarca, tuvo como objetivo diseñar e implementar herramienta computacional que establezca motilidad espermática de bovinos con visión artificial y procesamiento de vídeo. En metodología prepararon semen de toro según protocolos en laboratorios andrológicos y medicina veterinaria. Resultados mostraron diferencias de medida manual respecto a algoritmo con relación 0,966 confianza 95% de (3,058 – 3,19), medidas similares. La aplicación junto a técnicas de procesamiento digital de video puede generar tecnologías efectivas capaces de reemplazar mano humana y evitar inconvenientes.

Herrera & Medina (2015) con título “Diseño de sistema automático para seleccionar frutos de café con técnicas visión artificial” En Universidad Autónoma de Caribe, tuvo de objetivo diseñar sistema visión artificial para seleccionar frutos de café buen estado y separe de no aptos. Metodología deductiva. Resultado de análisis en frutos de café 87% de efectividad, con curva ROC. Concluyen que sistema discrimina entre café buenos de los malos, pero no es efectivo al 100%. Algoritmo para detectar brocas,

genero falsas lecturas.

(Rodríguez, 2016) En esta tesis se diseñó e implemento un software sobre una placa reducida, la misma que tiene la característica y la capacidad para identificar la madurez de granadilla, sistema de software consta de una placa reducida (Arduino) encargada de hacer procesamiento de imágenes, módulo de cámara con cual hace la recolección de imágenes. Del mismo modo el autor optó por utilizar lenguaje de programación Python y sus librerías de procesamiento en imágenes.

Según el autor los resultados obtenidos con el software desarrollado, se validaron comparando maduración de fruta, en 252 imágenes. Estas mismas arrojando las detecciones correctas hasta en 92% de exactitud.

(Hernández López, 2009) la tesis tiene como objetivo segmentar secuencia de video en tiempo real, para separar primer plano de fondo.

Está basada en la suposición de tener conocimiento del fondo, y que en toda la secuencia de video el fondo es estático. Para cumplir los requerimientos se utilizó diferentes tipos de algoritmos, basados en campos aleatorios Markovianos, además se agregaron tres partes importantes (“Modificación de la intensidad”, “Detección de sombras”, “y corrección de semejanza del color”).

CAPITULO III

MARCO TEÓRICO

Fundamentos teóricos de la palta

Generalidades

Posee forma periforme y redonda, verde oscuro a negro dependiendo tipo y madurez.

Pulpa consistente entre amarillo y verde claro; rico en calorías, minerales y vitaminas.

La industrialización de la palta genera mucho interés puesto que se utiliza en productos como el aceite y el puré. El aceite que se emplea en hacer cosméticos, jabones, cremas de belleza.

Posee contenido alto de aceites vegetales, aporte nutricional es una buena fuente de energía y grasas saludables, así mismo es antioxidante.

Tabla 1 Valor Nutricional de la palta

N.º	Componentes	Por 100 gramos
1	Energía	131 kcal
2	Agua	79.2 g
3	Proteína	1.7 g
4	Grasa Total	12.5 g
5	Carbohidratos	5.6 g
6	Fibra dietaría	6.7 g
7	Calcio	37 mg
8	Zinc	0.64 mg
9	Hierro	0.6 mg
10	Retinol	7 ug
11	Vitamina A	7 ug
12	Tiamina	0.09 mg
14	Niacina	0.82

Nota. Super Food Perú.

Norma técnica peruana de palta

(NTP, 2014) Documento con definición, requisito, especificación de calidad que debe reunir la palta para su exportación. Esta norma aplica para todos los cultivadores de la Persea americana (HASS) que habrán de proveerse frescos, pero no aplica a paltas de elaboración industrial.

Requisitos mínimos de calidad

Según la NTP 011.018:2014 en todas las categorías se deben cumplir:

- Que este sana, excluyendo los afectados no aptos para consumo.
- Que este entera.
- Limpias y exentas de cualquier materia.
- Exenta de plagas que altere aspecto.
- Sin daño por plagas.
- Sin humedad externa anormal.
- Sin daños por bajas y/o altas temperaturas.
- Péndulo de longitud no más a 10mm, corte limpio.

No obstante, ausencia no considera defecto, si lugar de inserción de péndulo seco e intacto.

Categorización de la palta según aspectos externos (Persea Americana MILL).

Según NTP 011.018:2014 las paltas de variedad Persea americana Mill (HASS) clasificadas en 3.

- Categoría extra.

Palta de calidad superior.

- 1) Aspecto característico de variedad.

2) Sin defectos, solo superficial leve sin afectar aspecto de producto, calidad, en envase.

3) Debe presentar pedúnculo e intacto

➤ Categoría I.

Paltas de buena calidad, con defectos leves que no afecte aspecto de producto, calidad, estado en envase:

1) Defecto leve en forma y color.

2) Defectos leves de cáscara (suberosidad, lenticelas ya sanadas) y quemaduras solares; no más de 4 cm².

3) Sin afectación de pulpa.

➤ Categoría II.

Paltas no clasificadas en categorías superiores, con requisitos mínimos especificados en 5.2. Permite defectos leves mientras conserve rasgos esenciales de calidad, conservación y presentación:

➤ Defectos en forma y color;

➤ Defectos leves de cáscara (suberosidad, lenticelas ya sanadas) y quemaduras solares. Superficie afectada no más 6 cm². Sin afectación de pulpa.

Clasificación por calibres.

La norma técnica nos proporciona una equivalencia entre el peso y calibre de palta para exportarlo a dos principales mercados mundiales (estadounidense y europea).

Clasificación según calibre tolera 10% de peso total correspondiente a calibre superior o inferior toda ésta aplicando para todas las categorías.

Tabla 2 *Clasificación según calibre*

Código de Calibre	Peso en gramos
4	781-1220
6	576-780
8	456-576
10	364-462
12	300-371
14	258-313
16	227-274
18	203-243
20	184-217
22	165-196
24	151-175
26	144-157
28	134-147
30	123-137
S3	80-123

Nota. Clasificación. Fuente: NTP

Tabla 3 *Clasificación según calibre indiferente*

Código de Calibre	Peso (en gramos)
24	447-532
28	390-447
32	333-397
40	298-354
48	269-326
60	177-213
70	135-177
84	106-134
96	99-106

Nota. Clasificación. Fuente: NTP

Fundamento en procesamiento de imágenes

Opencv

Librería OpenCv (Open Computer Vision Library) se creó el 2000 por Intel Corporación. Función Linux, Windows y Mac OSX, OpenCv es librería Open Source escrita C y C++ se diseñó para eficiencia computacional y para enfoque en aplicaciones de tiempo real en visión artificial, teniendo como objetivo promocionar uso comercial con investigación abierta y libre.

Imagen digital

Imagen expresada en función $f(x,y)$ X e Y con coordenadas en 2 dimensiones, f en punto par será intensidad o nivel gris de imagen, también se tiene el Pixel como unidad básica de imagen. Cuando X, Y y valores f finitos y discretos, imagen será digital y cada punto de imagen es pixel (Gonzales & Woods, 2008).

La manera de representación datos en archivo imagen, hay 2 modos principales para

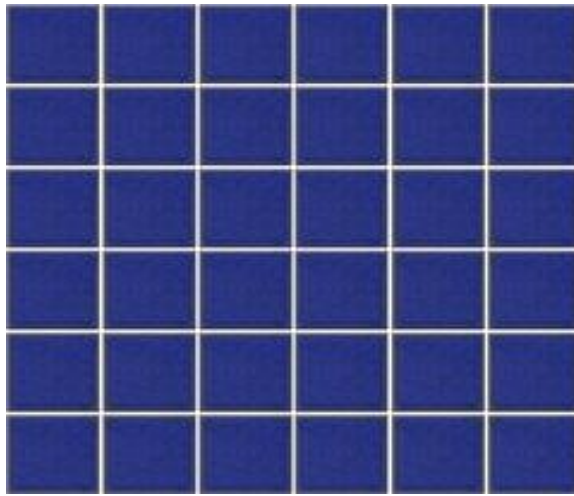
manejar información de imagen digital, uno imagen de mapa de bits, dos, vectoriales.

Imágenes de mapa de bits.

Bitmaps formada por rejilla de celdas. Cada celda es llamada pixeles, se le da valor de color y luminancia propio.

Al ver muchos píxeles, da ilusión de imagen con tono continuo.

Figura 1 *cuadro de pixeles*



Elaboración propia

En la imagen anterior se puede observar una rejilla de 6X6 pixeles, es así que al crear bitmap se crea rejilla específica de píxeles. Cuando cambia de tamaño, altera distribución y color de píxeles, por eso bitmaps se son de tamaño determinado y pierde calidad al altera dimensión.

Imágenes vectoriales.

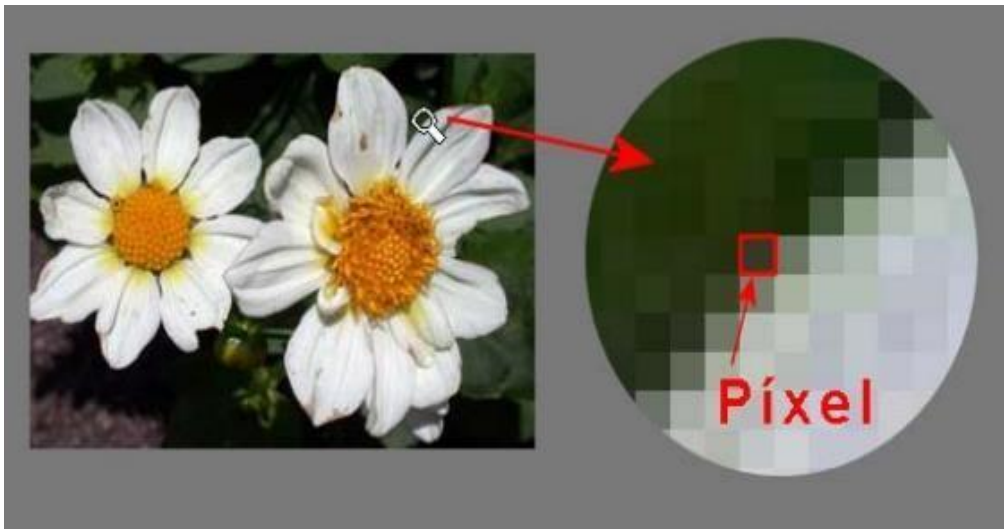
También llamados gráficos orientados a objetos, tiene la característica de que su tamaño es mucho más reducido con respecto a bitmap, porque como organiza información de imagen a más simple. Simplicidad producto de generar objetos como: curvas y líneas, con trazos geométricos establecidos con cálculos y matemáticas.

El Píxel.

Unidad mínima de visualización de imagen digital, capturadas con cámaras digitales

que capturan como cuadrícula de píxeles.

Figura 2 *pixel*



Nota. (Hernández 2009)

(Hernandez, 2009) Considera las siguientes características de un píxel.

- Forma cuadrada.
- Profundidad de color.
- posición relativa a otros píxeles de bitmap.

Profundidad de color

Cantidad de bits que codifican y guarda información de color de píxel en imagen. Bit es posición de memoria de 0 ó 1, a más profundidad imagen tendrá más colores.

Modos de color

Coordenadas que describen color de píxel usando números, modos de color más conocidos son:

Modo monocromático.

Tiene la característica de profundidad de color de 1 bit. Imagen blanco y negro.

Figura 3 *Imagen blanco y negro*



Nota: Elaboración propia

Modo escala de grises.

Administra canal negro con hasta 256 tonos de gris entre blanco y negro, la manera en que se convierten las imágenes de color a escala grises mediante equivalente E. la forma de establecer esta equivalencia es como promedio de valores de 3 componentes de color.

Ecuación 1 Equivalencia de Componente de Color

$$E(x, y) = \frac{R(x, y) + G(x, y) + B(x, y)}{3}$$

Figura 4 *Tono de gris entre blanco y negro*



Nota: Elaboración propia

Modo RGB.

Color producto de combinar 3 canales: Red (rojo), Green (verde), y Blue (azul).

Cada uno con valor 0 y 255, pudiendo tener hasta 16,7 millones de colores.

Figura 5 *Modo RGB.*



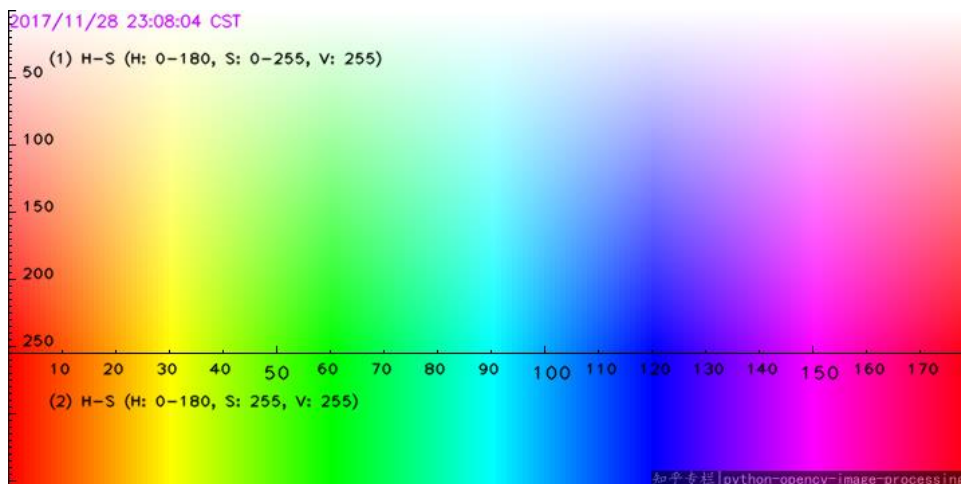
Nota: Extraído de google

Modo HSV.

Este modelo de color se define a través de tres componentes tono (H), saturación(S) y luminosidad (V) gracias a este modelo se parece más a la percepción del humano, y debido a ello es fácil de ajustar un color.

El tono se da como una representación angular en el rango de 0° - 360° , así mismo tanto la saturación y la luminosidad varían en una representación porcentual de 0-100, donde 100 para la saturación representa un color puro.

Figura 6 *Modo HSV*



Nota: Elaboración propia

Procesamiento digital de imágenes.

Es aplicar una gran cantidad de eventos que implica modificar valor de pixeles, con el objetivo de medir y cuantificar. En estos procesos intervienen la corrección de defectos, filtrado, realce de la imagen, la umbralización y las operaciones matemáticas.

“El procesamiento de las imágenes surge como consecuencia de la falta de perfección en la captura.”. Cabe destacar que ni aun teniendo los equipos más sofisticados se puede garantizar al cien por ciento la ausencia de defectos como: el ruido producido por factores climatológico, eléctrico, de iluminación. Del mismo modo se debe incluir las condiciones del ambiente de la muestra a capturar, ya que la luz incidente será distinta en diferentes horas del día, procesos para el cual se deberá aplicar diferentes técnicas de iluminación.

Una de las ramas de procesamiento digital más aplicada y avanzada es visión artificial, que trata sobre aprovechar la capacidad de procesamiento de la computadora para emular el ojo humano, incluyendo la capacidad de aprendizaje e interpretación. Visión por computadora emplea software de reconocimiento de patrones, clasifica objetos en imagen. (González & Wood, 2008).

Detectores de bordes

Un detector de bordes es la definición de fijar límites de un objeto a analizar, detectando discontinuidades. Estas discontinuidades son detectadas a través de filtros especiales.

Desplazamiento y sustracción.

2 píxeles adyacentes con intensidad diferente, por ello si se quita a imagen lo mismo desplazado horizontal o verticalmente se tiene imagen con bordes en eje perpendicular a desplazamiento.

Laplaciano.

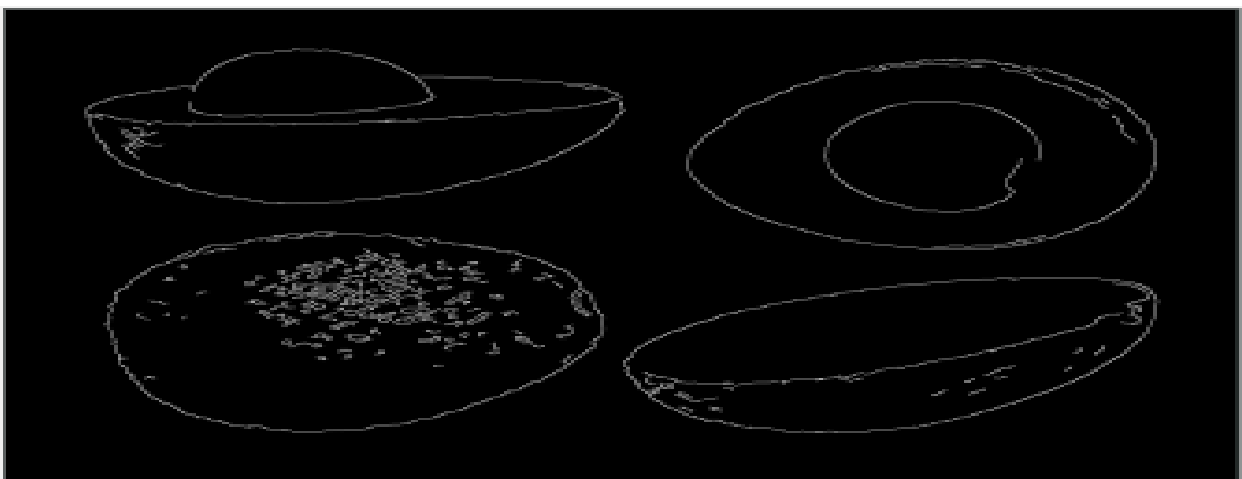
Estos filtros se basan en segunda derivada, se consideran como filtros direccionales que participan en toda dirección.

$$M = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Canny.

Método que emplea gradiente de imagen para realzar bordes.

Figura 7 *Extraído de Google*



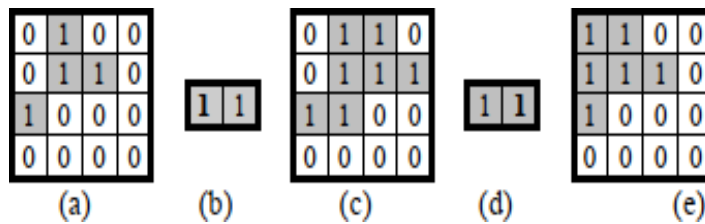
Operaciones Morfológicas

Simplifica datos de imagen, preservando característica esencial y quitando partes irrelevantes. Son las más comunes que dilatan y erosionan.

Dilatación

También conocida como adición de Minkowski, genera puntos, producto de sumar pares de puntos, uno de X y otro de porción desplazada. Origen o centro del último coincide con píxel de valor 1 de imagen (González & wood, 2008)

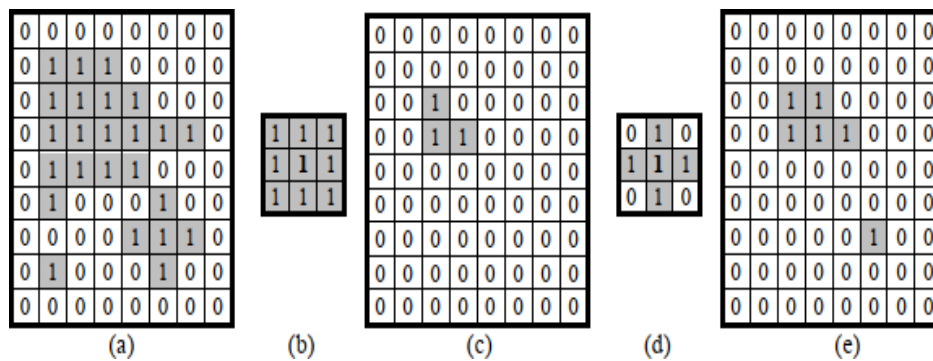
Figura 8 *Métodos y conceptos de la visión artificial*



Erosión

Operación de efecto contrario a dilatación, reduce tamaño de objetos moviendo elemento estructurante en imagen resultado será 1 por cada punto en estudio, si todos los puntos de elemento están dentro de conjunto de objeto.

Figura 9 *Métodos y conceptos de visión Artificial*

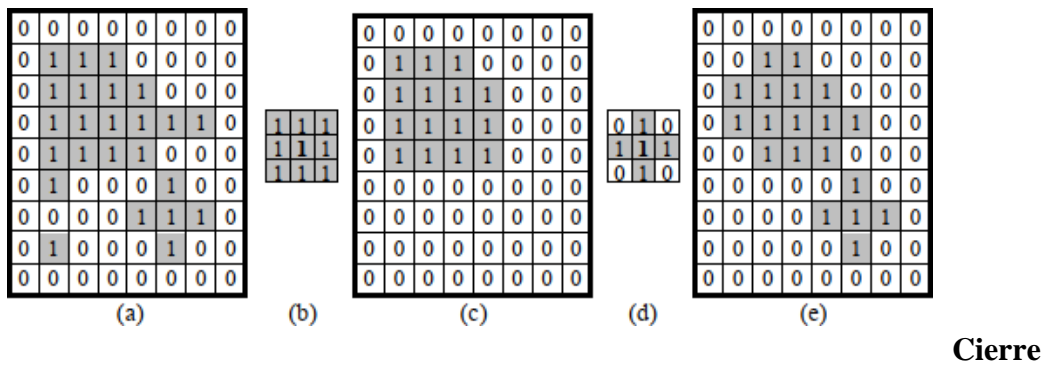


Apertura

Emplea erosión seguida de dilatación usando elemento estructurante. Erosionar hacer perder detalles mínimos, quita ruido, mengua imagen.

Luego al dilatar imagen crece, y detalles no perdidos resaltan.

Figura 10 *Apertura*



Proceso de dilatación seguida de erosión.

Ecuación 2 Cierre

$$cierre X \theta B = U\{B_v / B_v \cap X \neq \emptyset\}$$

Segmentación de imágenes

Implica fraccionamiento de imagen en segmento de píxeles que son homogéneos con respecto al criterio. Esto con la finalidad de adquirir partes relevantes de imagen. Partes adquiridas deben tener en común algunas características como el color, la textura, forma. Existen varios tipos de segmentación como:

Umbralización.

Define umbral como función que convierte imagen con tonalidades diversas en imagen blanco y negro. Si imagen original es $f(x,y)$, imagen umbralizada $g(x,y)$ y fija umbral $U(0 < U < 255)$.

Ecuación 3 Umbralización

$$(x, y) = 255 \text{ si } (x, y) > \text{umbral}$$

$$(x, y) = 0 \text{ si } (x, y) \leq \text{umbral}$$

Selecciona umbral que agrupe píxeles de imagen que pertenece a diferentes objetos de misma imagen diferenciándolo de fondo. Así segmentación según histograma basada

en elección de 1 o muchos umbrales que agrupen puntos de imagen en zonas con rasgos similares según nivel de gris.

Ecuación 4 Umbral de la Imagen

$$T = T[x, y, p(x, y), p(x, y)]$$

Donde $f(x,y)$ es intensidad (x,y) y $p(x,y)$ representan propiedad local medida en entorno de vecindad al punto. Imagen obtenida al procesar umbral obtenida con ecuación:

Ecuación 5 Segmentación

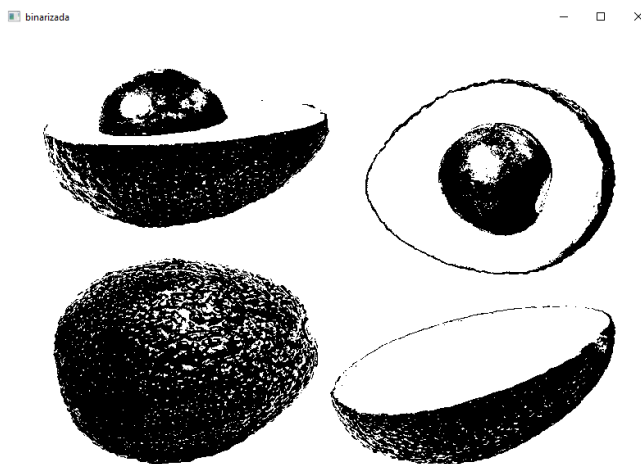
$$G(x,y) = 1 \text{ si } f(x,y) \leq T \text{ \& } f(x,y) \leq T$$

Píxeles marcados con 1 son objetos, píxeles marcados con 0 son fondo.

Umbrales más usados:

- **Umbral Global.** T depende de $f(x,y)$.
- **Umbral Local.** T depende de $f(x,y)$ y de $p(x,y)$.

Figura 11 *Extraída del libro métodos y conceptos de visión Artificial*

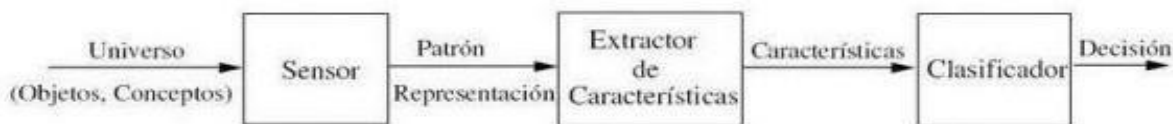


Reconocimiento de patrones

Procesar información de relevancia práctica que soluciona problemas, que puede hacerlo humano empleando computadoras, que tiende a ser difícil. (Bishop, 2007)

Se reducen dificultades haciendo métodos y algoritmos, objetivo primordial de reconocimiento de patrones como disciplina de Ingeniería de Sistemas. (Theodoridis & Koutroumbas, 1999)

Figura 12 Metodología de Desarrollo del Software.



Proyectos software y otros, por ser complejos necesitan técnicas para gestionarlos y ejecutarlos. Métodos de desarrollo de software son forma de interpretar realidad. Las metodologías destinadas al desarrollo de software se consideran un framework o marco de trabajo que consiste en planificar y controlar procedimientos de creación de un software.

En este campo de desarrollo software hay 2 grupos de metodología, tradicionales (formales) y ágiles.

Tradicionales son rígidas, porque piden definir el total de requisitos empezando proyecto. Ciclos del desarrollo son un poco flexibles y no dejan hacer cambios en el proceso de desarrollo de software. Metodologías Ágiles enfocan esfuerzo en capacidad de cambios, habilidad del equipo y buena relación con usuario.

Ambas con pros y contras; estos marcos de trabajos indican al desarrollador ¿Qué hacer? Más no ¿Cómo hacer? Método que se elija se debe adaptar a contexto de proyecto, según recursos técnicos, humanos; tiempo de desarrollo y tipo de sistema.

Metodologías ágiles.

Basado en pronta entrega de software incremental, deriva de desarrollo iterativo durante ciclo de vida de software.

Estas metodologías son flexibles y se modifican según realidad de equipo y proyecto. La característica principal es que proyectos grandes se subdividen en proyectos pequeños con lista de características. Los proyectos se tratan independientemente y crea características por un periodo de tiempo que esté en un intervalo de 2 a 6 semanas. En esta metodología la comunicación con el cliente es constante ya que para el inicio y todo el ciclo del desarrollo del software es inherente la presencia del cliente. Desarrollo de proyectos son colaborativo, y se adaptan a cambios; ya que los cambios en los requerimientos son una de las características esperadas, como entregas constantes a cliente y retroalimentación. (Agil Manifiesto, 2001)

➤ Scrum

Proceso empírico, iterativo e incremental de desarrollo de software, tiende a manejar procesos. Interactúa con las siguientes fases dividiendo los procesos.

1. Quien y que.

Identifica roles y responsabilidades de cada integrante del equipo de desarrollo.

Figura 13 *Extraído de platzi*



Equipo scrum data de 3 roles:

- Product owner/ Dueño de Producto. Voz de cliente y responsable de organizar, desarrollar, mantener tareas en desarrollo del proyecto.
- Scrum Master. Se asegura que trabajo de equipo vaya bien según bases del scrum.
- Development Team Members. Son los integrantes de equipo de desarrollo que escriben y prueban códigos.

Cabe mencionar que existen roles no centrales, quiere decir que la participación de estos es indispensable para la realización del proyecto, sin embargo, no depende de ellos el éxito y el fracaso del proyecto. Este rol viene representado principalmente por: Stakeholder (Cliente).

2. El Dónde y el Cuándo.

Dentro de esta Frase se puede encontrar al Sprint que básicamente se refiere a la unidad mínima de trabajo para el equipo de desarrolladores, es decir interacción entre el equipo; dura de 1 a 6 semanas donde scrum trabaja creando entregables. Este empieza con reunión para planear en el que se analiza según la prioridad y posteriormente se agrega al sprint Backlog, listado de tareas a implementar en sprint que empieza.

3. El por qué y el Cómo.

Esta fase engloba las herramientas que intervendrán en el desarrollo del proyecto, estas pueden ser un bug, requerimientos del proyecto, historias de los usuarios o el panel que muestra las tareas para el equipo de desarrollo del proyecto. (Platzi, 2015).

3.1 Lenguaje de programación Python

Python es lenguaje de programación creado por Guido van Rossum a inicios de los 90, siendo tecnología abierta y libre con ventajas sobre tecnología propietaria. (Ortiz, 2010)

Además, es lenguaje muy usado en aplicaciones web, desarrollo de software, ciencia de datos y machine learning (ML). Desarrolladores usan Python por ser eficiente y fácil aprender. Software Python se descarga gratis, se integra con cualquier sistema e incrementa velocidad. (A.W.S, 2023).

Principales características del lenguaje Python

- Para objetos, soporta estilos de programación procedural y funcional.
- Corre en muchas plataformas, Windows, Mac OS y Linux.
- Sintaxis y semántica sencilla.
- Usa tipos dinámicos.
- Bueno para programar scripts como aplicaciones de gran tamaño.
- Muy modular.
- Administración automática de memoria mediante recolección de basura.
- Amplia biblioteca de clases.
- Comunidad que promover desarrollo y adopción.
- Por ser interactiva, ideal para programación experimental y desarrollo rápido.

Desarrollo web de lado de servidor

Incluye funciones complejas backend que sitios web usan para mostrar información a usuario.

Python útil en escribir código de lado de servidor por su amplia biblioteca con código prescrito para hacer funciones backend complejas, por ejemplo, desarrolladores pueden hacer aplicación web esqueleto en segundos y se puede probar con herramientas de prueba de marco, sin necesidad de herramientas externas. (A.W.S, 2023)

Automatización con scripts de Python

Lenguaje scripting es lenguaje de programación para automatizar tareas que hace el humano. Programadores usan bastante scripts Python en automatización de tareas como:

- Cambiar nombre varios archivos a la vez
- Convertir archivo en otro tipo
- Quitar palabras duplicadas de archivo de texto
- Hacer operaciones matemáticas básicas
- Mandar mensajes por email
- Descargar contenido
- Análisis básicos de registros
- Descubrir errores en muchos archivos

Realizar tareas de ciencia de datos y machine learning

Extrae conocimientos relevantes a partir de datos, mientras enseña a computadoras que aprenden automáticamente de datos y hacer predicciones como:

Corregir y borrar datos incorrectos, limpieza de datos

- Extraen y seleccionan características de datos, agregando nombres relevantes a datos
- Buscar estadísticas a partir de datos
- Visualizar datos usando tablas y gráficos, como gráfico de línea, barras, circulares e histogramas

Desarrollo de software

Usan Python para hacer varias tareas de desarrollo y aplicación de software, como:

- Hacer seguimiento de errores en código de software
- Crear software automáticamente
- Administrar proyectos de software
- Hacer prototipos de software
- Hacer aplicaciones de escritorio con bibliotecas de interfaz gráfica de usuario (GUI)
- Crear juegos simples según texto a videojuegos complejos

Automatización de pruebas de software

- Se verifica si resultados reales de software coinciden con resultados esperados, garantizando que software no tiene errores.
- Desarrolladores usan marcos de prueba Python, como Unittest, Robot y PyUnit, para probar funciones.
- Para probar software usan Python para escribir casos de prueba en varios escenarios de prueba.
- Desarrolladores emplean herramientas para ejecutar scripts de prueba automáticamente, llamadas herramientas integración e implementación continuas (CI/CD), que ejecuta automáticamente scripts de prueba Python e informa resultado de pruebas.

3.2 SQLite: La Base de Datos Embebida

Usar base de datos se extendió de servidores a equipos móviles. Demanda de bases de datos en equipos móviles como PDAs y celulares creció bastante los últimos años por la necesidad de empresas de tener datos al instante y participar de la competencia. Esto generó que almacenamiento de datos dispositivos mejore capacidad de herramientas.

Por eso actualmente hay muchas bases de datos para celulares, como SQLite, enfocada por el artículo. (Filein, 2016).

¿Qué es SQLite?

Herramienta software libre, que almacena información en dispositivos, eficaz, potente, rápida y en equipos de bajo rendimiento de hardware, PDA o celular. SQLite emplea SQL92 y permite soporte desde consultas básicas a complejas de lenguaje SQL, pudiendo usarse en celulares y computadoras. (Filein, 2016).

Machine Learning

Machine learning es una sub área de inteligencia artificial, enfocada a extraer y aprender de los datos históricos características o patrones que permitan estimar y/o cuantificar.

(Amazon, 2020) “Machine learning busca crear sistemas que aprendan de datos y mejoren desempeño de tareas, sin programarlos explícitamente, y tomen decisiones con datos nuevos”.

Regresión lineal con Python

Modela relación entre variable continua y variables independientes ajustando ecuación lineal. Es regresión lineal simple con 1 variable independiente y lineal múltiple con más de una. Depende de contexto variable modelada será dependiente o de respuesta, y variables independientes como regresores, predictores o features. (Amat, 2021)

Definición matemática

Modelo regresión lineal (Legendre, Gauss, Galton y Pearson) conjunto de observaciones $\{y_i, x_{i1}, \dots, x_{in}\}_{i=1}^n$, media μ de variable respuesta y relaciona linealmente con variables regresores $x_1 \dots x_p$.

Ecuación 6 Regresión Lineal

$$U_y = b_0 + b_1x_1 + b_2x_2 + b_n$$

Resultado de ecuación llamada línea de regresión poblacional, recolecta relación entre predictores y media de variable respuesta.

Otra definición es:

Ecuación 7 Regresión Lineal Poblacional

$$U_y = b_0 + b_1x_1 + b_2x_2 + b_nx_n + \varepsilon$$

Hace referencia a valor de y para observación i concreta. Valor de observación puntual no será igual a promedio, por eso agrega término error ϵ .

En los casos, interpretación de elemento de modelo es igual:

b_0 : ordenada en origen, corresponde con promedio de variable respuesta y predictores es 0.

β_j : efecto promedio sobre variable respuesta incremento en unidad de variable predictora x_j , teniendo constantes las variables, llamadas coeficientes parciales de regresión.

ϵ : residuo o error, diferencia entre valor observado y estimado por modelo. Recoge efecto de variables que influyen en y pero no incluyen en modelo como predictor. (Amat, 2021)

Método más usado es ajuste por mínimos cuadrados ordinarios (OLS), identifica mejor modelo recta que reduce suma de desviaciones verticales entre datos entrenamiento y recta, elevadas a cuadrado. (Amat, 2021)

Análisis de normalidad con Python

Analiza si datos disponibles proceden de población con distribución normal. Hay 3 estrategias:

- Representación gráfica
- Método analítico
- Test hipótesis.

En el trabajo se muestra cómo usar estrategias para establecer si altura de grupo de personas tiene distribución normal. (Amat, 2021)

Métodos gráficos

El más usado en análisis de normalidad trata de representar datos en histograma y superponer curva de distribución normal con media y desviación estándar de datos.

Otro método son cuantiles teóricos (Gráficos Q-Q), que compara cuantiles de distribución con cuantiles teóricos de distribución normal con misma media y desviación estándar de datos. Mas aproximados los datos a una normal, más se alinean puntos entorno en recta. (Amat, 2021)

Relación lineal entre predictores numéricos y variable respuesta

Predictor numérico se relaciona linealmente con variable respuesta, mientras otros predictores son constantes, se comprueba representando residuos de modelo frente a cada predictor. En relación lineal, residuos distribuidos aleatoriamente entorno a 0.

Distribución normal de variable respuesta

Variable respuesta se distribuye normalmente, se comprueba en histogramas, cuantiles normales o test hipótesis normalidad.

Varianza constante de la variable respuesta (homocedasticidad)

Varianza de variable respuesta es constante en predictores. Se comprueba representando residuos de modelo con predictor. Si varianza es constante, distribuir aleatoriamente con misma dispersión. Distribución cónica indica ausencia de homocedasticidad. Se recurre también a contrastes homocedasticidad como test

Breusch-Pagan.

Razón de condición debido a: modelo lineal asume que variable respuesta Y tiene distribución normal, media μ se modela según variables (predictores) y varianza σ calculada con constante de dispersión y función $v(\mu)$. Representa varianza, no se modela en base a variables predictoras más bien indirectamente mediante relación con media y es único valor.

Ecuación 8 Varianza de la Homocedasticidad

$$\sigma^2 = \frac{\sum_i^n E^2}{n - p}$$

CAPITULO IV

METODOLOGÍA DE LA INVESTIGACIÓN

4.1 Tipo de investigación

Aplicada, estudia y aplica sobre trabajo en problemas concretos (Behar, 2008, p.20). Aquí se implementó sistema automático de reconocimiento de patrones para la clasificación de palta de los productores de la comunidad de Mollebamba.

4.2 Enfoque de investigación

Cuantitativo, según Hernández et al. (2014), recolecta datos basados en medición numérica y estudio estadístico

4.3 Diseño de investigación

No experimental, según Hernández et al. (2016) no manipula variables, solo observa fenómenos en ambiente natural para estudiarlos.

4.4 Nivel de investigación

Hernández, Fernández y Baptista (2016) en estudio descriptivo busca especificar propiedades, características y perfil de personas, grupos, comunidades o cualquier fenómeno sometido a estudio. Analizó las dimensiones, peso, tamaño, enfermedad y precio de la palta independiente o conjunta sobre conceptos o variables.

4.5 Población

Casos con especificaciones en común en espacio específico. Muchas veces no se puede analizar población por tiempo y recurso humano, por eso trabaja con “Muestra” (Chaudhuri, 2018). Población serán paltas recolectadas por productores Mollebamba – Chincheros.

4.6 Muestra

No probabilístico, elección de elementos no depende de probabilidad, si a propósitos (Hernández et al., 2014, p.176). Actividad elegida, por cercanía y fácil acceso de investigador, recolectar 114 frutos de palta. Para la evaluación.

4.7 Técnicas y análisis de datos

Para acopio de datos se hizo corpus que captura característica de paltas:

1. Obtener paltas de representativas a categoría (Primera, Segunda, Tercera).
2. Establecer características específicas, diferencias entre cada categoría de paltas.
3. Extraer características externas de tamaño, forma y color de palta con algoritmo procesamiento digital imágenes.
4. Construir corpus con estructura de rasgos extraídos y etiqueta de clase de paltas.

Estructura para construcción de corpus del crecimiento

- En vector de características indico atributos de imágenes de paltas.
- Rasgos de tamaño: Ancho, Altura, Perímetro, Área, Diámetro equivalente.
- Rasgos de forma: Relación de aspecto.
- Rasgos de color son medias de canales de color RGB, Gris, HSV y HSL de objeto en imagen.

4.7.1 Recolección de datos

Emplea librerías OPENCV y SCIKIT-LEARN OPENCV en lenguaje programación Python” (Andrade et al., 2019).

4.7.2 Análisis de datos

Empleo estadística descriptiva, inferencial e interpretación de gráficos estadísticos.

Según objetivos de estudio y escala de medición de variables, para evaluar normalidad de datos conseguidos, aplicó método kolmogorov smirnov.

4.8 Hipótesis de investigación

Hipótesis general

El sistema automático reconocimiento de patrones optimiza clasificación de palta de la comunidad de Mollebamba – Chincheros.

Hipótesis específicas

- El algoritmo aplicando la librería OPENCV en lenguaje programación Python automatiza estimación las dimensiones de palta.
- El algoritmo aplicando la librería OPENCV y SCIKIT-LEARN en lenguaje programación Python optimiza estimación el peso de palta.
- El algoritmo aplicando la librería OPENCV en lenguaje programación Python optimiza estimación el tamaño de palta
- El algoritmo aplicando librería OPENCV en lenguaje programación Python optimiza la estimación del porcentaje de enfermedad o daño de la palta.
- Automatizar la estimación del precio mediante sistema automático de reconocimiento de patrones para clasificación de la palta

4.9 Nivel de significancia y nivel de confianza

Gorgas et al. (2011) nivel de significancia es probabilidad de equivocación cuando rechaza hipótesis nula, cuando realmente es verdadera (error tipo I).

Significancia lo fija investigador y según Hernández et al. (2014), hay 2 niveles convenidos: 0.05 y 0.01, elegir 0.05, tiene 95% seguridad de no equivocarse y restante 5% en contra.

Este estudio toma 0.05 de nivel significancia, con confianza 95%.

4.10 Prueba de normalidad

Droppelmann (2018) para analizar normalidad de datos hay varios métodos, con más de 50 valores use prueba kolmogorov smirnov, p valor mayor igual 0,05 sí hay normalidad, menor distribución es no normal con significancia 5%. Según esto emplea kolmogorov smirnov, por muestra de más de 50.

4.11 Prueba estadística

Posibilidad de comparar variables cuantitativas continuas de distribución normal, con prueba t Student (Flores et al., 2017).

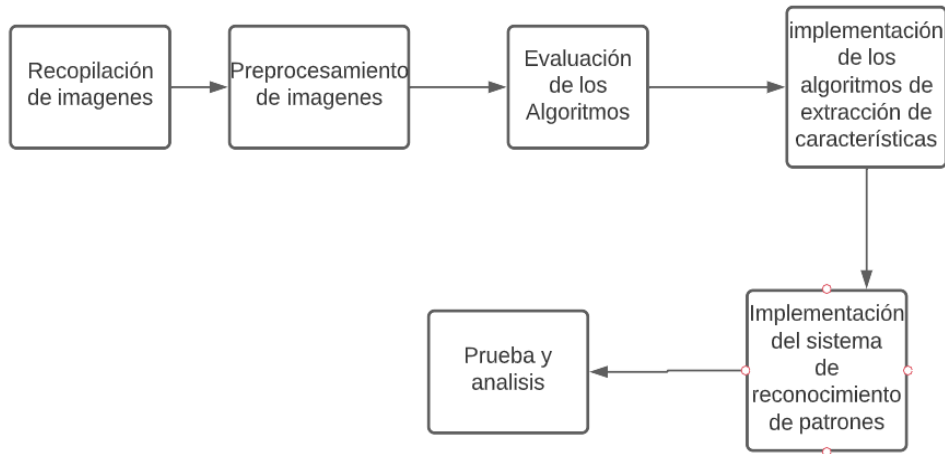
Hernández et al. (2014) indica que prueba t Student tiene nivel de libertad, para que dato varíen libremente y depende de cantidad de datos a comparar. Curva de distribución t se aproxima a curva normal si grado de libertad es más de 29 (Gorgas et al., 2011).

Tabla 4 *Operacionalización de variables*

Variable	Definición conceptual	Dimensión	Indicador
<p>V. Independiente</p> <p>Sistema automático reconocimiento de patrones</p>	<p>Procesar información de relevancia práctica que soluciona problemas, algunos de ellos resueltos por humanos con poco esfuerzo. A veces solucionarlos empleando computadoras se hace muy complicado. (Bishop, 2007)</p>	<p>Dimensiones</p> <p>Peso</p> <p>Tamaño</p> <p>Daño en porcentaje</p> <p>Precio</p>	<p>OPENCV</p> <p>SCIKIT-LEARN</p>
<p>V. Dependiente</p> <p>Clasificación de la palta</p>	<p>Según norma técnica peruana NTP 011.018:2014 las paltas de la variedad Persea americana Mill (HASS) hay 3 categorías</p>	<p>Categoría extra</p> <p>Categoría I</p> <p>Categoría II</p>	<p>.....</p>

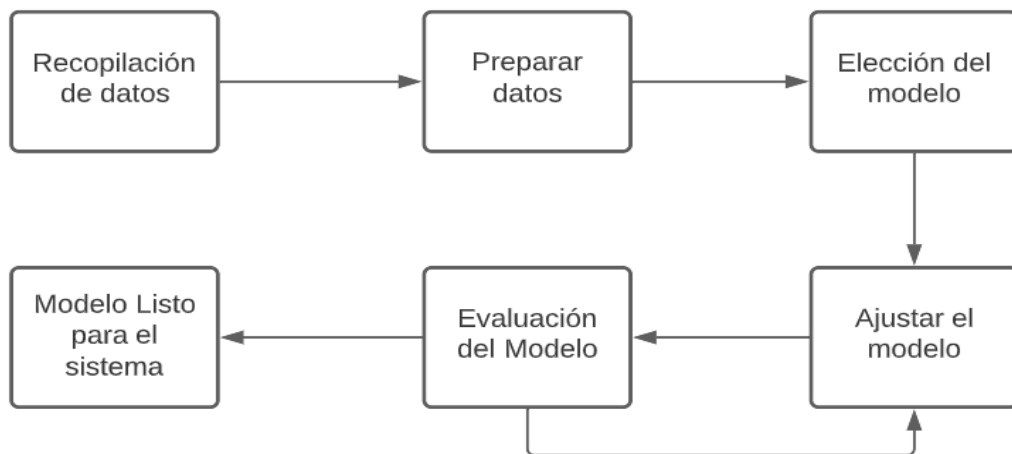
5.1 Diagramas de bloques de la propuesta de solución

Figura 14 Diagrama de bloques extracción de características y analisis de color



Nota. Fuente: Elaboración propia

Figura 15 Diagrama de bloques de entrenamiento del modelo de regresión lineal



Nota. Fuente: Elaboración propia

5.2 Recopilación de imágenes

Figura 16 *Cámara para capturar imágenes*

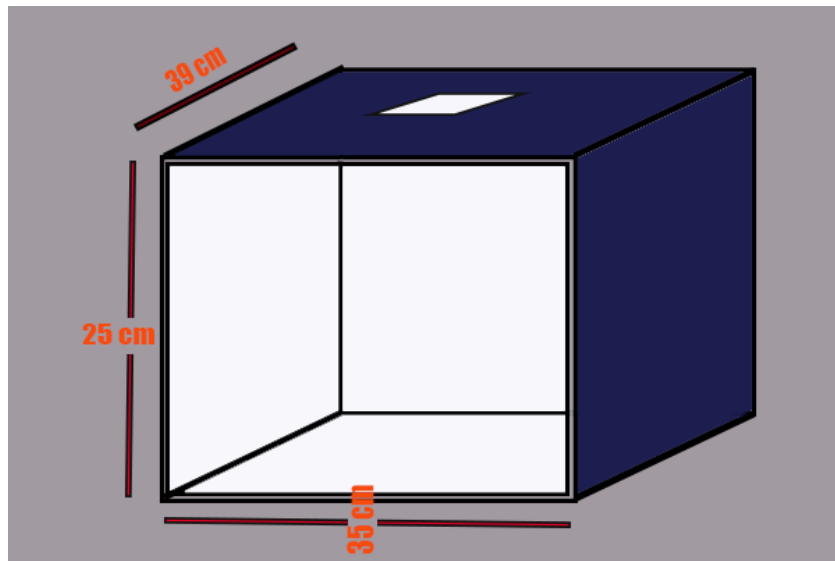


WEB CAM Logitech C922

5.3 Prototipo para captura de imágenes

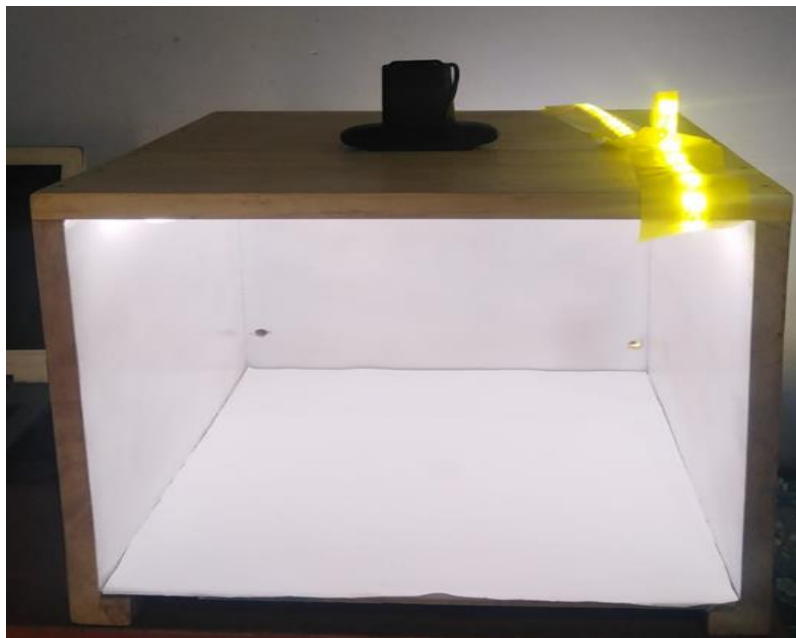
Para adquirir imágenes se hizo prototipo que consta de una caja cerrada, abierta por la parte frontal con un fondo blanco y en la parte superior una abertura para la cámara.

Figura 17 *Diseño del Prototipo*



Nota. Fuente: Elaboración propia

Figura 18 *Cámara de captura*



Nota. Fuente: Elaboración propia

5.4 Iluminación:

Para la iluminación se utilizó tira de leds, ya que la iluminación es más efectiva para nuestro caso.

Figura 19 *Iluminación led*



Nota. Fuente: Google

Unidad de central de proceso:

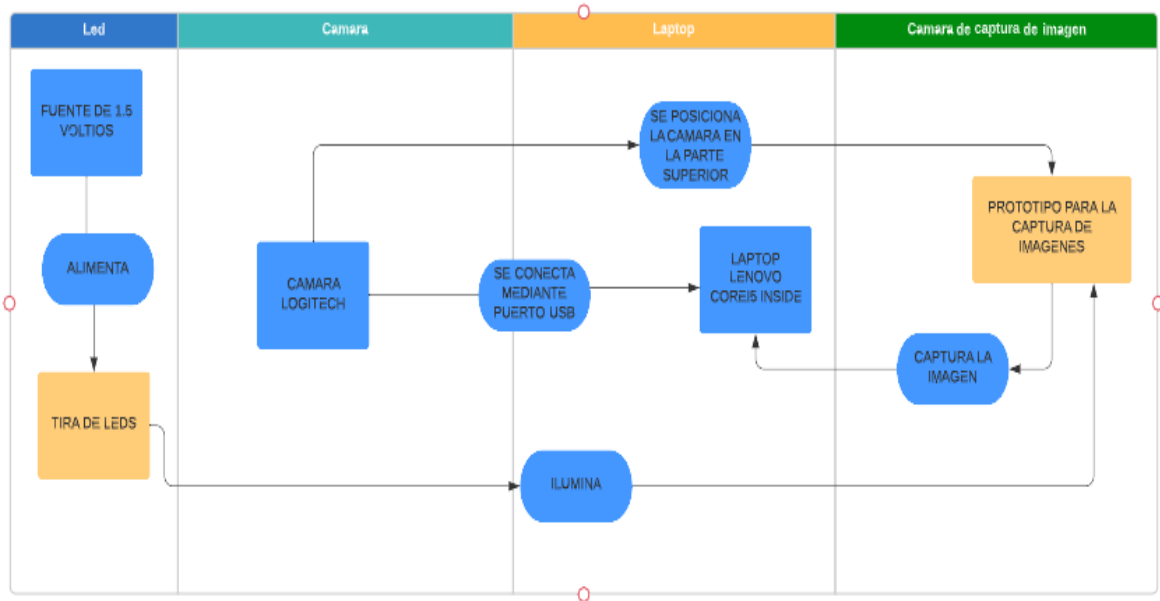
Para procesar imágenes y datos se empleó una laptop marca Lenovo, con las siguientes características.

Memoria: 4 GB.

CPU : Intel coreI5 inside.

Sistema operativo: Windows 10 de 64 bits.

Figura 20 *Flujograma de la instalación del hardware de sistema*

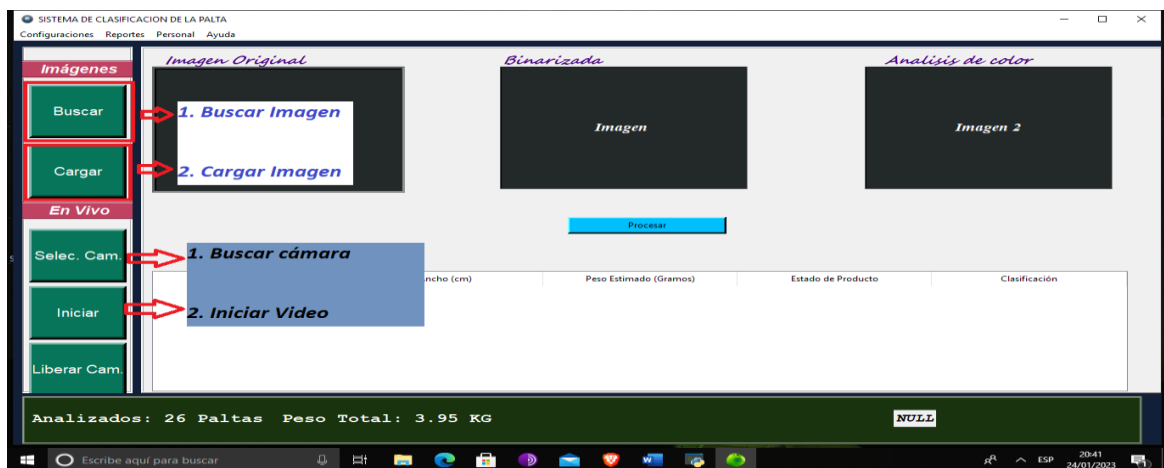


Nota. Fuente: Elaboración propia

Software para adquisición de imágenes

El software para adquisición de imágenes se hizo con el lenguaje de programación de Python v3.8.10 y la librería OPENCV 4.5.4

Figura 21 *Adquisición de imágenes*



Nota. Fuente: Elaboración propia

Código:

Buscar y abrir imagen

En el apartado imágenes el software permite buscar una imagen guardada en un disco externo, cargarla y posteriormente procesarla.

Figura 22 Apartado de imágenes

```
271 def abrirDireccion(self,hilo):
272     if self.EtiquetaUser.cget('text')!='NULL':
273         hilo=True
274         self.Address_video=filedialog.askopenfilename()
275     else:
276         msgI.showinfo('Alerta','Configure Propietario')
277
278 def abrirImagen(self):
279     if self.Address_video!=None:
280         self.imageComprobar=True
281         img_source=cv2.imread(self.Address_video)
282         img_source=cv2.cvtColor(img_source,cv2.COLOR_BGR2RGB)
283         #redimensionar para procesar
284         img_source=cv2.resize(img_source,(640,480),interpolation=cv2.INTER_AREA)
285         self.matriz_Image=img_source
286         img_source=cv2.resize(img_source,(int(img_source.shape[1]*0.5),int(img_source.shape[0]*0.5)),inter
287         #cv2.imshow('titulo111',img_source)
288         img_source=self.objVideo.lectura(img_source)
289         self.EtiquetaVideo.config(width="320",height="240")
290         self.EtiquetaVideo.config(image=img_source)
291         self.EtiquetaVideo.image=img_source
```

Nota. Fuente: Elaboración propia

Buscar cámara

En el apartado “en vivo” se puede seleccionar la cámara webcam a una velocidad de 30 fps, posteriormente mediante un botón permite procesar en tiempo real.

Figura 23 Apartado

```
FOLDERS
├── contratrafic
├── Mvc
├── _pycache_
├── Alcuatrec
├── Alcuatrecrv
├── images
├── images2
├── gitignore
├── Alcuatrecar
├── capturar.py
├── database.py
├── findCamera.py
├── fotogramas.py
├── generador.py
├── grafica.png
├── iconico
├── instaladorico
├── MITodoliteca.py
├── modelo.joblib
├── modelo1.joblib
├── modelo2.joblib
├── prueba1.py
├── prueba2.py
├── prueba3.spec
├── readme.txt
├── Report.py
├── Reporte.pdf
├── TopCamaras.py
├── usuario.py
├── VMMain.py
```

```
findCamera.py
4
5 class BuscarCamara:
6     """docstring for ClassName"""
7     ventana=None
8     def __init__(self):
9         self.numeroCamara=None
10
11     def Top_Camera(self):
12         self.top1 = tk.Toplevel(bg="Orange",relief="groove")
13         self.top1.title("Selección cámara")
14         self.top1.geometry("500x170+200+120")
15         self.top1.overridedirect(1)
16         self.top1.wm_attributes("-topmost",1)
17
18         labelCamara=tk.Label(self.top1,text="Seleccione una Cámara:")
19         labelCamara.config(width=20,bg="Orange",font=("",14),height=3)
20         labelCamara.place(x=0,y=20)
21
22         #agregando un select...
23         self.selectCamara=tk.Combobox(self.top1,font=("",14),height=3)
24         self.selectCamara["values"]=[0,1,2,3,4,5,6,7,8,9]
25         self.selectCamara.current(0)
26         self.selectCamara.place(x=250,y=40)
27
28         #agregando un boton...
29         botonAceptar=tk.Button(self.top1,text="Aceptar",bg="#FFF",font=("",14))
30         botonAceptar.config(width=9,command=self.event_Selection)
31         botonAceptar.place(x=160,y=100)
32         self.top1.grab_set()
33         #self.top1.mainloop()
34
35     def event_Selection(self):
36         self.numeroCamara=self.selectCamara.get()
37         self.top1.destroy()
```

Nota. Fuente: Elaboración propia

Leer video en video, capturar y procesar

La lectura de los fotogramas se realiza en el espacio de color BGR, posteriormente se realiza el escalamiento de la imagen para su posterior visualización en la pantalla

Figura 24 *Lectura de fotogramas*

```
232 def AbrirFotogramas(self, mirror=False):
233     if self.numeroCamara!=None:
234         self.cap=cv2.VideoCapture(int(self.numeroCamara), cv2.CAP_DSHOW)
235         while (self.cap.isOpened()):
236             ret,FrameMatriz_original=self.cap.read()
237             if ret:
238                 FrameMatriz_scalado=cv2.resize(FrameMatriz_original,(int(FrameMatriz_original.shape[1]*0.50),int(FrameMatriz_original.shape
239                 self.matriz_Image=FrameMatriz_original
240                 fotograma=self.objVideo.Lectura(FrameMatriz_scalado)
241                 self.EtiquetaVideo.config(width="320",height="240")
242                 self.EtiquetaVideo.config(image=fotograma)
243                 self.EtiquetaVideo.image=fotograma
244                 if self.hilo:
245                     self.hilo=False
246                     break
247                 self.cap.release()
248                 self.cap=None
249             else:
250                 msgI.showinfo("Alerta!!","Seleccione Camara!!")
251
```

Nota. Fuente: Elaboración propia

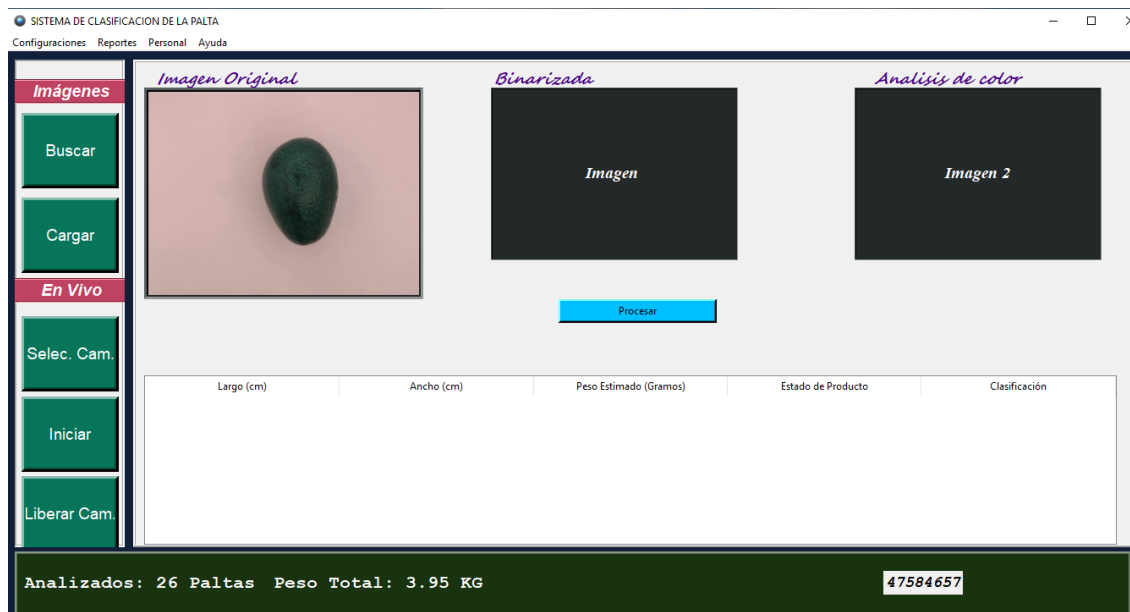
Conversión del fotograma al formato que reconoce la librería para su visualización en la pantalla

Figura 25 *Conversión de fotogramas*

```
22
23 def formato_Tkinter(self,matrix):
24     image=Image.fromarray(matrix)
25
26     image=ImageTk.PhotoImage(image)
27     return image
28
29
```

Nota. Fuente: Elaboración propia

Figura 26 *Interfaz de la captura de imagen*



Nota. Fuente: Elaboración propia

- a. Palta de categoría A
- b. Palta de categoría B
- c. Palta de categoría C

Figura 27 *Imágenes de las paltas sanas*



Nota. Fuente: Elaboración propia

Figura 28 *Imágenes de las paltas defectuosas*



Nota. Fuente: Elaboración propia

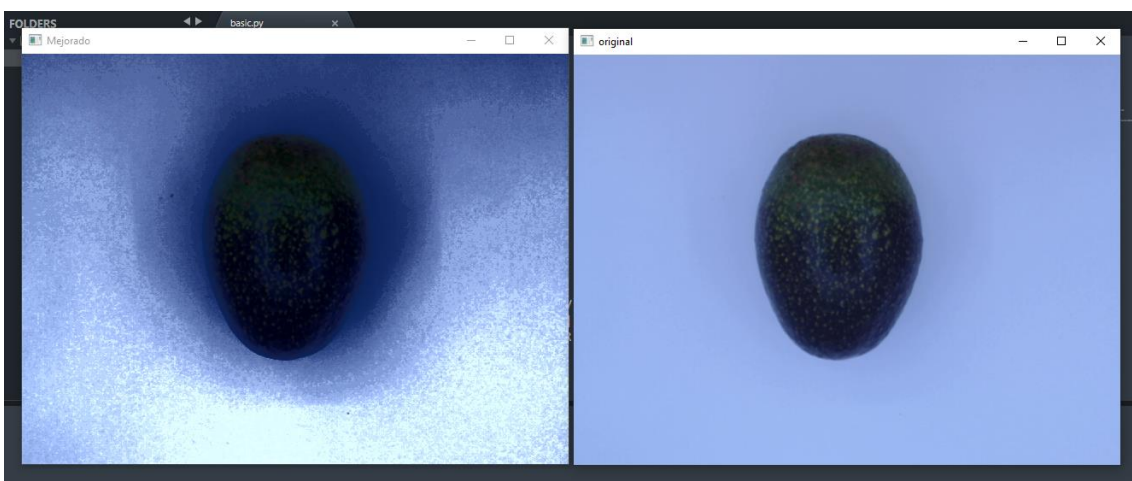
Preprocesamiento de Imágenes.

Mejoramiento del contraste de las imágenes.

```
33 def equalizar(self,img):  
34     #img=self.img.copy()  
35     img=cv2.cvtColor(img,cv2.COLOR_BGR2YUV)  
36     img[:, :, 0]=cv2.equalizeHist(img[:, :, 0])  
37     img=cv2.cvtColor(img,cv2.COLOR_YUV2BGR)  
38     return img  
39
```

Nota. Fuente: Elaboración propia

Figura 29 *Resultado:*



Nota. Fuente: Elaboración propia

Una vez mejorado la imagen se procedió a la conversión de imagen a escala grises,

seguidamente aplicó el filtro gaussiano para difuminar imagen con el objetivo de reducir y eliminar los ruidos que no son de interés, y finalmente se aplicó la umbralizacion para la binarizacion de imagen con objetivo de extraer región de interés.

Figura 30 *Imagen escala de grises*



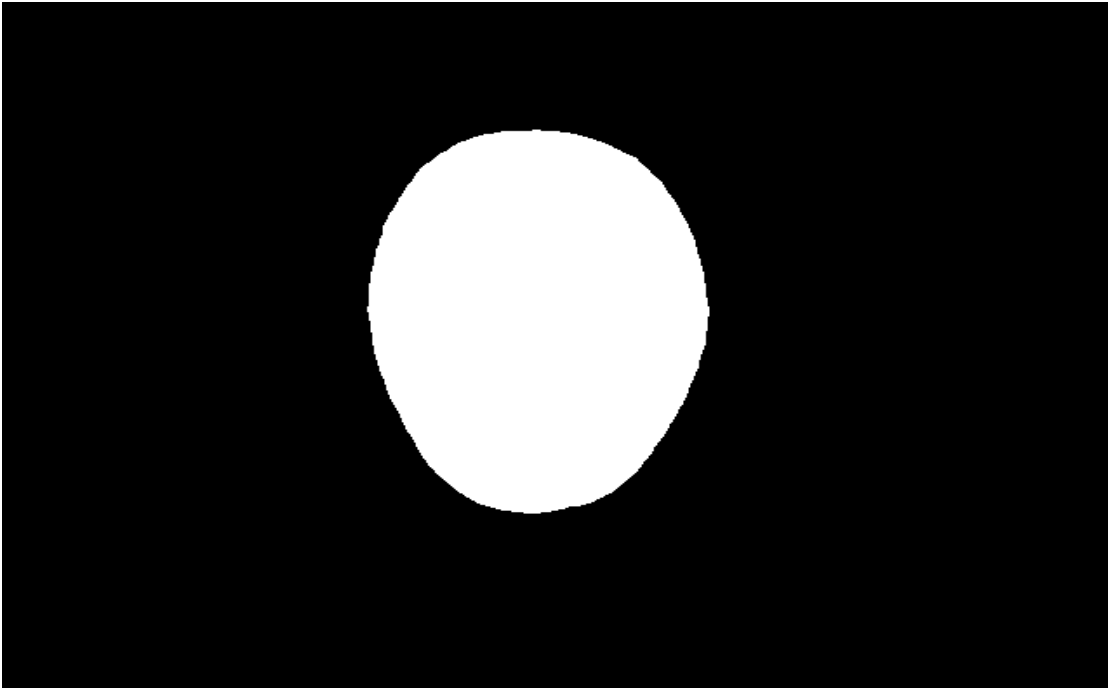
Nota. Fuente: Elaboración propia

Binarizado

Figura 31 *binarizar*

```
37
38
39 def detected_edges(self, img):
40     (T, thread)=cv2.threshold(img,100,255,cv2.THRESH_BINARY_INV)
41     (T1,thread1)=cv2.threshold(img,150,255,cv2.THRESH_BINARY_INV)
42     #realizar el bitwi
43     imagen_final=cv2.bitwise_or(thread,thread1)
44
45     #dilatacion
46     kernel=np.ones((5,5),np.uint8)
47     dilatado=cv2.dilate(imagen_final,kernel,iterations=2)
48     #detectando contorno
49     imagen_contorno=dilatado-cv2.erode(dilatado,kernel,iterations=1)
50     return imagen_contorno
51
```

Nota. Fuente: Elaboración propia



Nota. Fuente: Elaboración propia

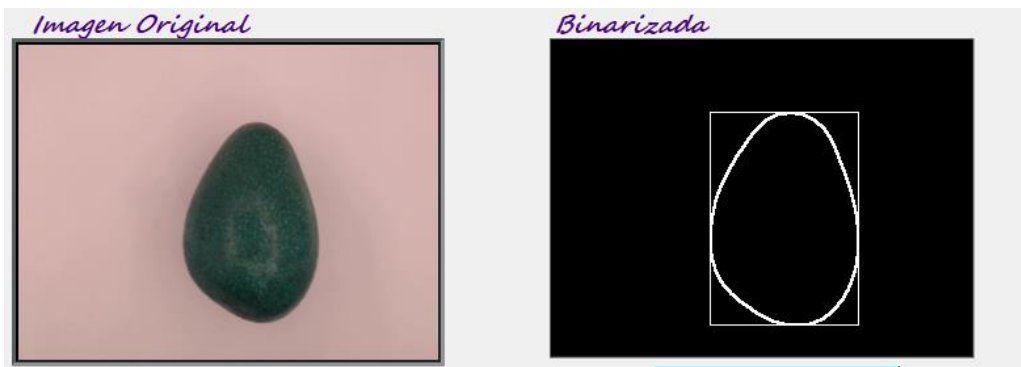
Contornos

Extracción de contornos y el cuadro mínimo delimitador (las medidas del cuadro delimitador servirán para estimar el peso)

Figura 32 *Extracción de contornos*

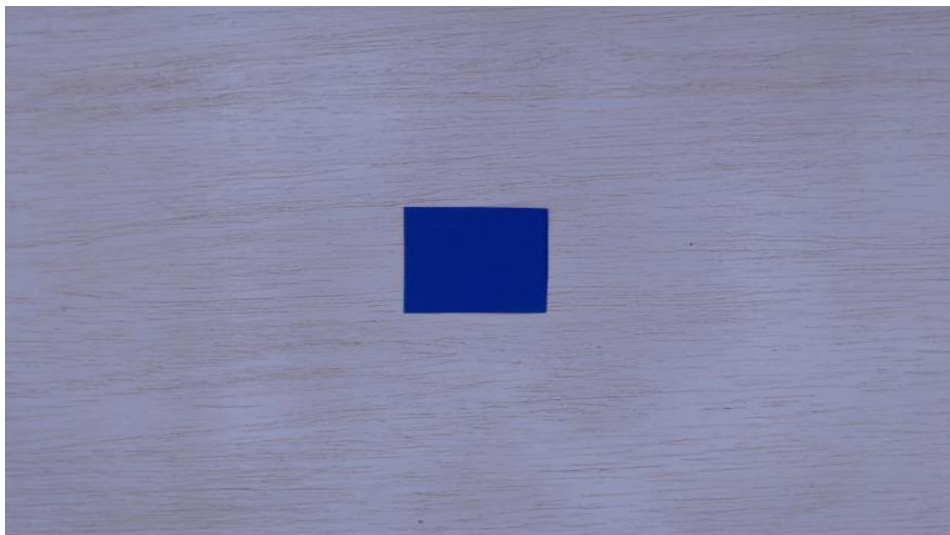
```
def encontrar_contorno(self, imgBinary,):
    (self.contorno, jerarquia) = cv2.findContours(imgBinary, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
    font = cv2.FONT_HERSHEY_SIMPLEX
    for c in self.contorno:
        rect = cv2.minAreaRect(c)
        box = cv2.boxPoints(rect)
        #normalizamos
        box = np.int0(box)
        #hallando alturas
        lado1 = ((box[0][0] - box[1][0])**2 + (box[0][1] - box[1][1])**2)**0.5
        lado2 = ((box[1][0] - box[2][0])**2 + (box[1][1] - box[2][1])**2)**0.5
        lado1 = round(lado1)
        lado2 = round(lado2)
        largo = 0
        ancho = 0
        #largo y el ancho
        if lado2 <= lado1:
            largo = lado1
            ancho = lado2
        else:
            largo = lado2
            ancho = lado1
        #print(lado1, lado2)
    return largo, ancho, box
```

Nota. Fuente: Elaboración propia



Nota. Fuente: Elaboración propia

Figura 33 *Conversión de las dimensiones a cm*



Nota. Fuente: Elaboración propia

El resultado de la extracción de características da como resultado el ancho y largo de la palta, para convertir o generar su equivalencia a cm se realizó el siguiente procedimiento

En el prototipo se tomó un pequeño objeto con dimensiones de 5x5 cm, al extraer los lados de este pequeño objeto se obtuvo una medida en pixeles. Haciendo uso la regla de 3 simples se obtuvo la equivalencia.

Formula

$$X = P / 5 \text{ cm} \quad X = 67 / 5 \text{ cm} > 13.4$$

Es decir, 1cm equivale a 13.4 pixeles

Análisis de color: análisis de color se hecho para detectar anomalías en la superficie de la palta estas anomalías pueden ser como:

Enfermedades, rasgaduras, putrefacción, etc. Que suele cambiar de color al presentar estas anomalías.

Para obtener el porcentaje del área sana de la palta se realizó los siguientes pasos

1. se obtuvo el área del color verde de la región de interés.
2. De la imagen binarizada se obtuve el área total de región de interés.
3. extrajo cociente de área total con respecto el área de color verde
4. finalmente se obtuvo el área de la zona afectada.

Zona Sana=AreaTotal/Are Verde (%)

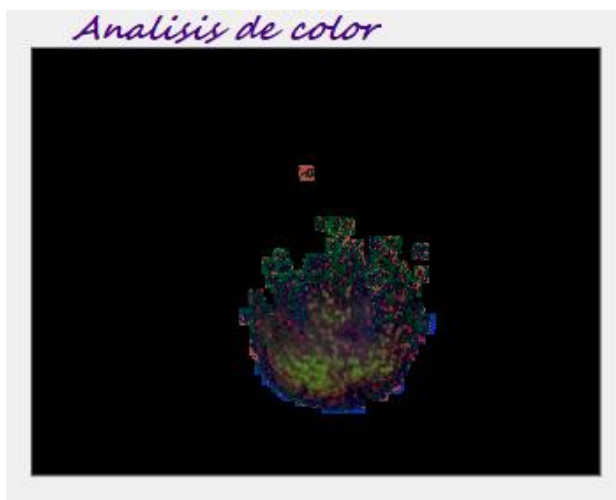
AreaAfectada=1-ZonaSana

Figura 34 *Análisis de color código*

```
116
117 def obtener_masqueraGreen(self,img):
118     imgo=img.copy()
119     #mejora el contraste
120     img_yuv=cv2.cvtColor(img,cv2.COLOR_BGR2YUV)
121     img_yuv[:, :,0]=cv2.equalizeHist(img_yuv[:, :,0])
122     img=cv2.cvtColor(img_yuv,cv2.COLOR_YUV2BGR)
123     adjusted = cv2.convertScaleAbs(img, alpha=1.5, beta=0)
124
125     #cambiar a escala de grises y se suaviza la imagen
126     img1=cv2.cvtColor(adjusted,cv2.COLOR_BGR2GRAY)
127     img_gauss=cv2.GaussianBlur(img1,(5,5),0)
128     #(T,thread)=cv2.threshold(img_gauss,50,255,cv2.THRESH_BINARY_INV)
129
130     (T,thread2)=cv2.threshold(img_gauss,94,255,cv2.THRESH_BINARY_INV)
131     res=cv2.bitwise_and(img,img,mask=thread2)
132     res=cv2.cvtColor(res,cv2.COLOR_BGR2HSV)
133     #verde
134     lower_color=np.array([40,0,0])
135     upper_color=np.array([85,255,255])
136     mask_green=cv2.inRange(res,lower_color,upper_color)
137     kernel=np.ones((5,5),np.uint8)
138     dilatado=cv2.dilate(mask_green,kernel,iterations=2)
139     #erosionado=cv2.erode(dilatado,kernel,iterations=1)
140     area=self.area_verde(dilatado)
141     return dilatado,area
```

Nota. Fuente: Elaboración propia

Figura 35 *Representación del área de la zona verde*



Nota. Fuente: Elaboración propia

Proceso del manejo del software.

Resultado del análisis

La finalidad de realizar la clasificación de los frutos de la Palta de acuerdo al peso y la coloración de la superficie del fruto así de esa manera brindar una herramienta automatizada a los productores de este fruto, permitiendo optimizar recursos (humanos y tiempo). Ya que el software brinda funciones como la clasificación automática (estimación de peso y detección de enfermedades y/o daños), reportes y exportación en archivo pdf.

1. Acceso al sistema.

El sistema ACUATEC es un sistema portable, quiere decir que no requiere una instalación previa, sin embargo, requiere un espacio mínimo de 20 MB en el disco, memoria del equipo de 4 GB como mínimo y finalmente procesador Core i5 o su equivalencia en AMD. Para acceder simplemente se le da doble clic sobre el icono del programa, visto en capturas.

2. Vista de la pantalla principal.

En la pantalla se visualiza 4 menú es superiores categorizado según sus funcionalidades.

- ✓ Configuración.
- ✓ Reportes.

- ✓ Personal.
- ✓ Ayuda.

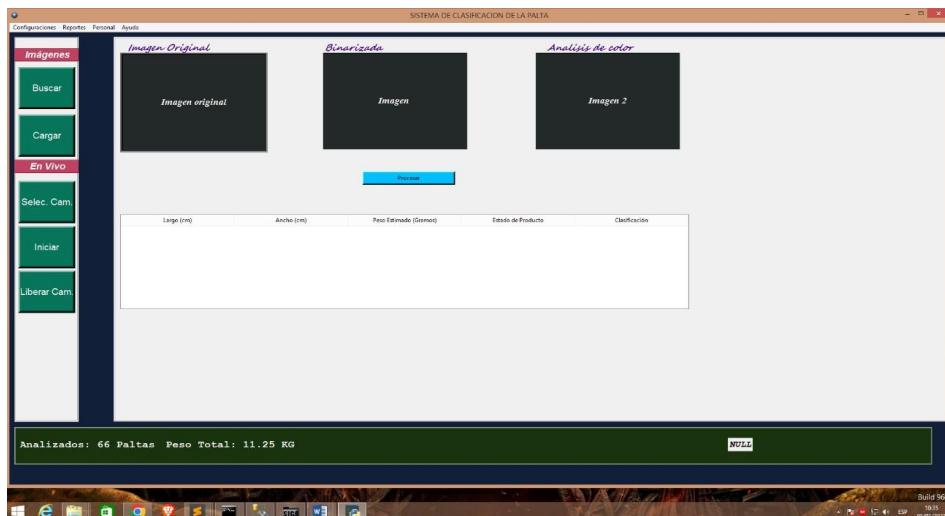
Así como también cuenta con controles laterales, que básicamente son los controles para la manipulación de la imagen y video.

Otro punto importante en la pantalla principal son los pequeños cuadros que se muestran en la pantalla, que básicamente es donde se muestra la imagen o video y la imagen procesada.

Y finalmente en la parte inferior hay un módulo en donde se muestra los datos del usuario y un reporte básico.

Figura 36 *menú principal*

Nota. Fuente: Elaboración propia



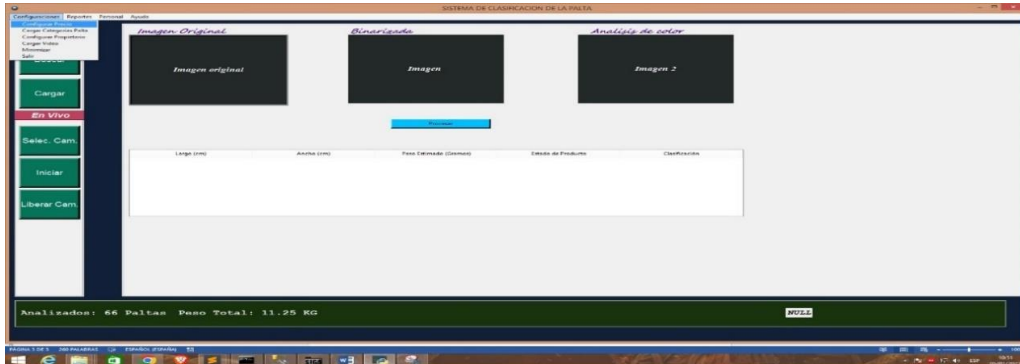
3. Menú configuración.

Este menú cuenta con 6 submenús:

- Configuración precio.
- Cargar categorías palta.
- Configuración propietaria.
- Cargar video.

- Minimizar.
- Salir.

Figura 37 Configuración

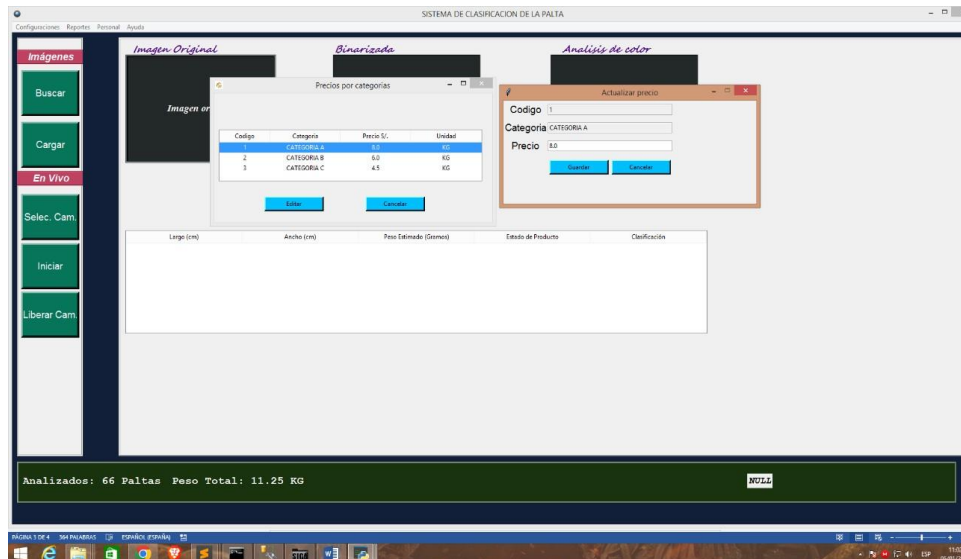


Nota. Fuente: Elaboración propia

3.1. Configuración precio.

En este sub menú se puede realizar la configuración del precio, y se aprecia en lasiguiente captura.

Figura 38 Sub menú



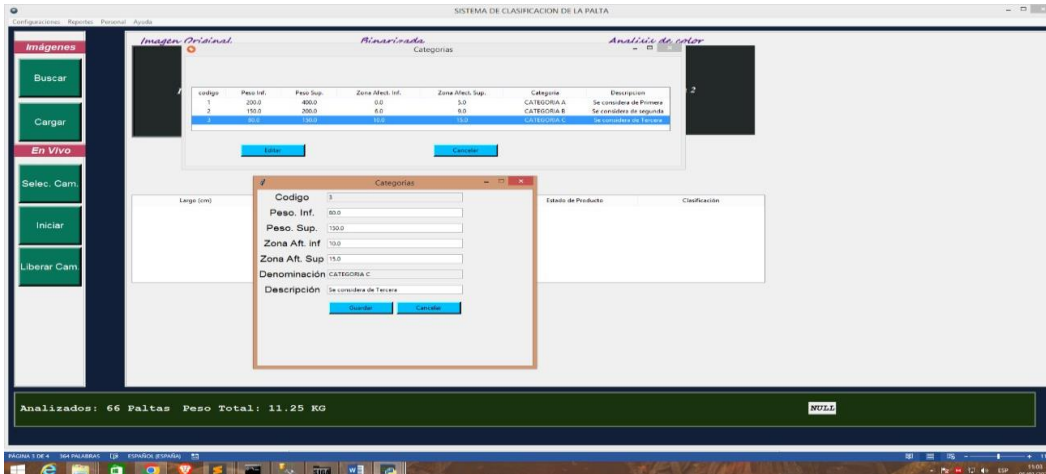
Nota. Fuente: Elaboración propia

3.2. Cargar categorías palta.

En esta categoría es en donde se modifica los pesos mínimos y máximos por categoría de la palta, así como también la superficie de la palta que

representa con algún daño o enfermedad para ser aceptada en las 3 categorías. (Se aprecia en la siguiente captura).

Figura 39 Categorización

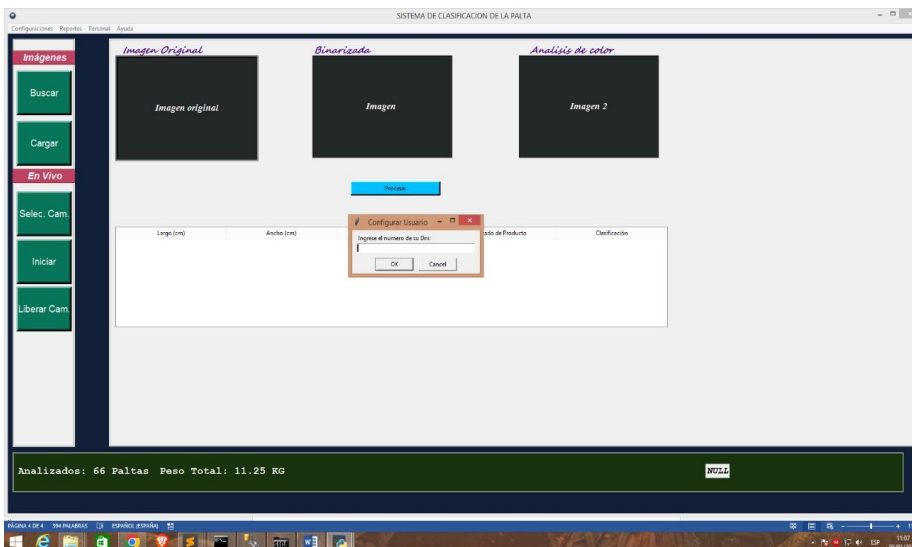


Nota. Fuente: Elaboración propia

3.3. Configuración propietaria.

Este menú es muy importante, ya que para realizar el procesamiento se debe ingresar un usuario, el cual representa el dueño o propietario del producto analizado. (Véase la siguiente imagen)

Figura 40 procesamiento



Nota. Fuente: Elaboración propia

3.4. Cargar video.

También se puede realizar el análisis en un video grabado, precisamente el presente menú permite cargar dicho contenido, para su posterior análisis.

3.5. Minimizar.

Minimiza el sistema.

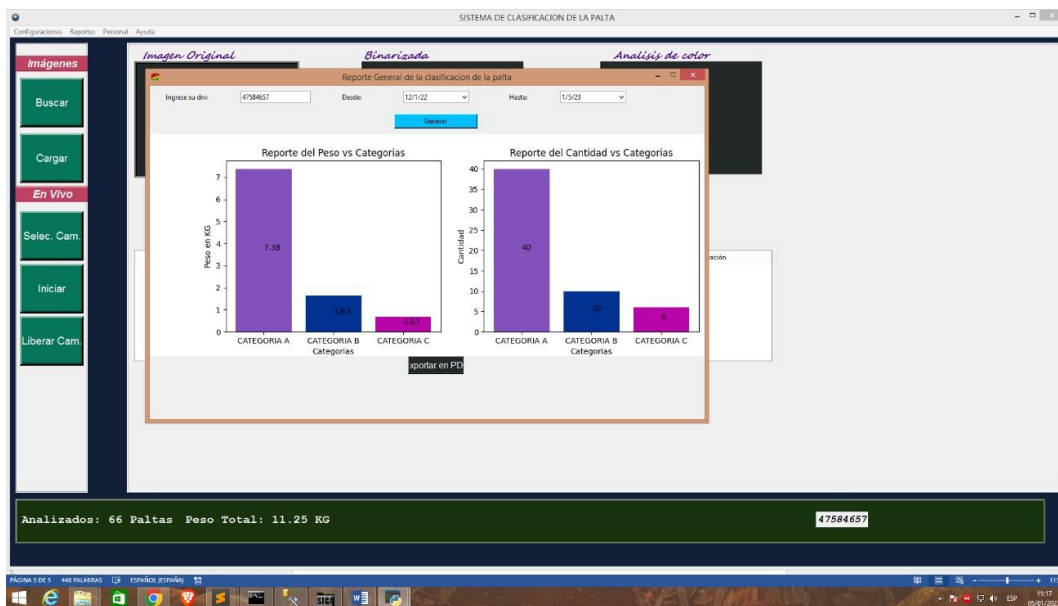
3.6. Cerrar.

Cierra el sistema.

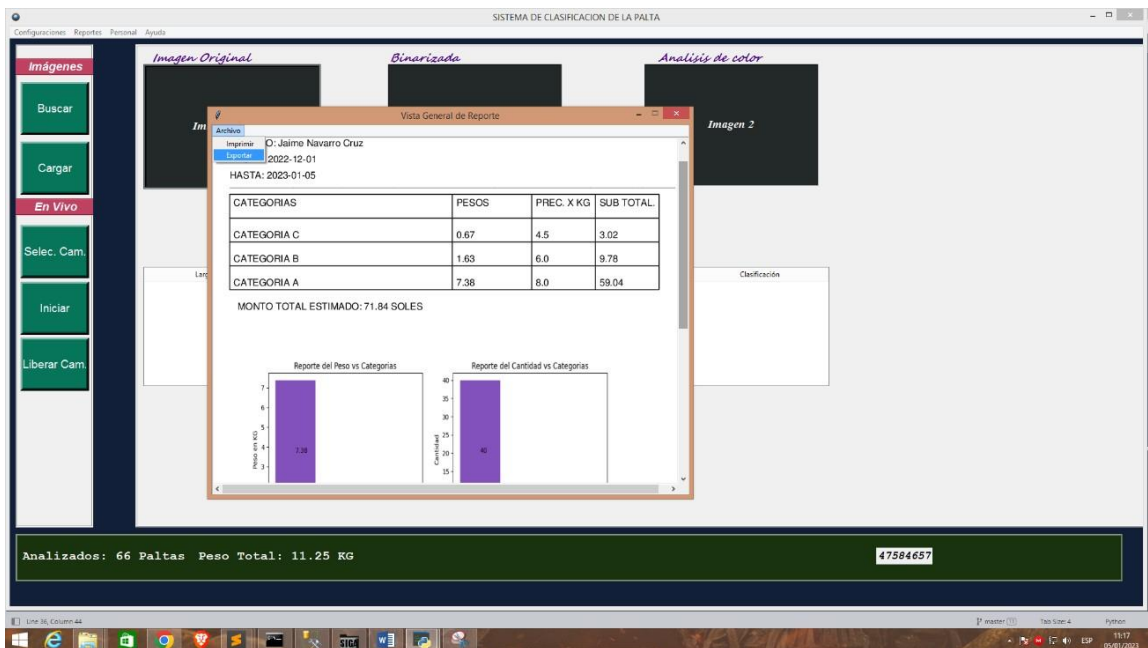
4. Menú reporte.

En este menú o apartado extrae los datos estadísticos, así como también cuenta con una opción de vaciar la data (reiniciar).

Figura 41 *datos estadísticos*



Nota. Fuente: Elaboración propia

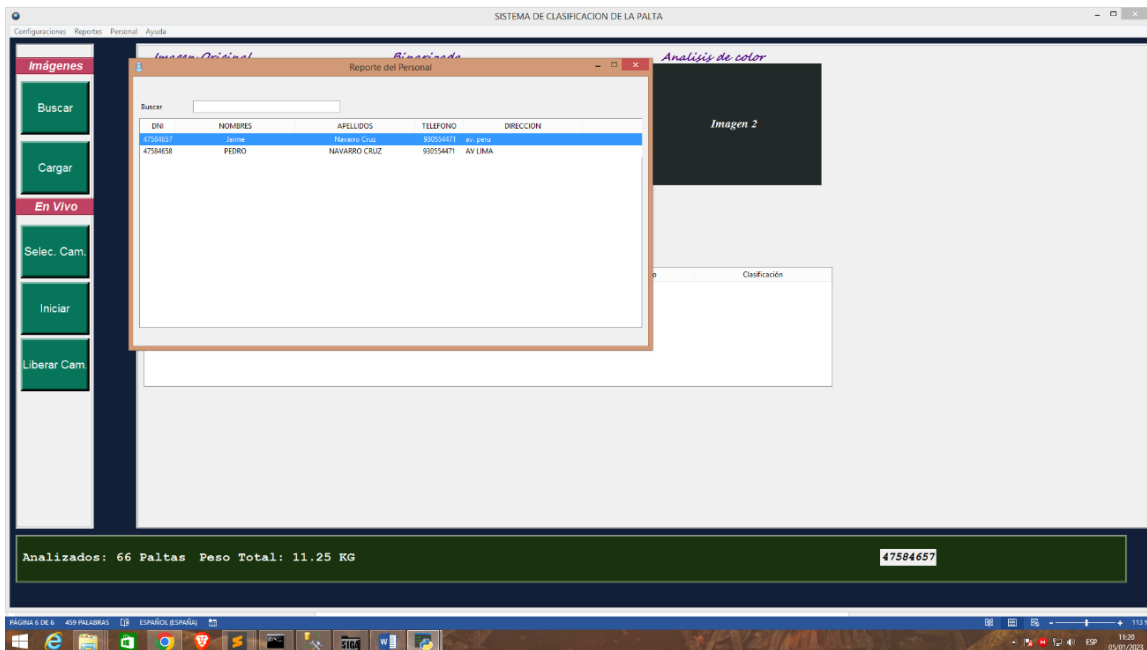


Nota. Fuente: Elaboración propia

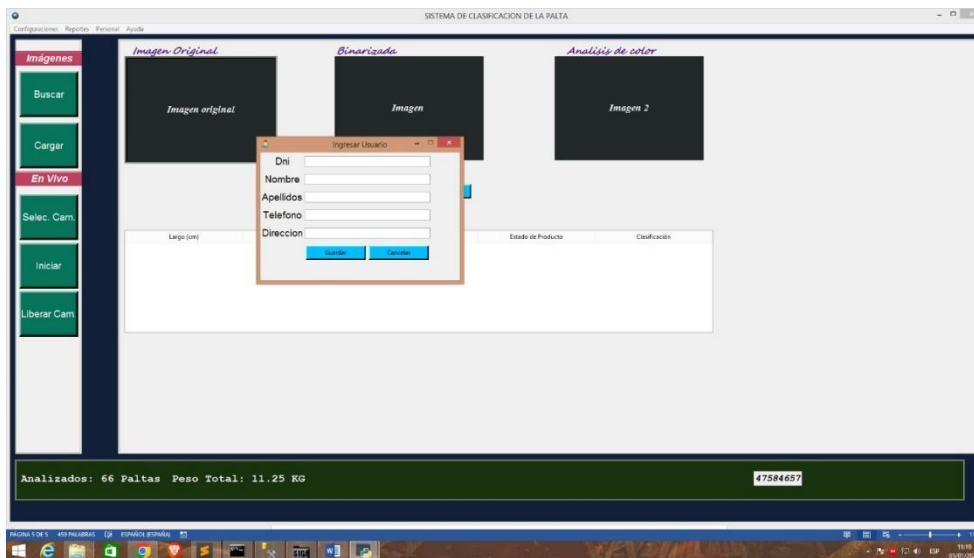
5. Personal

Se hace inserción de un nuevo usuario.

Figura 42 *Nuevo usuario.*



Nota. Fuente: Elaboración propia



Nota. Fuente: Elaboración propia

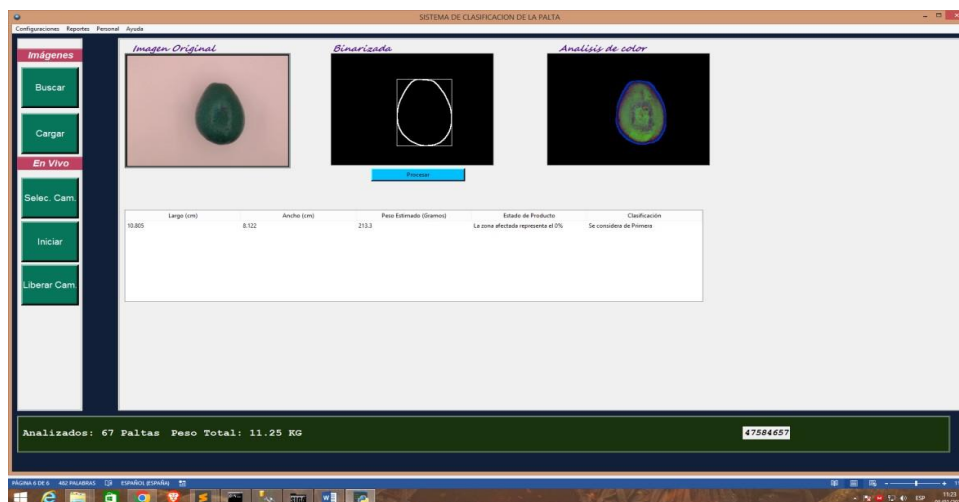
6. Ayuda

Y finalmente tenemos el menú ayuda, que brinda la información del desarrollador y las librerías usadas para el desarrollo.

7. Sobre el Procesamiento.

En este apartado es en donde se carga el video (tiempo real) o imagen con los controles que figuran en la barra lateral izquierdo, una vez realizado la carga se presiona en el botón procesar, y obtenemos el resultado deseado.

Figura 43 *Procesamiento*



Nota. Fuente: Elaboración propia

CAPITULO VI RESULTADOS

6.1 Entrenamiento del modelo de regresión lineal para estimación de peso

Modelo entrenado es modelo estadístico de regresión lineal con dos variables, un variable independiente=el área en cm cuadrados de la superficie de la palta y variable dependiente= el peso de la palta en gramos, para entrenar el modelo se utilizó la plataforma de **google colab**, así como también la librería de scikit learn que viene a ser una plataforma libre acceso.

Para el entrenamiento del modelo se trabajó con una muestra de 114, las cuales se etiquetaron de forma manual

Tabla 5 *Muestras*

Nro.	Área	Peso	Nro.	Área	Peso	Nro.	Área	Peso	Nro.	Área	Peso
1	54.4	146	31	32.83	98	61	97.2	234	91	48.38	160
2	39.75	120	32	41.04	140	62	65.72	170	92	35.77	120
3	68	197	33	57	150	63	96.39	230	93	101.15	260
4	51.92	163	34	52.8	150	64	65.1	167	94	63.36	159
5	50.4	149	35	52.29	165	65	51.6	147	95	75.97	187
6	51.92	153	36	37.26	95	66	112.05	330	96	56.76	150
7	39.78	120	37	55.76	148	67	94.6	231	97	78	183
8	52.8	160	38	51.75	159	68	76	180	98	79.8	184
9	48.6	155	39	33.39	94	69	75.05	170	99	65.28	156
10	52.29	160	40	82.5	230	70	77	184	100	81.36	195
11	40.56	140	41	62.3	176	71	50.16	159	101	78.44	184
12	57.8	168	42	61	169	72	45.82	147	102	58.65	175
13	58.5	156	43	63	175	73	46.75	151	103	89.38	210
14	49.77	173	44	64	174	74	59	160	104	77.52	170
15	82.95	200	45	69.58	175	75	56.05	166	105	76.32	172
16	37.5	115	46	81	250	76	87.5	259	106	77.7	175
17	53.4	160	47	58.28	154	77	63	180	107	82	205
18	38.5	134	48	67.2	170	78	63.65	169	108	78.75	184
19	47.4	160	49	83.74	200	79	48	160	109	80.56	185

20	54.81	143	50	101.7	320	80	50.02	142	110	69.3	182
21	41.25	140	51	81	220	81	39.6	136	111	63.36	159
22	52	158	52	102	250	82	44.82	146	112	62.08	162
23	48	160	53	75.26	185	83	65.92	179	113	66.66	172
24	49.8	135	54	90.06	230	84	56.32	155	114	83.2	200
25	44.25	139	55	78	185	85	65.55	170			
26	41.25	135	56	61	179	86	46.8	149			
27	33.32	100	57	69	187	87	48.4	130			
28	49.41	160	58	99.6	297	88	51.66	157			
29	40.5	140	59	65	168	89	61.74	158			
30	41.44	140	60	94.4	230	90	45.82	137			

Nota. Fuente: Elaboración propia

Modelo de regresión lineal y conclusiones derivadas sean válidas los datos y/o variable se deben cumplir ciertas condiciones:

1. relación lineal entre las variables respuesta o dependiente y la variable independiente
2. distribución normal de variable respuesta.

PRUEBA DE NORMALIDAD

Figura 44 Carga de datos (pesos de palta)

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import style
import seaborn as sns
import statsmodels.api as sm
from sklearn import stats
from google.colab import drive
drive.mount('/content/drive')

plt.style.use('ggplot')

import warnings
warnings.filterwarnings('ignore')

url = ('/content/drive/MyDrive/kagle/originalPP.csv')
datos = pd.read_csv(url, sep=';')
datos.info()

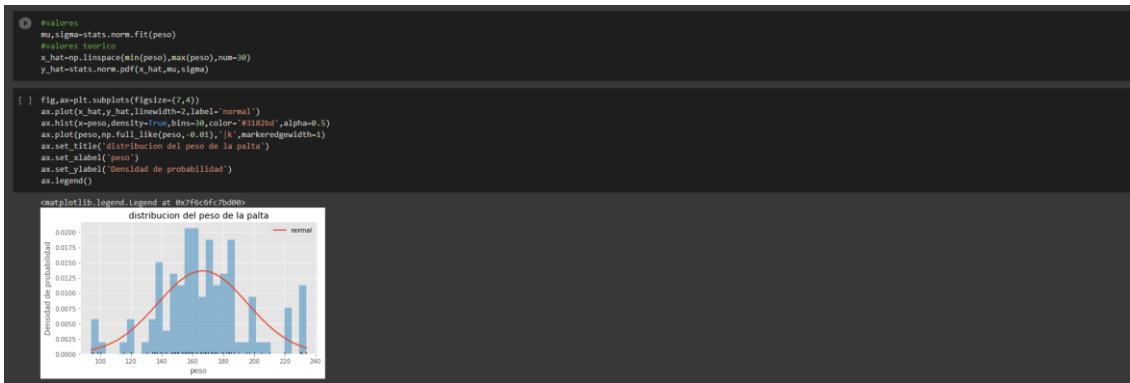
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 132 entries, 0 to 131
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  -
 0   area    132 non-null     float64
 1   peso    132 non-null     int64
dtypes: float64(1), int64(1)
memory usage: 2.2 KB

selection = datos[datos.peso > 250]
peso = selection['peso']

```

Nota. Fuente: Elaboración propia

Figura 45 *Grafica de la distribución de datos y su densidad de probabilidad*



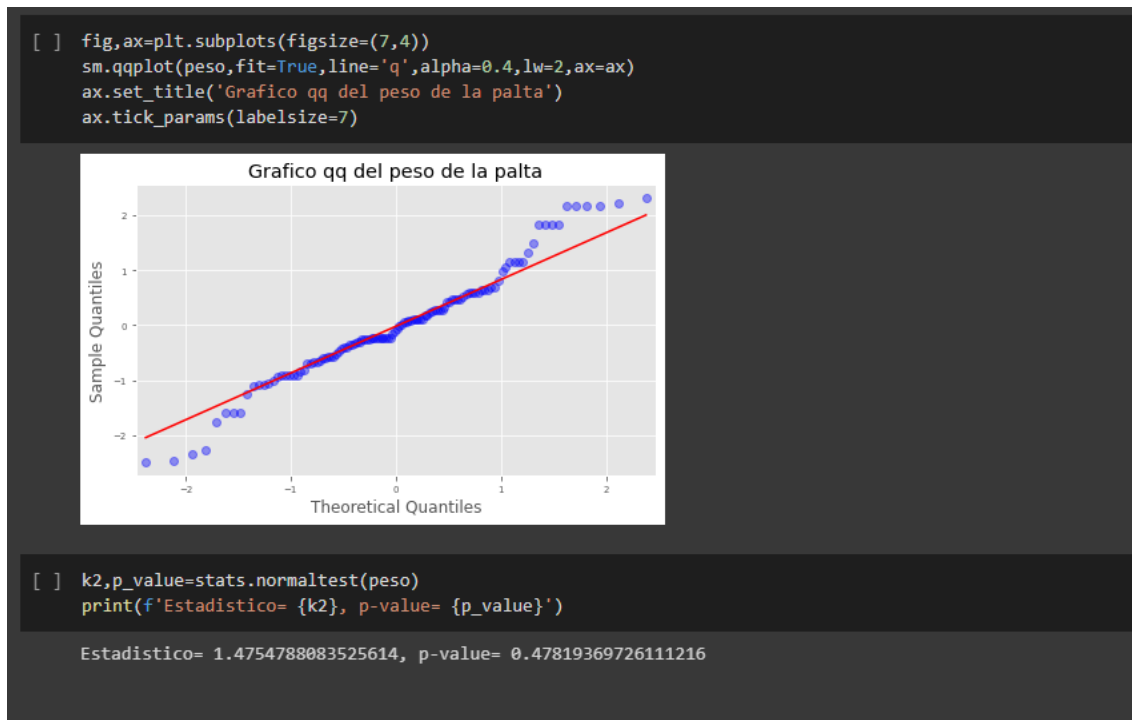
Nota. Fuente: Elaboración propia

Tabla 6 *Prueba de normalidad a los datos procesados*

	AREA	PESO
N	114	114
Parámetros normales^{a,b}	Media	171.44
	Desviación estándar	40.177
Máximas diferencias extremas	Absoluta	.166
	Positivo	.093
	Negativo	-.097
Estadístico de prueba	.093	.166
Sig. asintótica (bilateral)	,017 ^c	,000 ^c

Nota. Fuente: Elaboración propia

Figura 46 prueba de D' de agostino



Nota. Fuente: Elaboración propia

Se utilizó prueba D' Agostino en donde nos indica que:

H0: datos tienen distribución normal.

H1: datos no tienen distribución normal.

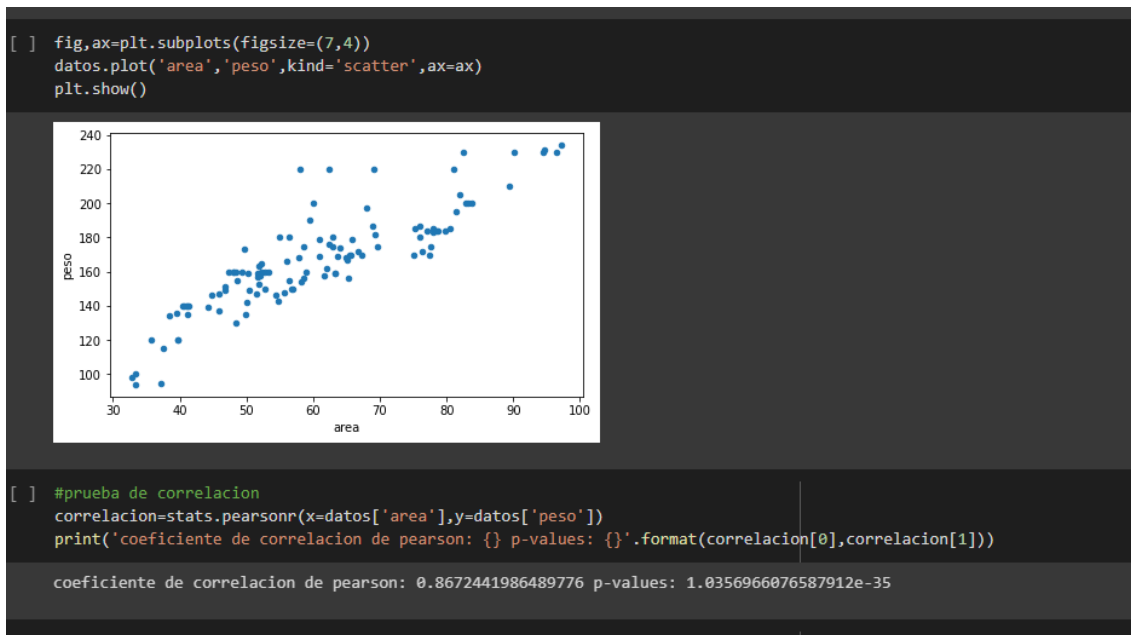
En donde p-valor mayor 0.05, acepta hipótesis nula, datos tienen distribución normal.

Prueba de correlación de las variables

Se verifica que tanto se relacionan las variables, para ello se utilizó prueba correlación

Pearson.

Figura 47 Prueba de correlación de



Nota. Fuente: Elaboración propia

Se ve en figura anterior tiene una correlación 86% quiere decir tiene una relación fuerte.

Una vez realizado las pruebas necesarias se ajusta modelo regresión lineal

Paso1: separa la data en dos partes, una de entrenamiento y la otra de prueba, si bien la proporción se puede tomar de manera arbitraria, en la mayoría de los modelos se suele tomar 80-20 es decir 80% de los datos para ajustar el modelo y el 20% restante para realizar el test.

Figura 48 Realización del test

```
[ ] X=datos['area']
    Y=datos['peso']
    x_train,x_test,y_train,y_test=train_test_split(X.values.reshape(-1,1),Y.values.reshape(-1,1),train_size=0.8,random_state=1234,shuffle=True)
```

Nota. Fuente: Elaboración propia

Luego realiza el entrenamiento del modelo:

Figura 49 *Entrenamiento del modelo*

```
#creamos el modelo y entrenamos
modelo=linear_model.LinearRegression()
modelo.fit(X=x_train.reshape(-1,1),y=y_train)

LinearRegression()

[ ] print(f'intercepto: {modelo.intercept_}   ::: coeficiente : {modelo.coef_}')

intercepto: [64.06930545]   ::: coeficiente : [[1.70041911]]
```

Nota. Fuente: Elaboración propia

Se ve en imagen anterior nos arroja un intercepto de 64.06 y un coeficiente de 1.7.

Finalmente, nuestra recta de regresión vendría a ser:

$$\text{Peso}=64.06+\text{area}*1.7$$

Entrenado modelo, evalúa capacidad predictiva usando conjunto de test.

Figura 50 *Capacidad predictiva*

```
[ ] #precision del modelo
rmse=mean_squared_error(y_true=y_test,y_pred=predicciones,squared=False)
precision=modelo.score(x_train,y_train)
#el error del test es
print(f'error: {rmse}')
print(f'la precision es de: {precision}')

error: 13.991228465027108
la precision es de: 0.7706227067565214
```

Nota. Fuente: Elaboración propia

Como se puede visualizar nos arroja un error de 13, quiere decir que el valor real de las predicciones puede oscilar en:

$$\text{Peso}=64.06+\text{area}*1.7\pm 13$$

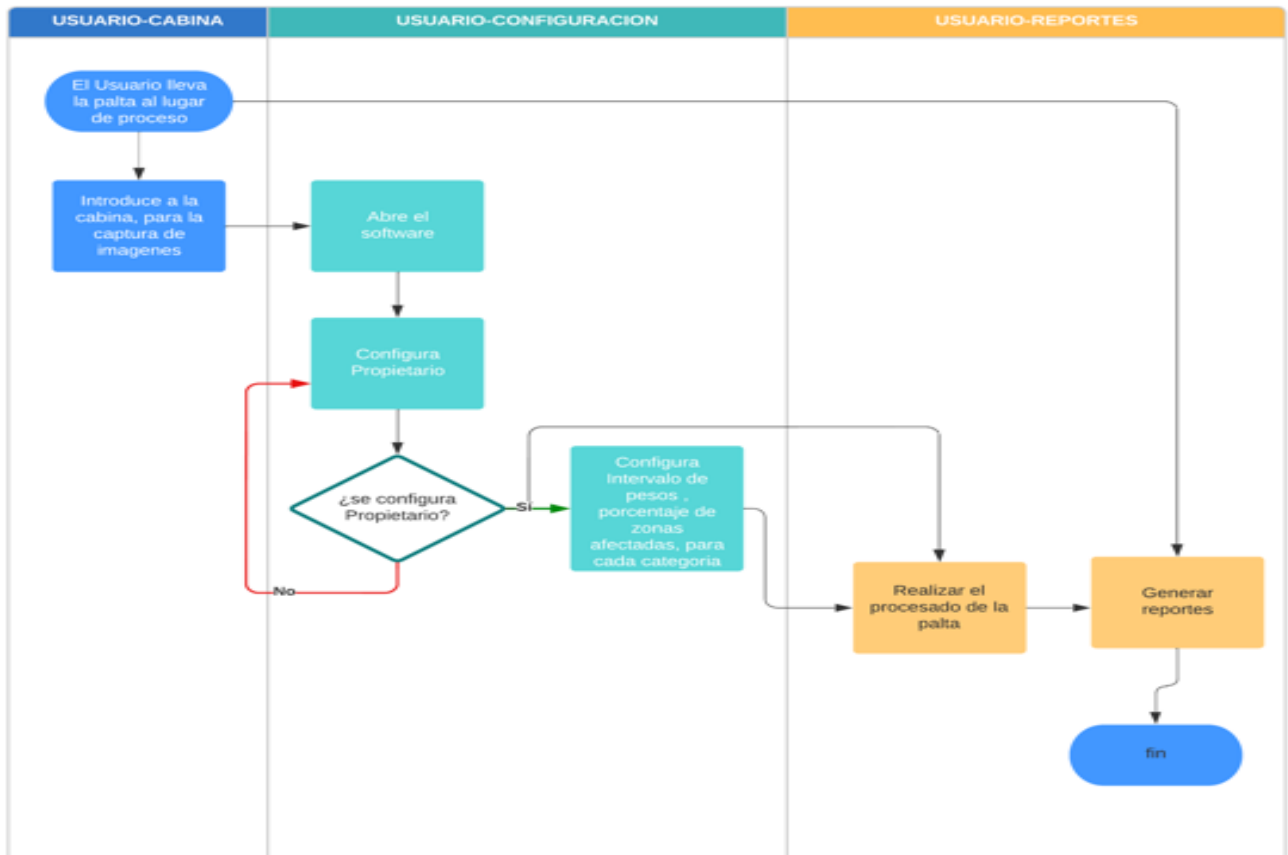
Finalmente modelo generado se guardó en archivo externo con la extensión *.joblib esto con la finalidad de incrustar el modelo en el software, para las predicciones de los pesos de la palta.

6.2 Objetivos

El software cumple la función de realizar el procesamiento y posterior análisis para tomar decisiones. Para ello toma como recurso principal la imagen capturada del prototipo diseñado, posteriormente con un simple click en un botón realiza el procesamiento y análisis correspondiente.

Figura 51 Flujo de procesamiento de palta

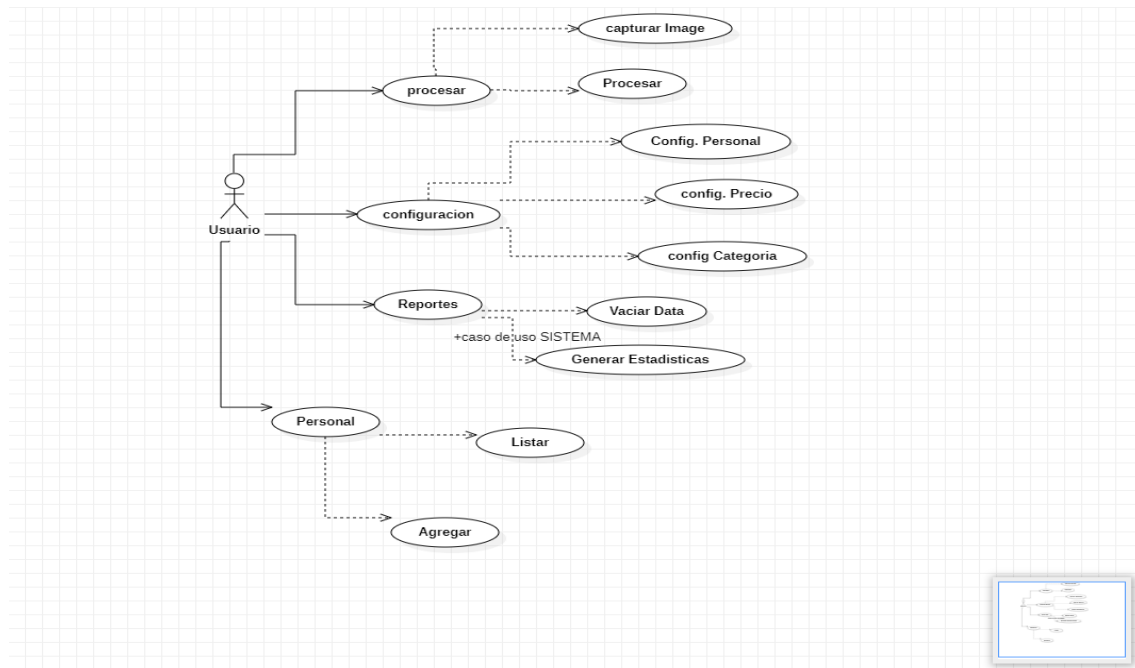
Diagrama de flujo de proceso:



Nota. Fuente: Elaboración propia

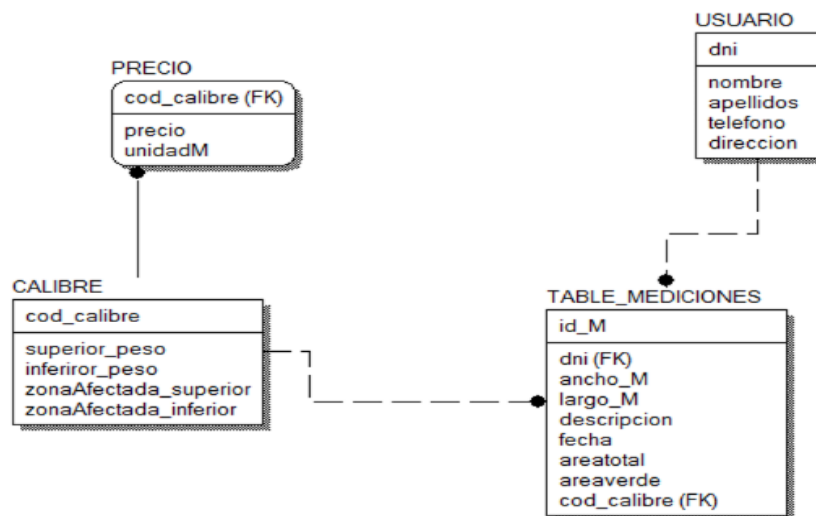
A continuación, el diagrama de caso de uso.

Figura 52 Diagrama caso de uso



Nota. Fuente: Elaboración propia

Figura 53 Diagrama Entidad – Relación de la base de datos



Nota. Fuente: Elaboración propia

CONCLUSIONES

- Se pudo diseñar sistema automático de bajo costo según necesidad de productores de palta de la comunidad de Mollebamba y se podría adecuar también para PyMe's que clasifica paltas Hass en 4 calidades según calibre y calidad de piel.
- Se hizo prototipo de sistema automático de selección paltas Hass por índices de calidad rozamiento, quemadura solar o frío, negras o lenticelas.
- Se definió sistema de clasificación según visión de computadora que disgrega producto final según categorías preestablecidas, para reducir pérdidas e incrementar productividad de proceso.
- Por dificultad de identificación de imágenes basado en contenido, resultados de este estudio son aceptables porque identifica imágenes en buenas condiciones. Demostró que infraestructura hecha es adecuada para propósito del estudio donde logro objetivos buscados.
- Se debe estudiar más a fondo clasificación e identificación de cada tipo de imagen, ejemplo, agregar tipos de momentos invariantes como Flusser u otra característica que ayude ayudar clasificación y mejorar nivel de acierto.

RECOMENDACIONES

- El estudio diseño sistema de reconocimiento de patrones basado en visión de computadora
- Luz tiene rol relevante al evaluar imágenes, por ende mejor usar cámaras infrarrojas o que ambiente este iluminado para identificación.
- Uso de métodos de procesamiento de imágenes basados en las librerías de clasificación como OPENCV y SCIKIT-LEARN, relativo a aplicación específica; hay mucha evidencia para dar juicio y decir que proceso de reconocimiento automático es mejor a otros.
- Recomienda hacer más estudios donde se aplique métodos computacionales de áreas de inteligencia artificial: machine learning, data mining, etc, y llevarlo al campo de agricultura.
- Aborde alternativas novedosas para implementar algoritmos aprovechando multiprocesamiento paralelo de computadoras modernas.

REFERENCIAS BIBLIOGRÁFICAS

- A.W.S. (2023). *¿Qué es Python?* Pagina web: <https://aws.amazon.com/es/what-is/python/>.
- Agencia Peruana de Noticias. (2022). *Exportaciones de palta Hass crecerán 8% este año*. Lima: <https://andina.pe/agencia/noticia-exportaciones-palta-hass-creceran-8-este-ano-906120.aspx>.
- Agil Manifiesto. (2001). *Principios del Manifiesto Ágil*. Obtenido de <https://agilemanifesto.org/iso/es/manifiesto.html>
- Aguilar, J., & Castaño, V. (2016). *Diseño e implementación de un sistema para el reconocimiento de patrones Utilizando Videocamaras*. Pereira: <https://repositorio.utp.edu.co/server/api/core/bitstreams/54ccb194-1a04-4acc-8870-39c9662317a3/content>.
- Amat, J. (2021). *Análisis de normalidad con python*. [cienciadedatos.net](https://www.cienciadedatos.net): <https://www.cienciadedatos.net/documentos/pystats06-analisis-normalidad-python.html>.
- Armando, C. (2019). *La situación del mercado internacional de la palta*. Lima: Ministerio de Agricultura y Minas. Obtenido de <https://bibliotecavirtual.midagri.gob.pe/index.php/analisis-economicos/estudios/2019/28-la-situacion-del-mercado-internacional-de-la-palta/file>
- Ataucusi, E. (2021). *Influencia de un sistema con reconocimiento facial y medición de temperatura en el control de acceso para participantes del programa trabaja Peru en el distrito de Talavera*. Apurímac: https://repositorio.unajma.edu.pe/bitstream/handle/20.500.14168/666/Edison_Tesis_Bachiller_2021.pdf?sequence=1&isAllowed=y.
- Basantes, N. (2018). *Desarrollo de una plataforma tecnológica para la gestión de seguridad en una institución educativa de grado inicial mediante el uso de sistemas móviles, reconocimiento facial y sistemas de alertas*. Guayaquil: <http://repositorio.ug.edu.ec/bitstream/redug/32717/1/B-CISC-PTG-1572%20Basantes%20Verdugo%20Nathalia%20Lilibeth.pdf>.
- Bermeo, E., & Zorrilla, P. (2019). *Desarrollo de un sistema automático de selección de paltas Hass por sus índices de calidad para la empresa Agroindustrias Verdeflor S.A.C*. Lima: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/629958/Bermeo_VE.pdf

- ?sequence=3.
- Bermeo, V. (2019). *Desarrollo de un sistema automático de selección de paltas Hass por sus índices de calidad para la empresa Agroindustrias Verdeflor S.A.C.* Lima: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/629958/Bermeo_VE.pdf?sequence=3.
- Bishop, C. M. (2007). *Pattern Recognition and Machine Learning*.
- Caceres, E. (2018). *Aplicacion Movil de reconocimiento facial en personas con antecedentes de abuso sexual en la provincia de Andahuaylas, Apurimac- 2018.* Andahuaylas: https://repositorio.unajma.edu.pe/bitstream/handle/20.500.14168/358/Ervin_Lewis_Tesis_Bachiller_2018.pdf?sequence=1&isAllowed=y.
- Espinoza, G. (2015). *Sistema de reconocimiento de patrones en placas vehiculares para el acceso automático de visitas a un edificio.* Lima: https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/5753/ESPINOZA_GERARDO_PLACAS_VEHICULARES_AUTOMATICO_VISITAS_EDIFICIO.pdf?sequence=1&isAllowed=y.
- Filein, R. (2016). *SQLite: La Base de Datos Embebida.* SG #17: <https://sg.com.mx/revista/17/sqlite-la-base-datos-embebida>.
- Gonzales & Woods. (2008). *Digital Image Processing*.
- Hernandez, L. F. (2009). *Segmentacion binaria del video de una webcam en tiempo real.*
- Herrera, J., & Medina, S. (2015). *Diseño de un sistema Automatico de seleccion de Frutos de cafe mediante Tecnicas de Vision Artificial.* Barranaquilla, Colombia: <http://repositorio.uac.edu.co/bitstream/handle/11619/1656/TMCT%200022C.pdf?sequence=1&isAllowed=y>.
- INEI. (Octubre de 2020). *Instituto Nacional de Estadística e Informática.*
- MINAGRI. (17 de Marzo de 2021). *Ministerio de Desarrollo Agrario y Riego.* Obtenido de <https://www.gob.pe/institucion/midagri/noticias/348556-midagri-ultima-campana-de-exportacion-de-palta-peruana-supero-las-400-mil-toneladas>
- Montoya, C., Cortes, J., & Chaves, J. (2014). *Sistema automático de reconocimiento de frutas basado.* Chile: <https://www.scielo.cl/pdf/ingeniare/v22n4/art06.pdf>.
- NTP. (2014). *Norma Técnica de la Palta.* Lima.
- Ortiz, A. (2010). *Python como primer lenguaje de programación.* Mexico:

- https://arielortiz.info/publicaciones/primer_lenguaje_30_jun_2010.pdf.
- Palomino, C. J. (2014). *Software automático de reconocimiento de patrones para la discriminación de granos de arroz pulido basado en procesamiento digital de imágenes*. Universidad Nacional Jose Maria Arguedas, Andahuaylas.
- Palomino, J. (2015). *Software automatico de reconocimiento de patrones para la discriminacion de granos de arroz pulido basado en procesamiento digital de imagenes*. Peru: <https://repositorio.unajma.edu.pe/bitstream/handle/20.500.14168/181/04-2014-EPIS-Palomino%20Caceres-Software%20automatico%20de%20reconocimiento%20de%20patrones.pdf?sequence=1&isAllowed=y>.
- Peru, S. F. (Diciembre de 2021). *Super Food Peru*. Obtenido de <https://peru.info/es-pe/superfoods/detalle/super-palta#:~:text=La%20palta%20tiene%20potasio%20y,dar%20soporte%20al%20sistema%20inmunol%C3%B3gico>
- Platzi. (2015). *Platzi*. Obtenido de <https://platzi.com/blog/metodologia-scrum-fases/>
- Portafolio. (2018). *Cultivos del futuro, con los pies sobre la*. <https://www.portafolio.co/economia/cultivos-del-futuro-con-los-pies-sobre-la-tierra-521749>.
- Rivas, D. (2020). *Diseño conceptual de un sistema automatico para la clasificacion de paltas Hass*. Lima: https://tesis.pucp.edu.pe/repositorio/bitstream/handle/20.500.12404/18521/RIVAS_ROMERO_DEGLAN_DISE%c3%91O_CONCEPTUAL_SISTEMA.pdf?sequence=1&isAllowed=y.
- Rodriguez, D. (2016). *Diseño e implementacion de una herramienta computacional para el analisis asistido de motilidad espermatica en bovinos mediante Tecnicas de Vision Artificial*. Colombia: <https://repositorio.ucundinamarca.edu.co/bitstream/handle/20.500.12558/348/DISE%c3%91O%20E%20IMPLEMENTACION%20DE%20UNA%20HERRAMIENTA%20COMPUTACIONAL%20PARA%20EL%20ANALISIS%20ASISTIDO%20DE%20MOTILIDAD%20ESPERMATICA%20EN%20BOVINOS%20MEDIANTE%20TECNICAS%20DE>

- Sanchez, H. (2018). *Diseño e implementacion de un controlador basado en reconocimiento de patrones para mejorar el abastecimiento de insumos alimentarios en galpones avicolas*. Ecuador: <http://dspace.esPOCH.edu.ec/bitstream/123456789/9233/1/20T01097.pdf>.
- Santa maria, J., Rios, C., Rosdriguez, C., & Garcia, C. (2021). *Reconocimiento de patrones de imágenes a través de un sistema de visión artificial en MATLAB*. Lima: <http://portal.amelica.org/ameli/journal/535/5352640003/html/>.
- Theodoridis & Koutroumbas. (1999). *Pattern Recognition*.

ANEXO

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLE
Problema General	Objetivo General	Hipótesis General	V. Independiente
¿Cuáles es el diseño de sistema automático de reconocimiento de patrones para clasificación de la palta en la comunidad Mollebamba- Chincheros?	Diseñar sistema automático de reconocimiento de patrones para clasificación de la palta en la comunidad de Mollebamba- Chincheros.	El sistema automático de reconocimiento de patrones optimiza clasificación de palta de la comunidad de Mollebamba – Chincheros	Sistema automático de reconocimiento de patrones
Problemas Específicos	Objetivos Específicos	Hipótesis Específicas	V. Dependiente
<p>1. ¿Cuál es el algoritmo aplicando librería OPENCV en lenguaje de programación Python para estimar las dimensiones de la palta?</p> <p>2. ¿Cuál el algoritmo aplicando librería OPENCV y SCIKIT-LEARN en lenguaje de programación Python para estimar el peso de la palta?</p> <p>3. ¿Cuál el algoritmo aplicando librería OPENCV en lenguaje de programación Python para estimar el tamaño de la palta?</p> <p>4. ¿Cuál el algoritmo aplicando librería OPENCV en lenguaje de programación Python para estimar el porcentaje de enfermedad o daño de la palta?</p> <p>5. ¿De qué manera sistema automático de reconocimiento de patrones para clasificación de la palta estima el precio?</p>	<p>1.Diseñar el algoritmo aplicando la librería OPENCV en lenguaje de programación Python para estimar dimensiones de la palta.</p> <p>2.Diseñar el algoritmo aplicando librería OPENCV y SCIKIT-LEARN en lenguaje de programación Python para estimar el peso de la palta.</p> <p>3.Diseñar el algoritmo aplicando librería OPENCV en lenguaje de programación Python para estimar el tamaño de la palta.</p> <p>4.Diseñar el algoritmo aplicando librería OPENCV en lenguaje de programación Python para estimar el porcentaje de enfermedad o daño de la palta</p> <p>5.Estimar el precio mediante sistema automático de reconocimiento de patrones para clasificación de la palta</p>	<p>1.El algoritmo aplicando la librería OPENCV en lenguaje de programación Python automatiza la estimación dimensiones de la palta</p> <p>2. El algoritmo aplicando la librería OPENCV y SCIKIT-LEARN en lenguaje de programación Python automatiza estimación el peso de palta</p> <p>3.El algoritmo aplicando la librería OPENCV en lenguaje de programación Python automatiza estimación el tamaño de palta.</p> <p>4.El algoritmo aplicando librería OPENCV en lenguaje de programación Python automatiza la estimación del porcentaje de enfermedad o daño de la palta</p> <p>5. Automatizar la estimación del precio mediante el sistema automático de reconocimiento de patrones para clasificación de palta</p>	Clasificación de la palta

