



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TITULO DEL TRABAJO DE TITULACIÓN

**Desarrollo de una interfaz gráfica para la detección de anomalías
en el plasma usando visión artificial y Random Forest**

AUTOR

Malavé Ramírez, Cindy Ivonne

EXAMEN COMPLEXIVO

**Previo a la obtención del grado académico en
INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN**

TUTOR

Fís. Castellanos Omar, PhD.

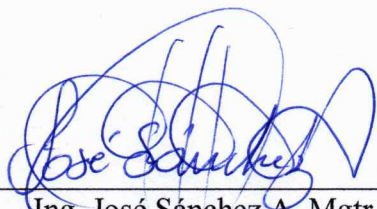
Santa Elena, Ecuador

Año 2026



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

TRIBUNAL DE SUSTENTACIÓN



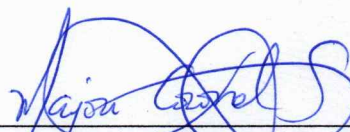
Ing. José Sánchez A. Mgtr.
DIRECTOR DE LA CARRERA



Fís. Omar Castellanos B. PhD.
TUTOR



Ing. Jaime Orozco I. Mgtr.
DOCENTE ESPECIALISTA



Ing. Marjorie Coronel S. Mgtr.
DOCENTE GUÍA UIC



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

CERTIFICACIÓN

Certifico que luego de haber dirigido científica y técnicamente el desarrollo y estructura final del trabajo, este cumple y se ajusta a los estándares académicos, razón por el cual apruebo en todas sus partes el presente trabajo de titulación que fue realizado en su totalidad por CINDY IVONNE MALAVÉ RAMÍREZ, como requerimiento para la obtención del título de Ingeniero en Tecnologías de la Información.

La Libertad, a 28 días del mes julio del año 2026

TUTOR

Fís. Omar Castellanos, PhD.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, CINDY IVONNE MALAVÉ RAMÍREZ

DECLARO QUE:

El trabajo de Titulación *Desarrollo de una interfaz gráfica para la detección de anomalías en el plasma usando visión artificial y Random Forest* previo a la obtención del título en Ingeniero en Tecnologías de la Información, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

La Libertad, a 28 días del mes julio del año 2026

EL AUTOR

Cindy Malavé Ramírez



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES
CERTIFICACIÓN DE ANTIPLAGIO**

Certifico que después de revisar el documento final del trabajo de titulación denominado Examen Complexivo, presentado por el estudiante, CINDY IVONNE MALAVÉ RAMÍREZ fue enviado al Sistema Antiplagio, presentando un porcentaje de similitud correspondiente al 3%, por lo que se aprueba el trabajo para que continúe con el proceso de titulación.



TUTOR

Fís. Omar Castellanos, PhD.



**UNIVERSIDAD ESTATAL PENÍNSULA
DE SANTA ELENA
FACULTAD DE SISTEMAS Y TELECOMUNICACIONES**

AUTORIZACIÓN

Yo, CINDY IVONNE MALAVÉ RAMÍREZ

Autorizo a la Universidad Estatal Península de Santa Elena, para que haga de este trabajo de titulación o parte de él, un documento disponible para su lectura consulta y procesos de investigación, según las normas de la Institución.

Cedo los derechos en línea patrimoniales del presente trabajo de titulación con fines de difusión pública, además apruebo la reproducción dentro de las regulaciones de la Universidad, siempre y cuando esta reproducción no suponga una ganancia económica y se realice respetando mis derechos de autor.

La Libertad, a 28 días de mes julio del año 2026

EL AUTOR

Cindy Malavé Ramírez

AGRADECIMIENTO

Agradezco a Dios por permitirme contar con la fortaleza, conocimiento y bienestar necesarios para lograr esta meta, además de concederme constancia y serenidad para enfrentar las dificultades surgidas a lo largo de este camino y alcanzar mis objetivos.

Agradezco profundamente a mis padres, quienes han sido mi principal apoyo en cada etapa de mi vida. A mi padre, por ser un ejemplo de esfuerzo y superación; y a mi madre, por su amor, apoyo incondicional y valores que han contribuido a mi formación personal y profesional.

A mis hermanos, por su apoyo, orientación y compañía a lo largo de este camino académico. Asimismo, agradezco a los docentes de la Carrera en Tecnologías de la Información y la Comunicación por compartir sus conocimientos, experiencias y enseñanzas que enriquecieron mi aprendizaje.

Agradezco al Grupo de Plasma del Departamento de Física Aplicada de la Universidad de Cantabria, España por facilitarme el acceso a los datos de la Santander Linear Plasma Machine SLPM, esenciales para la realización de este estudio.

Finalmente, agradezco a mi tutor, Fís. Omar Castellanos, por su orientación y apoyo durante el desarrollo de este proyecto.

Cindy Ivonne, Malavé Ramírez

DEDICATORIA

Dedico este trabajo a Dios, por brindarme la fortaleza, sabiduría y perseverancia necesarias para superar cada desafío y permitirme culminar esta importante etapa de mi formación profesional.

A mis padres, por acompañarme en cada etapa de este camino, creer siempre en mí y brindarme la fortaleza para alcanzar esta meta. A mi padre, por enseñarme con su ejemplo el valor del esfuerzo y la constancia; y a mi madre, por su amor incondicional, apoyo permanente y por impulsarme siempre a alcanzar mis metas.

A mis hermanos, por acompañarme en este camino con su cariño, confianza y palabras de aliento, brindándome motivación para seguir adelante en los momentos de mayor dificultad.

Finalmente, dedico este logro a todas las personas que, de una u otra manera, formaron parte de este proceso y contribuyeron a mi crecimiento personal y académico. Este trabajo representa el fruto del esfuerzo, la dedicación y el compromiso asumidos para alcanzar una de las metas más importantes de mi vida.

Cindy Ivonne, Malavé Ramírez

ÍNDICE GENERAL

TITULO DEL TRABAJO DE TITULACIÓN	I
TRIBUNAL DE SUSTENTACIÓN	II
CERTIFICACIÓN	III
DECLARACIÓN DE RESPONSABILIDAD	IV
CERTIFICACIÓN DE ANTIPLAGIO	V
AUTORIZACIÓN	VI
AGRADECIMIENTO	VII
DEDICATORIA	VIII
ÍNDICE GENERAL	IX
ÍNDICE DE TABLAS	XI
ÍNDICE DE FIGURAS	12
RESUMEN	14
ABSTRACT.	14
INTRODUCCIÓN	15
CAPÍTULO 1. FUNDAMENTACIÓN	16
1.1 Antecedentes	16
1.2 Descripción del Proyecto	17
1.3. Objetivos del Proyecto	23
1.4. Justificación del proyecto	24
1.5. Alcance del Proyecto	26
CAPÍTULO 2. MARCO TEÓRICO Y METODOLOGÍA DEL PROYECTO	31
2.1. Marco conceptual	31
2.2 Marco teórico	39

2.3. Metodología del proyecto	41
2.3.1. Metodología de Investigación	41
2.3.2. Metodología de desarrollo.	42
CAPÍTULO 3. PROPUESTA	44
3.1. Requerimientos	44
3.1.1. Requerimientos Funcionales	44
3.1.2. Requerimientos no Funcionales	49
3.2. Componente de la Propuesta	50
3.2.1. Arquitectura del Sistema	51
3.2.2. Diagrama	53
3.2.3. Modelado de datos	57
3.2.4. Entrenamiento y evaluación del modelo	62
3.3. Diseño de interfaz	65
3.4. Pruebas	68
CONCLUSIONES	74
RECOMENDACIONES	75
BIBLIOGRAFÍA	76
ANEXOS	81

ÍNDICE DE TABLAS

Tabla 1:Requerimiento Funcional – Módulo de Adquisición	45
Tabla 2: Requerimiento Funcional – Módulo de Procesameinto:	45
Tabla 3: Requerimiento Funcional - Módulo de entrenamiento del modelo	46
Tabla 4:Requerimiento Funcional - Módulo de Detección de anomalía	47
Tabla 5: Requerimiento Funcional - Visualización de gráficos	48
Tabla 6: Requerimiento Funcional - Módulo de resultados	49
Tabla 7: Requerimientos no funcionales	50
Tabla 8: Comparación entre el enfoque tradicional y el sistema propuesto	50
Tabla 9: Entrenar modelo	54
Tabla 10: Detectar anomalías	55
Tabla 11: Resultados	56
Tabla 12: Variables de Identificación	58
Tabla 13:Variables de energía	58
Tabla 14:Variables estadísticas	59
Tabla 15: Variables de activida	59
Tabla 16: Variables de agrupamiento	60
Tabla 17: Variables temporales	60
Tabla 18: Variable objetivo	60
Tabla 19: Reglas utilizadas para la generación de la variable objetivo.	62
Table 20: Caso de prueba importación del video	69
Tabla 21: Caso de prueba preprocesamiento	70
Tabla 22: Caso de prueba entrenamiento del modelo	71
Tabla 23: Caso de prueba detección de anomalías	72
Table 24: Caso de prueba de Generación del reporte	73

ÍNDICE DE FIGURAS

Figura 1: Metodología incremental	43
Figura 2: Arquitectura del sistema	52
Figura 3: Caso de uso del sistema	53
Figura 4: Matriz de confusión del experimento 17 características	63
Figura 5: Características más importantes del experimento 17 características	64
Figura 6: Matriz de confusión del experimento 25 características	64
Figura 7: Características más importantes del experimento 25 características	65
Figura 8: Pantalla de importación de datos	66
Figura 9: Pantalla de preprocesamiento	66
Figura 10: Pantalla de entrenamiento	67
Figura 11: Pantalla de detección de anomalías	67
Figura 12: Pantalla Resultados	68
Figura 13: Resultado de la prueba de importación del video	69
Figura 14: Resultado de la prueba de preprocesamiento	70
Figura 15: Resultado de la prueba entrenamiento del modelo	71
Figura 16: Resultado de la prueba detección de anomalías	72
Figura 17: Resultado de la prueba de Generación del reporte	73
Figura 18: Mensaje de error al cargar un archivo cih	81
Figura 19: Selección de archivo	81
Figura 20: Archivo cargado correctamente	81
Figura 21: Extracción de característica	82
Figura 22: Entrenamiento del modelo	82
Figura 23: Selección de archivo pantalla anomalías	82
Figura 24: Archivo cargado correctamente	83
Figura 25: Seleccionar al menos una región de interés	83
Figura 26: Probabilidad del modelo en las regiones de interés	83
Figura 27: Energía absoluta de una región de interés	84
Figura 28: Energía normalizada de una región de interés	84
Figura 29: Gráfica de z-score de una región de interés	84

Figura 30: gráfica de probabilidad de modelo	85
Figura 31: Guardado en análisis en su respectivo archivo	85
Figura 32: Vizualización pantalla resultado	85
Figura 33: Reporte de detección	86

RESUMEN

El presente trabajo desarrolla una interfaz gráfica para la detección de anomalías en el plasma usando visión artificial y Random Forest. El objetivo principal es identificar patrones inusuales a partir del análisis de vídeo, lo que facilita un seguimiento automatizado y preciso del fenómeno. Para el desarrollo del sistema se utiliza la metodología incremental, integrando módulos de adquisición de datos, procesamiento de imágenes y detección mediante modelos de aprendizaje automático. Los datos se derivan de grabaciones experimentales obtenidas en condiciones controladas. Como resultado, el sistema puede detectar diferencias significativas en la intensidad y los patrones del plasma, lo que permite una identificación eficiente de anomalías. Se concluye que el uso de técnicas de visión por computadora y aprendizaje automático mejora la precisión y confiabilidad del análisis de plasma, facilitando la investigación.

Palabras claves: Detección de anomalías, visión por computadora, aprendizaje automático.

ABSTRACT.

This work develops a computer-based system using machine learning techniques for the detection of anomalies in plasma behavior within a linear non-confining machine. The main objective is to identify unusual patterns through video analysis, enabling automated and accurate monitoring of the phenomenon. The system is developed using an incremental methodology, integrating data acquisition, image processing, and detection modules based on machine learning models. The data are obtained from experimental recordings collected under controlled conditions. As a result, the system is capable of detecting significant variations in plasma intensity and patterns, allowing for efficient anomaly identification. It is concluded that the application of computer vision and machine learning techniques improves the accuracy and reliability of plasma analysis, thereby facilitating scientific research.

Keywords: Anomaly detection, computer vision, machine learning.

INTRODUCCIÓN

Actualmente, el uso de herramientas tecnológicas se ha convertido en un soporte fundamental para el análisis de fenómenos científicos, ya que permite procesar de manera más rápida y precisa grandes cantidades de información. En el caso del estudio del plasma, entender su comportamiento es un serio desafío porque sus patrones varían y hay cambios que no siempre son fáciles de identificar a simple vista. Por esta razón existe la necesidad de incluir métodos que faciliten el análisis.

El análisis del comportamiento del plasma implica varios pasos, comenzando con la adquisición de datos experimentales, generalmente mediante grabaciones de vídeo, y continuando con el procesamiento e interpretación. Este proceso se realiza de forma tradicional mediante herramientas especializadas, pueden surgir dificultades, como detección tardía de anomalías, errores en la interpretación de patrones o incluso pérdida de información relevante.

Este proyecto propone el desarrollo de un sistema computacional basado en aprendizaje automático para detectar anomalías en plasma mediante análisis de imágenes. Su objetivo es automatizar parte del proceso de análisis, permitiendo identificar comportamientos inusuales de forma más eficiente y reducir la dependencia de la observación manual.

Este documento se estructura en tres capítulos, recomendaciones y conclusiones. El Capítulo 1 presenta el contexto general del proyecto, incluyendo antecedentes, problemática y justificación. El Capítulo 2 aborda las bases teóricas y conceptuales, junto con los métodos y herramientas de recopilación de datos. Finalmente, el Capítulo 3 describe los aspectos técnicos del sistema, como su arquitectura, herramientas utilizadas, requisitos, diseño y pruebas realizadas para verificar su funcionamiento.

CAPÍTULO 1. FUNDAMENTACIÓN

1.1 Antecedentes

En los últimos años, el desarrollo de la visión por computadora y el aprendizaje automático ha impulsado diversas investigaciones dirigidas al análisis automatizado de información visual, estas tecnologías han sido aplicadas en diversos contextos académicos y científicos, permitiendo mejorar la interpretación de datos de secuencias de imágenes y videos, así como optimizar procesos que antes se realizaban con la metodología tradicional [1].

La integración de métodos de procesamiento de imágenes y modelos de aprendizaje automático ha permitido desarrollar sistemas capaces de identificar patrones, detectar anomalías y automatizar procesos de análisis en diversos tipos de datos visuales, estas herramientas han demostrado ser particularmente útiles en entornos donde es necesario analizar grandes cantidades de información de manera eficiente y precisa, reduciendo la intervención humana y mejorando la confiabilidad de los resultados [2].

A continuación, se presentan algunos trabajos previos relacionados con esta investigación:

En la ciudad de Guayaquil se desarrolló un proyecto que tuvo como objetivo utilizar técnicas de visión artificial para determinar las características de los objetos a través del procesamiento de imágenes. En este estudio se utilizó el entorno LabVIEW para identificar viales con turbidez, permitiendo automatizar el proceso de inspección visual y mejorar la precisión de la detección de anomalías [3]. Este proyecto es similar a esta propuesta en aspectos como el uso de herramientas de visión artificial para el análisis automatizado de información visual.

Se desarrolló un estudio en Quito, Ecuador, con el objetivo de implementar el modelo yolo para la detección de matrículas de vehículos en imágenes y videos para automatizar la identificación de personajes mediante técnicas ocr, esta investigación se centra en la aplicación de la visión por computadora y el

aprendizaje profundo, incluida la capacitación por transferencia, para optimizar la capacitación del modelo sin comenzar desde cero, mejorando la eficiencia y precisión del análisis de secuencias de video. Este enfoque es relevante para este proyecto porque comparte una base metodológica para usar yolo para detectar objetos en sistemas. análisis visual utilizado en un entorno de monitoreo [4].

En el Reino Unido se desarrolló una investigación para evaluar una serie de algoritmos de aprendizaje automático aplicados a problemas de clasificación de datos reales, el estudio analiza el rendimiento de varios modelos, incluidas las máquinas de vectores y otros clasificadores supervisados, para determinar su eficacia en tareas de predicción y reconocimiento de patrones. Los resultados muestran que los métodos basados en árboles como random forest proporcionan un alto rendimiento debido a su capacidad para manejar datos complejos y reducir el sobreajuste [5]. Esta propuesta adquiere especial relevancia en el marco del presente proyecto, dado que permite constatar la eficacia del aprendizaje automático en sistemas automatizados de análisis de información y clasificación de datos.

Finalmente, el uso de técnicas de visión por computadora y aprendizaje automático ha sido ampliamente utilizado en diversos contextos, su uso en el análisis de fenómenos físicos como el plasma es aún limitado, lo que justifica el desarrollo de propuestas encaminadas a integrar estas herramientas en este tipo de entornos.

1.2 Descripción del Proyecto

Este proyecto consiste en el desarrollo de un sistema inteligente para el análisis del comportamiento del plasma mediante técnicas de visión artificial, procesamiento digital de imágenes, análisis estadístico y aprendizaje automático. La solución propuesta tiene como objetivo procesar y analizar automáticamente secuencias de datos experimentales obtenidas a partir de archivos en formato .cih e imágenes RAW, que contienen información visual relacionada con el fenómeno físico estudiado.

El sistema realiza un procesamiento cuadro a cuadro de las secuencias experimentales, permitiendo extraer características relevantes como intensidad del

plasma, energía, variabilidad, entropía, contraste, desviación estándar, varianza y comportamiento temporal. El sistema mantiene un historial temporal independiente para cada región de interés (ROI), permitiendo analizar la evolución energética y estadística del plasma a lo largo del tiempo. Esto facilita un análisis cuantitativo y cualitativo más preciso del fenómeno estudiado.

En los enfoques tradicionales, la detección de anomalías en el plasma suele basarse en el análisis de variables estadísticas, como la energía o el z-score, comparando sus valores con umbrales previamente establecidos para determinar si existe un comportamiento anómalo. Este procedimiento puede requerir ajustes manuales y presenta limitaciones cuando los datos muestran patrones más complejos. En la propuesta desarrollada, estas variables continúan siendo calculadas durante la extracción de características, pero la decisión final es realizada por un modelo de aprendizaje automático basado en Random Forest, el cual clasifica automáticamente cada región de interés como normal o anómala.

El sistema contará con una interfaz gráfica de usuario que permitirá el monitoreo, control y supervisión del procesamiento de datos experimentales. A través de esta interfaz, el usuario podrá cargar secuencias de video, visualizar imágenes originales y procesadas, definir regiones de interés (ROI), observar gráficos dinámicos relacionados con las mediciones obtenidas y analizar los resultados de la clasificación automática, incluyendo la detección de posibles anomalías. Asimismo, el sistema permitirá almacenar resultados, generar conjuntos de datos y registrar información relevante para facilitar posteriores análisis y documentación de los experimentos realizados.

El sistema consta de los siguientes módulos:

El módulo de adquisición de datos: Es el encargado de cargar y preparar secuencias de video en formatos CIH y RAW, que representan información experimental del plasma. Su función principal es convertir estos archivos en una secuencia de frames que puedan ser utilizados posteriormente en las etapas de preprocesamiento y

análisis. Además, permite al usuario seleccionar el archivo desde una interfaz gráfica y visualizar su contenido de forma inicial.

Una vez cargado el video, el módulo proporciona una vista previa interactiva basada en frames individuales, permitiendo la navegación mediante un control deslizante. También muestra información relevante del archivo, como la ruta, la resolución, el FPS original y el número total de frames, lo que facilita la validación de los datos antes de iniciar el procesamiento. Adicionalmente, el usuario puede definir parámetros como FPS objetivo y número de frames a procesar.

Este módulo gestiona el flujo inicial del sistema, asegurando que solo se pueda avanzar al siguiente módulo cuando un video válido ha sido cargado. También incluye funciones de limpieza que permiten reiniciar el estado del sistema para cargar nuevos datos sin conflictos. De esta manera, el módulo de adquisición actúa como la base de análisis del sistema Plasma

Módulo de preprocesamiento: Reúne varias funciones en una sola etapa del sistema, donde se prepara todo lo necesario antes del entrenamiento o la detección. Aquí no solo se trabaja con la imagen del plasma, sino también con la selección de zonas específicas, el cálculo de variables y la organización de los datos.

En la parte de segmentación se trabaja con las regiones de interés (ROI), que el usuario define directamente desde la interfaz sobre el video. Estas regiones permiten enfocarse solo en zonas relevantes del plasma. Además, el sistema puede manejar una selección más automática de áreas activas, lo que ayuda a detectar cambios sin depender totalmente de la intervención manual.

Luego viene la parte de extracción de características, donde cada frame se convierte en información numérica. Aquí se calculan valores como energía, intensidad media, desviación estándar, varianza, z-score, entropía, contraste y otros indicadores del comportamiento del plasma. También se analizan patrones como la actividad de píxeles, clusters energéticos y variaciones en el tiempo, lo que permite describir el fenómeno de forma más completa.

También cumple la función de generar y organizar los datos para el modelo. Toda esta información se guarda en archivos estructurados como CSV, que luego se usan para entrenar el sistema de aprendizaje automático. A la vez, se realiza un análisis estadístico básico para detectar desviaciones o comportamientos fuera de lo normal, usando medidas como el z-score y comparaciones con valores de referencia

El módulo de entrenamiento: El sistema toma los datos generados en el preprocesamiento para construir el modelo de detección de anomalías. Aquí se usa principalmente un flujo basado en Scikit-learn, donde el algoritmo Random Forest aprende a partir de las características del plasma como energía, variación, clusters y otras métricas ya calculadas en los frames.

En la interfaz se tiene un panel lateral donde se muestra el estado del modelo y la información básica del entrenamiento, como el número de muestras utilizadas, las características del dataset y la cantidad de anomalías presentes. Desde ahí también se controla todo el ciclo del modelo con botones para entrenar, reentrenar, guardar o cargar un modelo previamente entrenado, lo que permite manejarlo sin depender siempre de volver a entrenar desde cero.

En la parte derecha se muestran los resultados del entrenamiento. Se visualizan las métricas principales como accuracy, precision, recall y F1-score, que permiten evaluar qué tan bien está funcionando el modelo. También se muestra la importancia de las variables para entender qué características influyen más en la predicción, y una matriz de confusión que resume los aciertos y errores del modelo al clasificar entre estados normales y anomalías.

El módulo de detección de anomalías: integra el procesamiento del video de plasma con un modelo de aprendizaje automático previamente entrenado. A partir de la interfaz, el usuario carga el video, define regiones de interés (ROI) y ajusta parámetros como umbral de anomalía y sensibilidad. Estas configuraciones permiten que el sistema analice cada cuadro del video de forma controlada, enfocándose únicamente en las zonas relevantes para el estudio del plasma. La vista

central del sistema muestra el video con superposición de información, facilitando la observación directa del comportamiento del fenómeno.

El análisis se basa en la extracción de características como energía, desviación estándar, número de clusters y z-score, que se muestran en tiempo real en el panel derecho. Estas variables son evaluadas junto con la probabilidad generada por el modelo para determinar si un cuadro corresponde a un comportamiento normal o anómalo. Las variables estadísticas calculadas forman parte del conjunto de características utilizadas por el modelo Random Forest para clasificar automáticamente cada ROI. Los eventos detectados se registran en una tabla con información detallada como frame, tiempo, ROI y tipo de anomalía.

También actúa como una interfaz de monitoreo del sistema, permitiendo visualizar el estado del modelo, los eventos detectados y las características actuales del análisis. Toda la información se presenta de forma estructurada para facilitar la interpretación del comportamiento del plasma durante la ejecución.

El módulo de resultados se encarga de organizar y presentar la información generada durante todo el flujo del sistema de análisis de plasma. Aquí se agrupan los datos más importantes del procesamiento, como la cantidad de frames analizados, el número de anomalías detectadas, el modelo utilizado y su nivel de precisión. Esta información se muestra en la parte superior mediante indicadores tipo “cards”, lo que permite tener una visión rápida del estado general del sistema sin necesidad de revisar cada proceso por separado.

En la parte central se visualiza el detalle de los eventos detectados durante la ejecución del sistema. Cada evento contiene información como el frame, la región de interés (ROI), la probabilidad de anomalía, el tipo de evento y su estado. Además, se incluye una vista previa del frame donde ocurrió la anomalía, lo que facilita la interpretación visual de los resultados. Esta sección funciona como el núcleo de análisis posterior, ya que permite revisar con mayor detalle lo que el modelo fue detectando en el tiempo.

Tecnologías utilizadas

Para el desarrollo del sistema propuesto se utiliza principalmente el lenguaje de programación Python, debido a su amplia capacidad para el desarrollo de sistemas científicos. El lenguaje de programación Python se utiliza para desarrollar el sistema propuesto principalmente por su capacidad integral para desarrollar sistemas científicos, es fácil de integrar con bibliotecas especializadas y es eficiente en la gestión de datos. Python permite la implementación de lógica de procesamiento de imágenes, análisis estadístico y algoritmos de aprendizaje automático, que forman el núcleo del sistema desarrollado.

La visión por computadora y el procesamiento de imágenes digitales utilizan la biblioteca OpenCV para realizar operaciones básicas como leer y manipular secuencias de video, conversión de imágenes, filtrado, normalización de intensidad, cálculo de diferencia de cuadros y segmentación usando regiones de interés (ROI). Estas características son esenciales para el análisis del comportamiento del plasma.

NumPy se utiliza para gestionar datos numéricos y operaciones matemáticas, lo que le permite trabajar con estructuras multidimensionales y realizar cálculos eficientes relacionados con la energía, la variación del tiempo y el análisis estadístico. Además, Pandas se utiliza para organizar, estructurar y almacenar datos en formato tabular, lo que facilita la generación de conjuntos de datos utilizados en el entrenamiento de modelos.

El aprendizaje automático utiliza la biblioteca Scikit-learn, que proporciona herramientas para entrenar, evaluar y validar modelos predictivos. Este sistema implementa el algoritmo Random Forest, elegido por su capacidad para manejar datos complejos y su robustez al ruido y la variabilidad de los datos experimentales. Jolib se utiliza para almacenar un modelo entrenado de modo que el modelo pueda guardarse y reutilizarse en ejecuciones futuras sin necesidad de volver a entrenar.

Matplotlib se utiliza para la visualización de datos, lo que permite la generación de gráficos dinámicos y estáticos relacionados con la evolución del plasma, como energía, intensidad y probabilidad de anomalías. Estas visualizaciones están integradas en la interfaz gráfica para facilitar la interpretación de los resultados.

La interfaz gráfica del sistema se desarrolla utilizando PyQt6, que le permite crear una aplicación de escritorio con componentes visuales. A través de esta interfaz, el usuario puede cargar datos, definir regiones de interés, controlar el procesamiento, ver resultados en tiempo real y analizar gráficos dinámicos. La integración de herramientas de visualización en la interfaz permite una experiencia de usuario más completa y eficiente.

En cuanto al almacenamiento de información, el sistema implementa un enfoque basado en archivos que permite guardar los resultados en formatos estructurados como CSV, así como guardar imágenes relacionadas con los eventos relevantes detectados durante el análisis. El sistema permite generar reportes PDF que integran métricas, gráficas y resultados del análisis, facilitando un mayor análisis del comportamiento del plasma.

El entorno de desarrollo utilizado es Visual Studio Code, que proporciona un entorno flexible y potente para escribir, depurar y organizar el código fuente. Este entorno permite el trabajo modular, lo que facilita la integración de varios componentes del sistema. Además, ofrece soporte para extensiones de Python que optimizan el desarrollo, la depuración y la gestión de librerías utilizadas en el proyecto.

Finalmente, el sistema se implementa como una aplicación de escritorio que se ejecuta localmente en el ordenador del usuario sin necesidad de servidores externos. Esta arquitectura reduce la complejidad del sistema, mejora el rendimiento computacional y permite su uso en entornos experimentales controlados donde se requiere un análisis eficiente.

1.3. Objetivos del Proyecto

Objetivo General

- ✓ Desarrollar una interfaz gráfica que utiliza un sistema computacional basado en visión artificial, análisis estadístico y aprendizaje automático, capaz de analizar secuencias de video de plasma para detectar anomalías.

Objetivos Específicos

- ✓ Diseñar una interfaz gráfica de usuario (GUI) que integre todos los módulos del sistema, permitiendo cargar secuencias de video de plasma, visualizar cuadros originales y procesados, y controlar los parámetros de preprocesamiento para acondicionar los datos de entrada.
- ✓ Integrar una interfaz gráfica de usuario (GUI) con librerías de procesamiento de imágenes y aprendizaje automático, garantizando la comunicación eficiente entre el usuario y el sistema.
- ✓ Aplicar un método de segmentación basado en regiones de interés (ROI) que permita analizar de forma localizada el comportamiento del plasma y permitir que desde la GUI la selección de los ROIs.
- ✓ Generar un dataset a partir de las características extraídas del plasma y utilizarlo para entrenar un modelo de aprendizaje automático basado en Random Forest para la detección de anomalías.

1.4. Justificación del proyecto

El análisis de fenómenos físicos complejos como el plasma requiere herramientas que permitan procesar e interpretar grandes cantidades de información visual generada durante experimentos científicos, que son difíciles de resolver utilizando métodos tradicionales [6]. En muchos casos, estas pruebas dependen de la observación manual o del uso de equipos especializados, lo que significa altos costos, condiciones controladas y personal altamente capacitado, limitando su uso a ambientes académicos con recursos limitados.

El desarrollo basado en visión por computadora, análisis estadístico y técnicas de aprendizaje automático es una alternativa efectiva para automatizar el análisis de secuencias de video [7]. Estas tecnologías permiten la identificación de patrones, variaciones energéticas y comportamientos dinámicos que no siempre pueden ser detectados por la observación humana, mejorando la precisión y reduciendo la subjetividad del análisis [8].

Este proyecto está motivado por la necesidad de una herramienta que integre procesamiento de imágenes digitales, análisis estadístico y modelos de aprendizaje automático para estudiar el comportamiento del plasma. Este enfoque permite no sólo caracterizar el fenómeno, sino también detectar anomalías utilizando indicadores cuantitativos como la intensidad de la luz, variaciones entre fotogramas y análisis estadístico, brindando una solución desde un enfoque de tecnologías de la información y las comunicaciones.

Además, la inclusión de mecanismos de validación temporal nos permite reducir los falsos positivos debido a eventos constantes en el tiempo, lo que aumenta la confiabilidad del sistema en el análisis dinámico de datos [9]. Asimismo, la generación automática de datos a partir del procesamiento de vídeo permite mejorar el modelo de aprendizaje automático, facilitando su adaptación a diversos escenarios experimentales y fortaleciendo su capacidad de generalización [10].

El desarrollo de este sistema contribuye a fortalecer el uso de tecnologías informáticas en el análisis de fenómenos físicos, ofreciendo una solución accesible, automatizada y utilizable en el entorno académico y de investigación, lo que es un importante aporte al campo de las TIC aplicadas a la ciencia.

Esta propuesta se orienta hacia el Plan de Desarrollo Ecuador No Se Detiene 2025-2026 específicamente en los siguientes puntos

- Eje social en concordancia con el objetivo 2: Potenciar las capacidades de la ciudadanía con acceso universal a una educación inclusiva de calidad, acceso a espacios de intercambio cultural y una vida activa [11]. Correspondiendo a la política 2.1, y objetivo 2.2.
- Eje Institucional relacionado al objetivo 8 Fortalecer la institucionalidad publica de forma eficiente, transparente y participativa: en cumplimiento de la política 8.3 [11].

1.5. Alcance del Proyecto

El objetivo de este trabajo es desarrollar un sistema inteligente de monitoreo y análisis del comportamiento del plasma a partir de datos experimentales obtenidos de un dispositivo de plasma lineal no confinado. Actualmente, el análisis de estos datos suele realizarse de forma manual o mediante herramientas limitadas, lo que dificulta la identificación temprana de anomalías y la correcta interpretación de los resultados experimentales. Para abordar esta problemática, se propone el desarrollo de una aplicación capaz de automatizar el procesamiento, análisis y visualización de los datos experimentales.

Los datos utilizados corresponden a mediciones reales obtenidas de una máquina lineal de plasma en condiciones controladas de laboratorio. Cada archivo tiene formato .cih con resolución de 128x80 píxeles. Los datos fueron proporcionados por el Laboratorio de Física de Plasma, del Departamento de Física Aplicada, de la Facultad de Ciencias, de la Universidad de Cantabria, España.

El sistema estará orientado al ámbito de investigación, permitiendo al usuario acceder de manera clara y organizada a la información procesada, facilitando la identificación de patrones, variaciones temporales y comportamientos anómalos del plasma. Además, el proyecto contempla la generación incremental de conjuntos de datos a partir de videos experimentales, los cuales servirán como base para el entrenamiento, evaluación y actualización de modelos de aprendizaje automático implementados en el sistema.

Módulo de adquisición de datos(Interfaz gráfica con librerías): Este módulo se encarga de la carga y preparación de secuencias de video en formatos CIH y RAW para su posterior análisis dentro del sistema. Su objetivo principal es convertir los archivos experimentales en una secuencia de frames utilizables por los demás módulos del flujo de procesamiento.

- Carga de videos experimentales de plasma desde una interfaz gráfica mediante selección de archivos.

- Extracción de fotogramas a partir del video para su procesamiento individual.
- Lectura y visualización de información del archivo como ruta, resolución, FPS y número total de frames.
- Navegación de frames mediante un control deslizante para inspección visual del video.
- Configuración de parámetros de procesamiento como FPS objetivo a analizar.
- Validación básica de los archivos cargados para asegurar su compatibilidad con el sistema.
- Reinicio y limpieza de datos cargados para permitir nuevas sesiones sin conflictos.
- Control del flujo del sistema, habilitando el avance al siguiente módulo únicamente cuando existe un video válido cargado.

Módulo de preprocesamiento (Segmentación de ROI y generación de dataset):

Este módulo reúne en una sola etapa todo el proceso de preparación de los datos del plasma antes de su análisis mediante modelos de aprendizaje automático. Su función principal es transformar los frames del video en información estructurada, integrando procesos de segmentación, extracción de características, generación de datos y análisis estadístico, con el objetivo de dejar la información lista para etapas posteriores del sistema.

- Conversión de imágenes a escala de grises para facilitar el análisis del contenido visual.
- Normalización de intensidad para estandarizar los valores entre frames.
- Mejora de contraste para resaltar estructuras relevantes del plasma.
- Cálculo de diferencias entre fotogramas para identificar cambios energéticos en el tiempo.
- Selección de regiones de interés (ROI) desde la interfaz gráfica para el análisis localizado del plasma.

- Análisis por regiones para enfocar el estudio en zonas específicas del fenómeno.
- Cálculo de características del plasma como energía, intensidad media, desviación estándar, varianza, z-score, entropía y contraste.
- Análisis de actividad de píxeles, porcentaje de actividad y formación de clusters energéticos.
- Cálculo de variaciones temporales de energía para identificar comportamiento dinámico.
- Almacenamiento de las características extraídas en archivos CSV estructurados.
- Organización de datos para su uso en entrenamiento y validación de modelos de aprendizaje automático.
- Etiquetado de datos como estados normales o anómalos.

Módulo de entrenamiento (Random forest): Este módulo se encarga de construir el modelo de detección de anomalías a partir de los datos generados en el preprocesamiento. Utiliza un enfoque basado en Scikit-learn, donde el algoritmo Random Forest aprende a partir de las características extraídas del plasma como energía, variación temporal, clusters y demás métricas calculadas por frame.

- Construcción del modelo de detección de anomalías a partir de datos preprocesados del sistema.
- Entrenamiento del modelo utilizando Scikit-learn con el algoritmo Random Forest.
- Uso de características del plasma como energía, variación, clusters y métricas estadísticas como entrada del modelo.
- Clasificación del comportamiento del plasma en estados normales y anómalos.
- Gestión del ciclo de entrenamiento, incluyendo entrenamiento, reentrenamiento y guardado.

- Almacenamiento del modelo entrenado para su reutilización sin necesidad de reentrenamiento desde cero.
- Visualización del estado del modelo desde la interfaz gráfica, incluyendo número de muestras, características del dataset y presencia de anomalías.
- Evaluación del modelo mediante métricas como accuracy, precision, recall y F1-score.
- Análisis de la importancia de las variables para determinar su influencia en la predicción.
- Representación de la matriz de confusión para visualizar aciertos y errores del modelo.

Módulo de detección de anomalías: Este módulo integra el procesamiento del video de plasma con un modelo de aprendizaje automático previamente entrenado. Permite analizar el comportamiento del plasma a partir de la selección de regiones de interés (ROI) y la configuración de parámetros como la sensibilidad de modelo.

- Integración del video de plasma con un modelo de aprendizaje automático previamente entrenado.
- Carga de video experimental y procesamiento por frames.
- Selección y gestión de regiones de interés (ROI) desde la interfaz gráfica.
- Visualización del video con superposición de información del análisis.
- Evaluación conjunta de métricas estadísticas y salida del modelo para detección de anomalías.
- Clasificación de cada frame en estados normales o anómalos.
- Registro de eventos detectados con información de frame, tiempo, ROI, probabilidad y tipo.
- Presentación de resultados en tabla para análisis detallado del comportamiento del plasma.
- Monitoreo del estado del modelo durante la ejecución del sistema.

Módulo de resultados: Este módulo se encarga de organizar y presentar toda la información generada durante el flujo completo del sistema de análisis de plasma.

Su función principal es consolidar los resultados del procesamiento, y detección de anomalías para facilitar su interpretación y análisis.

- Organización de los resultados generados durante todo el flujo del sistema de análisis de plasma.
- Visualización de métricas generales como número de frames analizados, anomalías detectadas y precisión del modelo.
- Visualización de eventos detectados con información detallada como frame, ROI, probabilidad, tipo y estado.
- Presentación de una vista previa del frame donde se detecta la anomalía para análisis visual.
- Gestión de sesiones de trabajo desde un panel lateral.
- Selección de sesiones anteriores para revisión de resultados históricos.
- Grafica de z-score, energía absoluta, energía normalizada y probabilidad del modelo
- Eliminación de sesiones del sistema.
- Exportación pdf de la sesión.

CAPÍTULO 2. MARCO TEÓRICO Y METODOLOGÍA DEL PROYECTO

2.1. Marco conceptual

Sistema Computacional

Un sistema computacional consta de un conjunto de elementos de hardware y software que trabajan juntos para procesar información, ejecutar programas y gestionar datos, permitiendo la automatización de tareas, almacenando información y realizando procesos que facilitan el análisis en una variedad de aplicaciones, estos elementos se organizan de manera jerárquica, donde el hardware ejecuta las instrucciones definidas por el software [12].

Interfaz gráfica

Una interfaz gráfica de usuario permite la interacción entre el usuario y el sistema utilizando elementos visuales como botones, ventanas y gráficos que facilitan el acceso a las funciones del sistema y brindan una mejor experiencia de usuario al simplificar la comunicación entre el usuario y el software este tipo de interfaz reduce la necesidad de memorizar comandos o instrucciones complejas [13].

Visual Studio Code

Visual Studio Code es un editor de código fuente desarrollado por Microsoft que le permite escribir, editar y depurar programas en una variedad de lenguajes de programación, ofreciendo un entorno de desarrollo liviano pero altamente funcional que incluye herramientas como resaltado de sintaxis, autocompletado inteligente, depuración integrada y soporte de control de versiones que facilita la búsqueda de errores y la organización de extensiones de código [14].

Python

Python es un lenguaje de programación de alto nivel ampliamente utilizado en el desarrollo de aplicaciones científicas, análisis de datos y aprendizaje automático, caracterizado por una sintaxis clara y simple que facilita la comprensión del código, también cuenta con bibliotecas especializadas que permiten implementar de manera

eficiente algoritmos complejos y trabajar con procesamiento de imágenes, análisis numérico y modelos predictivos en un mismo entorno de desarrollo [15].

OpenCV

OpenCV es una biblioteca de código abierto que permite trabajar de manera eficiente con imágenes y videos utilizando varios algoritmos ya implementados, puede realizar tareas como detección de objetos, segmentación, filtrado y análisis de movimiento, lo que lo hace muy útil en sistemas que necesitan interpretar automáticamente información visual. También integra fácilmente lenguajes como Python y C, facilitando su uso tanto en proyectos académicos como aplicados donde se procesan grandes cantidades de datos visuales [16].

Numpy

NumPy es una de las bibliotecas más importantes para trabajar con datos numéricos en Python, ya que ofrece estructuras como matrices multidimensionales que permiten un procesamiento de información más rápido y eficiente, también incluye funciones matemáticas que ayudan a realizar operaciones complejas sin tener que programarlas desde cero, lo cual es especialmente útil en áreas como análisis de datos, procesamiento de señales y aprendizaje automático [17].

Pandas

Pandas es una biblioteca de Python que tiene como objetivo analizar y manipular datos estructurados, brindando herramientas eficientes para trabajar con tablas utilizando estructuras como DataFrames que permiten organizar la información de forma clara y flexible, facilitando tareas como limpiar, transformar, filtrar y resumir datos. su integración con otras bibliotecas científicas lo convierten en una herramienta clave en los flujos de trabajo de ciencia de datos [18].

Scikit-learn

Scikit-learn es una biblioteca de aprendizaje automático de Python que contiene una variedad de algoritmos de clasificación, regresión y agrupamiento que le permiten resolver diferentes tipos de problemas relacionados con datos, también

ofrece una forma sencilla de entrenar y evaluar modelos, lo que facilita su uso incluso para quienes son nuevos en el campo, también incluye herramientas para la preparación de datos, selección de funciones y validación de resultados, lo que ayuda a crear modelos más confiables y fáciles de implementar [19].

Matplotlib

Matplotlib es una biblioteca de visualización de datos de Python que permite generar gráficos estáticos, animados e interactivos, facilitando la representación visual de información numérica, se utiliza ampliamente en el análisis de datos de investigación, ya que le permite crear gráficos como histogramas, diagramas de dispersión, gráficos de líneas y visualizaciones personalizadas, su flexibilidad y compatibilidad con otras bibliotecas científicas [20].

PyQt6

PyQt6 es una herramienta que le permite crear interfaces gráficas en Python utilizando el marco Qt. Con él se pueden desarrollar aplicaciones de escritorio que incluyen ventanas, botones, menús y otros elementos visuales que facilitan al usuario la interacción con el sistema, también se puede integrar con aplicaciones que procesan datos o imágenes, lo que lo hace útil en proyectos donde se necesita mostrar resultados y controlar el sistema a través de una interfaz [21].

Dataset

Un dataset es un conjunto de datos ordenados que se utilizan para entrenar y evaluar modelos de aprendizaje automático, generalmente estructurados en filas y columnas que representan observaciones y variables, lo que permite su uso en procesos de análisis, modelado y validación de algoritmos, y facilita la gestión y organización de la información entre aplicaciones [22].

Los archivos CSV

Los archivos CSV (Comma-Separated Values) son un formato común para almacenar datos textuales en los que la información se organiza en filas y columnas que separan cada valor mediante comas u otros separadores, lo que permite una

representación sencilla de datos estructurados y facilita el intercambio entre diferentes sistemas y herramientas de análisis de datos [18].

Frame

Un fotograma es una unidad única en una secuencia de video digital que representa una imagen estática que forma parte de una serie organizada en el tiempo, donde cada fotograma contiene información visual que puede procesarse de forma independiente para analizar cambios, detectar variaciones o identificar patrones, permitiendo el estudio de fenómenos dinámicos mediante el análisis continuo de imágenes secuenciales y facilitando el procesamiento de video a nivel computacional [23].

Intensidad de píxel

Se refiere a un valor numérico que representa el nivel de brillo de un punto específico en una imagen digital, y es un elemento fundamental del procesamiento de imágenes, ya que permite la cuantificación de información visual y el análisis de diferencias de brillo entre diferentes áreas, facilitando la detección de cambios, la identificación de patrones y la extracción de características de la imagen [24].

Escala de grises

La escala de grises es un tipo de representación de imágenes en la que cada píxel contiene información únicamente sobre la intensidad de la luz, eliminando componentes de color y reduciendo la complejidad de los datos, facilitando el procesamiento y análisis de imágenes, trabajando con una dimensión de información, permitiendo resaltar variaciones de intensidad y simplificando la aplicación de métodos computacionales para el análisis visual [24].

Procesamiento de video

El procesamiento de video consiste en analizar una secuencia de imágenes digitales ordenadas en el tiempo conocidas como fotogramas, cada una de las cuales representa una única unidad que puede procesarse de forma independiente para identificar cambios, variaciones o patrones en un fenómeno, permitiendo así

estudiar el comportamiento dinámico facilitando la extracción de información a partir de datos visuales, esenciales para aplicaciones de análisis que requieren imágenes [25].

Procesamiento digital de imágenes

El Procesamiento digital de imágenes incluye un conjunto de técnicas que tienen como objetivo mejorar, transformar y analizar imágenes para obtener información relevante, abarcando operaciones como filtrado, conversión de formato, ajuste de intensidad y eliminación de ruido, lo que resulta en datos más consistentes y adecuados para el análisis, estas técnicas se utilizan en este proyecto mediante procesos como conversión de escala de grises, normalización de valores y reducción de ruido [24].

Diferencia entre frames.

La diferencia entre frames es una técnica utilizada para detectar cambios entre cuadros sucesivos en una secuencia de video, donde los valores de intensidad de cada píxel se comparan entre cuadros sucesivos para resaltar áreas de variación, facilitando así la detección de cambios en un fenómeno dinámico y permitiendo el análisis de variaciones visuales en secuencias de tiempo [16].

Vision por computadora

Permite a los sistemas procesar e interpretar información visual a partir de imágenes o secuencias de video, con el objetivo de transformar datos visuales en información útil utilizando algoritmos que identifican patrones, formas o cambios en el entorno, utilizando técnicas que van desde el procesamiento básico de imágenes hasta el aprendizaje automático avanzado que permiten tareas como la detección de objetos, el reconocimiento de patrones de movimiento, como la investigación de banda ancha, la segmentación, la seguridad y la detección de aplicaciones industriales [1].

Segmentación de imágenes

La segmentación de imágenes es la división de una imagen en regiones con características similares con el fin de facilitar su análisis y permitir una mejor

interpretación de la información visual, pues al separar las áreas relevantes de aquellas que no agregan valor, es posible centrar la atención solo en las partes de interés, lo que ayuda a reducir la complejidad del procesamiento, mejorar la precisión del análisis y optimizar el uso de la organización de la información y el uso de recursos específicos [24].

Regiones de interés

Las regiones de interés, conocidas como ROI, son áreas de una determinada imagen o secuencia de vídeo que se seleccionan para centrar el análisis en las partes más relevantes, reduciendo así el procesamiento innecesario de información y mejorando la eficiencia del sistema, también facilitan la identificación de características importantes como bordes, formas o movimientos, que se utilizan ampliamente en aplicaciones de visión por computadora [26].

Extracción de características

La extracción de características es el proceso de obtención de valores numéricos que representan las propiedades relevantes de un conjunto de datos, lo que permite convertir información visual en variables cuantificables que pueden ser interpretadas por diversos algoritmos que facilitan el análisis, comparación e identificación de patrones de datos, ya que la calidad de estos patrones depende en gran medida de los modelos utilizados [27].

Visualización de datos

La visualización de datos le permite representar información gráficamente utilizando cuadros, gráficos y otras herramientas visuales que facilitan la interpretación de los resultados, de modo que pueda identificar de manera más clara y comprensible patrones, tendencias y relaciones en los datos, especialmente cuando se trabaja con una gran cantidad de información [28].

Plasma

El plasma es considerado el cuarto estado de la materia, caracterizado por la presencia de partículas ionizadas como electrones e iones libres, que ocurren

cuando un gas es sometido a suficiente energía que provoca la separación de sus componentes atómicos, y exhibe un comportamiento colectivo dominado por interacciones electromagnéticas que le permiten exhibir dinámicas complejas y variaciones temporales [6].

Dinamica del plasma

La dinámica del plasma estudia su comportamiento en el tiempo y el espacio, teniendo en cuenta variables como la energía, la densidad y las interacciones entre partículas cargadas, donde sus fluctuaciones pueden observarse como cambios de intensidad o distribución espacial, permitiendo interpretar patrones que facilitan la identificación del estado del sistema [29].

Variación de intensidad

La variación de intensidad corresponde a cambios en los niveles de brillo o valores de píxeles en una imagen o secuencia de video, que pueden ser causados por cambios en la iluminación, movimiento de objetos o transformaciones en la escena, lo cual es un elemento clave en el análisis de imágenes ya que permite identificar bordes, identificar regiones de diferentes propiedades y analizar patrones visuales, la investigación es importante en las aplicaciones de visión por computadora, donde es necesario interpretar cambios en la información visual y obtener características adecuadas para el procesamiento automático de datos [30].

Análisis temporal

El análisis temporal permite estudiar el comportamiento de un fenómeno a lo largo del tiempo, identificando patrones, tendencias y variaciones en las secuencias de datos que facilitan la comprensión de la evolución de los sistemas dinámicos mediante el estudio de eventos secuenciales, es fundamental en los sistemas de procesamiento de información porque nos permite interpretar cambios [31].

Machine learning

El machine learning es una rama de la inteligencia artificial que permite a los sistemas aprender de los datos y reconocer patrones sin programación especial para

cada situación, utilizando algoritmos que ajustan sus parámetros en función de la información disponible, permitiéndoles mejorar el rendimiento ante nuevos datos y adaptarse a diferentes problemas de análisis de información [32].

Random forest

Random Forest es un algoritmo de clasificación que combina varios árboles de decisión para generar un modelo más preciso, basado en la construcción de varios clasificadores independientes que operan sobre subconjuntos de datos y cuyas decisiones se integran mediante votación, lo que permite reducir la redundancia, mejorar la generalización del modelo y obtener resultados más estables frente a datos variables [33].

Clasificación binaria

La clasificación binaria es un tipo de problema de aprendizaje supervisado en el que el modelo debe asignar cada dato a una de dos categorías posibles en función de patrones obtenidos durante el entrenamiento a partir de datos preetiquetados, una técnica ampliamente utilizada en diversas áreas de la toma de decisiones automatizada porque permite distinguir entre dos estados bien definidos en función de propiedades derivadas de la información analizada [8].

Anomalía

Una anomalía es una observación, evento o patrón que se desvía significativamente del comportamiento esperado o normal dentro de un conjunto de datos, de tal manera que genera la sospecha de haber sido producido por un mecanismo diferente al que origina el resto de las observaciones [34]. Estas desviaciones pueden clasificarse como anomalías puntuales, cuando un único registro se aparta claramente del resto de las observaciones, cuando un valor resulta atípico únicamente dentro de un intervalo temporal o condición específica [35]

Z-score

El z-score es una medida estadística de cuántas desviaciones estándar tiene un valor de la media de un conjunto de datos, y es particularmente útil para comparar datos

que pertenecen a diferentes distribuciones o escalas, y se usa ampliamente para detectar anomalías porque puede identificar valores que se desvían significativamente del comportamiento esperado de la distribución [36].

Análisis estadístico

El análisis estadístico permite la interpretación de datos utilizando medidas descriptivas como media, varianza y desviación estándar, que facilitan resumir una gran cantidad de información en valores representativos, permitiendo identificar patrones, tendencias y comportamientos inusuales en un conjunto de datos que son útiles para una comprensión más clara de las características del fenómeno analizado [37].

Matriz de confusión

Una matriz de confusión es una herramienta utilizada para evaluar el desempeño de un modelo de clasificación, permitiendo comparar predicciones con valores reales usando una tabla que muestra los éxitos y fracasos del modelo en cada categoría, facilitando el cálculo de métricas como exactitud, recuperación y precisión, proporcionando una visión clara del comportamiento del sistema en diferentes escenarios y su capacidad para clasificar correctamente datos en diferentes escenarios [38].

2.2 Marco teórico

Aplicación de la visión por computadora en el análisis de fenómenos dinámicos

La visión por computadora se ha convertido en una herramienta clave en las tecnologías de la información y la comunicación, permitiendo a los sistemas interpretar información visual a partir de imágenes y secuencias de video mediante algoritmos especializados, facilitando la transformación de datos visuales en información estructurada, lo cual es particularmente útil en el estudio de fenómenos dinámicos con variaciones constantes en el tiempo, ya que permite detectar cambios, extraer patrones a partir de características visuales e identificarlos [1]. En este proyecto se utiliza para analizar secuencias de vídeo de plasma, permitiendo

identificar variaciones en su comportamiento, detectar cambios de intensidad y extraer características que serán utilizadas en el proceso de análisis y clasificación.

Procesamiento digital de imágenes

El procesamiento de imágenes digitales es la base del análisis de la información visual, ya que permite mejorar, transformar y preparar las imágenes mediante técnicas como la conversión de escala de grises, la reducción de ruido y el ajuste de intensidad, facilitando la adquisición de datos más consistentes y analíticos [24]. Este proyecto aplica estas técnicas al vídeo plasma para mejorar la calidad visual, reducir el ruido y resaltar las variaciones de intensidad, facilitando la extracción de la información necesaria para el análisis.

Machine Learning en la clasificación de datos:

El aprendizaje automático permite desarrollar modelos capaces de aprender de los datos y reconocer patrones sin programación explícita, utilizando algoritmos útiles para distinguir entre diferentes estados o comportamientos [22]. Se utiliza el aprendizaje automático para analizar las propiedades extraídas del plasma y clasificar su comportamiento, permitiendo identificar condiciones normales y posibles anomalías.

Enfoque basado en extracción de características

En los sistemas de análisis de imágenes, el enfoque basado en la extracción de características permite representar la información visual mediante variables numéricas que describen las propiedades relevantes del fenómeno observado, estas características pueden incluir mediciones estadísticas, espaciales y de intensidad que facilitan la transformación de datos visuales en estructuras que pueden analizarse mediante algoritmos computacionales [39]. Se aplica este enfoque obteniendo características como energía normalizada, desviación estándar, variaciones de intensidad y propiedades geométricas de las áreas expuestas, que permiten modelar el comportamiento del plasma y facilitan su posterior análisis mediante métodos de aprendizaje automático.

Análisis basado en regiones de interés

El análisis basado en regiones de interés (ROI) implica seleccionar regiones específicas en una secuencia de imagen o vídeo para centrar el procesamiento solo en las regiones relevantes, esto permite reducir el ruido de áreas irrelevantes y mejorar la precisión de la detección, el uso de múltiples roi permite analizar diferentes áreas del escenario de forma independiente, lo que facilita la localización de cambios o anomalías [40]. En este proyecto, el uso de roi permite dividir el área de análisis de plasma en múltiples regiones, lo que contribuye a una detección más precisa.

Análisis temporal y persistencia de eventos

El análisis temporal permite estudiar cómo evoluciona un fenómeno a partir de la información obtenida en diferentes instantes de tiempo, considerando la relación existente entre observaciones consecutivas [35]. En la detección de anomalías, este tipo de análisis ayuda a diferenciar cambios momentáneos de aquellos que representan un comportamiento persistente, disminuyendo la posibilidad de generar falsas alarmas por variaciones transitorias [41]. Con base en este principio, el sistema desarrollado incorpora un mecanismo de persistencia temporal que evalúa de forma continua el comportamiento de las regiones de interés y confirma una anomalía únicamente cuando esta se mantiene durante varios instantes consecutivos.

2.3. Metodología del proyecto

2.3.1. Metodología de Investigación

En este contexto, la investigación exploratoria permite identificar trabajos relacionados con el uso de visión artificial y aprendizaje automático para el análisis automatizado de información visual [42]. Estos antecedentes constituyen una base metodológica para el desarrollo de un sistema capaz de procesar videos de plasma, extraer características relevantes y detectar anomalías de manera automática, contribuyendo a una evaluación más eficiente y objetiva del fenómeno estudiado.

2.3.2. Metodología de desarrollo.

El proyecto se desarrollará utilizando el modelo incremental, que es un enfoque de desarrollo de software en el que el sistema se construye de forma incremental utilizando subversiones funcionales que incorporan nuevas características en cada iteración para que el sistema pueda dividirse en componentes o módulos que se desarrollan, prueban e integran continuamente, promoviendo la detección temprana de errores y reduciendo el riesgo en el proceso de desarrollo, este modelo favorece la inclusión de mejoras y nuevos requisitos sin necesidad de reconstruir completamente el sistema [43].

La metodología se organiza en cuatro fases principales: análisis, diseño, desarrollo y pruebas, como se muestra en la figura 1, permitiendo mantener un seguimiento adecuado del progreso del sistema y facilitando la incorporación de mejoras funcionales en cada iteración del desarrollo.

- ✓ Fase de análisis: En esta fase se identifican y definen los requerimientos funcionales y no funcionales del sistema, considerando las necesidades relacionadas con el análisis de secuencias de video de plasma, la automatización del procesamiento y la interacción del usuario mediante la interfaz gráfica. Además, se investigan técnicas de visión artificial, procesamiento digital de imágenes y aprendizaje automático que servirán como base para la extracción de características, construcción del dataset y funcionamiento general del sistema.
- ✓ Fase de diseño: En esta fase se establece la arquitectura general del sistema y la organización modular de la aplicación, definiendo el flujo de procesamiento desde la adquisición de datos hasta la visualización de resultados. También se diseña la interfaz gráfica de usuario, la estructura de almacenamiento de métricas, el manejo de regiones de interés (ROI) y la integración entre los módulos de procesamiento, análisis estadístico y machine learning.

- ✓ Fase de desarrollo: En esta fase se implementan progresivamente los diferentes módulos de la aplicación, incluyendo la GUI, carga y reproducción de secuencias experimentales, preprocesamiento de imágenes, segmentación mediante ROIs, extracción de características y generación automática del dataset. Asimismo, se integra el modelo Random Forest para la detección de anomalías, junto con herramientas de visualización, almacenamiento de métricas y generación de alertas visuales.
- ✓ Fase de pruebas: En esta fase se realizan pruebas funcionales, de integración y rendimiento para verificar el correcto funcionamiento de cada componente implementado. Se evalúa la estabilidad de la interfaz gráfica, la precisión del análisis visual, el desempeño del modelo de aprendizaje automático y la correcta representación de resultados mediante tablas, gráficas e indicadores visuales de anomalías.

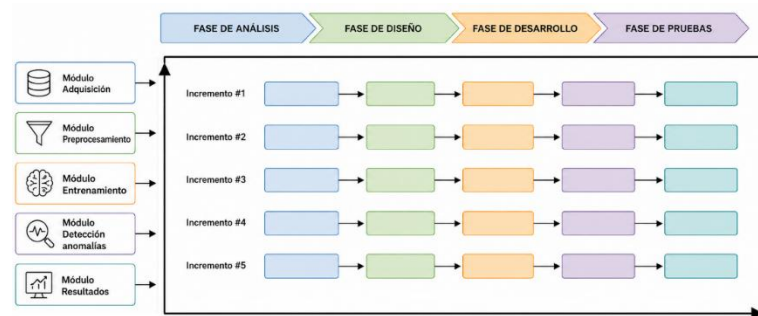


Figura 1: Metodología incremental

CAPÍTULO 3. PROPUESTA

3.1. Requerimientos

Para garantizar que la solución propuesta responda a los objetivos establecidos, resulta imprescindible definir de forma precisa las prestaciones que debe incorporar el sistema. Estas funcionalidades se estructuran en cinco módulos que cubren todo el flujo de trabajo, desde la carga de datos experimentales hasta la fase de aprendizaje del modelo. Las tablas siguientes recogen los requerimientos funcionales (RF) identificados para cada uno de estos bloques, los requerimientos no funcionales permiten que la plataforma funcione de manera estable y eficiente durante el análisis de los datos.

3.1.1. Requerimientos Funcionales

Módulo de Adquisición de datos

Código	Requerimiento Funcional
RF-01	El sistema permitirá cargar secuencias experimentales de plasma a partir de archivos en formato .cih.
RF-02	El sistema permitirá extraer automáticamente los fotogramas asociados a las secuencias experimentales cargadas.
RF-03	El sistema permitirá obtener metadatos de las secuencias experimentales, incluyendo resolución, FPS y duración de captura.
RF-04	El sistema verificará que los archivos cargados sean válidos y compatibles antes de iniciar el procesamiento.
RF-05	El sistema permitirá reducir la cantidad de fotogramas procesados para optimizar el rendimiento computacional.
RF-06	El sistema permitirá configurar el FPS objetivo a procesar.
RF-07	El sistema permite previsualizar el video mediante un slider para navegar entre fotogramas.
RF-08	El sistema habilitará o deshabilitará el avance al siguiente módulo según exista un video válido cargado
RF-09	El sistema mostrara un mensaje si no se seleccionó el archivo
RF-10	El sistema mostrara un mensaje si no se pudieron cargar los frames del archivo

Código	Requerimiento Funcional
RF-11	El sistema mostrara un mensaje si no se leen los metadatos
RF-12	Al seleccionar un archivo no valido el sistema mostrara que seleccione un archivo cih
RF-13	El sistema mostrara con la barra de progreso cuando el archivo ya este cargado
RF-14	El sistema permite al usuario reiniciar todos los campos de la pantalla a su estado inicial.
RF-15	El sistema permite avanzar a la siguiente pantalla únicamente si se ha cargado un archivo correctamente.

Tabla 1:Requerimiento Funcional – Módulo de Adquisición

Módulo de preprocesamiento

Código	Requerimiento Funcional
RF-16	El sistema muestra información general del video CIH cargado, incluyendo metadatos como resolución, FPS y duración.
RF-17	El sistema permite la selección manual de tres regiones de interés (ROI) sobre los frames del video.
RF-18	El sistema permite dibujar, visualizar y reiniciar las regiones de interés (ROI) seleccionadas.
RF-19	El sistema permite aplicar filtros de preprocesamiento (Gauss y CLAHE) para mejorar la calidad de las imágenes de forma opcional.
RF-20	El sistema muestra el resultado de los filtros aplicados en la vista de salida.
RF-21	El sistema ejecuta el procesamiento del video al presionar el botón correspondiente, calculando características de cada ROI en cada frame.
RF-22	El sistema muestra una tabla con las características extraídas, incluyendo frame, ROI, energía, Z-score, entropía y clustering.
RF-23	El sistema genera un dataset estructurado en formato CSV al finalizar el procesamiento y notifica su creación exitosa.
RF-24	El sistema muestra mensajes de estado durante el procesamiento, incluyendo errores, advertencias y confirmaciones.

Tabla 2: Requerimiento Funcional – Módulo de Procesameinto:

Módulo entrenamiento

Código	Requerimiento Funcional
RF-25	El sistema carga automáticamente el modelo anterior desde disco al ingresar a la pantalla (si existe)
RF-26	El sistema muestra la información del modelo anterior, incluyendo tamaño del dataset, número de características, total de anomalías y métricas generales.
RF-27	El sistema muestra las métricas de rendimiento del modelo (Accuracy, Precision, Recall y F1-Score).
RF-28	El sistema muestra la matriz de confusión del modelo entrenado.
RF-29	El sistema muestra la importancia de las características en una tabla ordenada
RF-30	El sistema permite iniciar el entrenamiento del modelo mediante un botón de acción.
RF-31	El sistema ejecuta el proceso completo de entrenamiento cuando el usuario lo solicita.
RF-32	El sistema deshabilita el botón de entrenamiento durante el proceso de ejecución.
RF-33	El sistema muestra el estado del entrenamiento.
RF-34	El sistema compara el desempeño del nuevo modelo con el modelo anterior.
RF-35	El sistema guarda automáticamente el nuevo modelo solo si su rendimiento (accuracy) es mayor al modelo anterior, y lo actualiza en el sistema.

Tabla 3: Requerimiento Funcional - Módulo de entrenamiento del modelo

Módulo de detección

Código	Requerimiento Funcional
RF-36	El sistema permite cargar archivos de plasma en formato .CIH para su análisis

Código	Requerimiento Funcional
RF-37	El sistema muestra la información del video cargado en el panel lateral.
RF-38	El sistema permite visualizar el video en reproducción durante el proceso de análisis.
RF-39	El sistema permite seleccionar hasta tres regiones de interés (ROI) sobre el video
RF-40	El sistema permite limpiar completamente las ROI y reiniciar el estado del sistema.
RF-41	El sistema valida la existencia de un video antes de iniciar el proceso de detección.
RF-42	El sistema permite iniciar, pausar y continuar el proceso de detección de anomalías
RF-43	El sistema ejecuta la detección de manera secuencial sobre los frames del video.
RF-44	El sistema permite ajustar parámetros de sensibilidad y velocidad de detección del modelo.
RF-45	El sistema muestra controles de configuración como sliders de sensibilidad y velocidad.
RF-46	El sistema informa al usuario sobre el impacto de la sensibilidad en la precisión de la detección.
RF-47	El sistema informa al usuario sobre el impacto de la sensibilidad en la detección.
RF-48	El sistema muestra el progreso de la detección, incluyendo el frame actual respecto al total.
RF-49	El sistema muestra la probabilidad de anomalía calculada por el modelo.
RF-50	El sistema muestra y almacena los eventos detectados con información como ROI, frame, tiempo, probabilidad de anomalía y tipo de evento.
RF-51	El sistema presenta los eventos detectados en una tabla de resultados y permite visualizar gráficas asociadas a cada ROI.

Tabla 4:Requerimiento Funcional - Módulo de Detección de anomalía

Visualización gráfica de la evolución temporal de las características extraídas.

Código	Requerimiento Funcional
RF-52	El sistema permite seleccionar una región de interés (ROI) para analizar su comportamiento durante el proceso de detección.
RF-53	El sistema muestra la evolución temporal de las características extraídas por ROI, incluyendo energía, Z-score, probabilidad de anomalía y variables asociadas.
RF-54	El sistema genera gráficas dinámicas actualizadas en tiempo real durante el proceso de detección.
RF-55	El sistema muestra la sensibilidad configurada por el usuario utilizada durante el proceso de detección de anomalías.
RF-56	El sistema ajusta el criterio de detección de anomalías de acuerdo con el nivel de sensibilidad seleccionado por el usuario.
RF-57	El sistema permite visualizar la influencia de la sensibilidad configurada en la clasificación de anomalías mediante la probabilidad generada por el modelo.
RF-58	El sistema guarda las gráficas asociadas a cada ROI dentro de la sesión de resultados.

Tabla 5: Requerimiento Funcional - Visualización de gráficos

Módulo de Resultados

Código	Requerimiento Funcional
RF-59	El sistema muestra las sesiones de análisis almacenadas previamente.
RF-60	El sistema permite seleccionar una sesión para consultar sus resultados.
RF-61	El sistema permite eliminar sesiones almacenadas del historial.
RF-62	El sistema muestra los eventos detectados en una tabla organizada por sesión.
RF-63	El sistema organiza los eventos en registros individuales asociados a cada detección.
RF-64	El sistema permite visualizar la imagen asociada al evento seleccionado.

Código	Requerimiento Funcional
RF-65	El sistema permite navegar entre las imágenes almacenadas dentro de una sesión.
RF-66	El sistema permite recorrer los eventos de una sesión mediante controles de navegación.
RF-67	El sistema actualiza dinámicamente la imagen y la información asociada al evento seleccionado.
RF-68	El sistema permite cambiar el tipo de gráfica visualizada para el análisis de eventos.
RF-69	El sistema permite exportar los resultados de una sesión en formato PDF.
RF-70	El sistema permite consultar resultados históricos sin necesidad de ejecutar nuevamente el proceso de detección.

Tabla 6: Requerimiento Funcional - Módulo de resultados

3.1.2. Requerimientos no Funcionales

Código	Requerimiento no funcional
RNF-01	El sistema permite proporcionar una interfaz gráfica intuitiva que facilite la navegación entre los módulos de adquisición, preprocesamiento, entrenamiento, detección y resultados.
RNF-02	El sistema permite procesar los datos de entrada y actualizar la información mostrada al usuario sin afectar significativamente la fluidez de la interfaz gráfica.
RNF-03	El sistema permite garantizar la correcta ejecución de los procesos de carga, preprocesamiento, entrenamiento y detección de anomalías.
RNF-04	El sistema permite preservar la consistencia de la información generada durante el procesamiento, entrenamiento y detección de anomalías.
RNF-05	El sistema permite estar organizado en módulos independientes para facilitar futuras modificaciones y correcciones.
RNF-06	El sistema debe permitir la incorporación de nuevas características, métricas o algoritmos de detección sin modificar significativamente la estructura general.

Código	Requerimiento no funcional
RNF-07	El sistema permite ejecutarse en equipos compatibles con Python y PyQt6.
RNF-08	El sistema permite permitir la ejecución continua de los módulos sin requerir reiniciar la aplicación durante el flujo normal de trabajo.
RNF-09	El sistema permite permitir la generación de reportes en formato PDF con los resultados obtenidos durante el análisis.
RNF-10	El sistema permite gestionar errores de entrada y condiciones inesperadas mediante mensajes informativos al usuario.

Tabla 7: Requerimientos no funcionales

3.2. Componente de la Propuesta

Comparación entre el enfoque tradicional y el sistema propuesto

Aspecto	Enfoque tradicional	Sistema propuesto
Método de detección	Métodos estadísticos basados en umbrales	Modelo de aprendizaje automático (Random Forest)
Criterio de decisión	Comparación de variables estadísticas con valores límite	Clasificación mediante un modelo entrenado
Uso de las variables estadísticas	Se utilizan para decidir si existe una anomalía	Se utilizan como características de entrada del modelo
Adaptabilidad	Limitada a reglas predefinidas	Aprende patrones a partir de los datos
Automatización	Parcial	Automática

Tabla 8: Comparación entre el enfoque tradicional y el sistema propuesto

En los métodos tradicionales, la detección de anomalías se realiza a partir del análisis de variables estadísticas y de la comparación de sus valores con umbrales definidos previamente. En el sistema desarrollado, estas variables no determinan

directamente la presencia de una anomalía, sino que forman parte del conjunto de características utilizado por el modelo Random Forest.

3.2.1. Arquitectura del Sistema

La arquitectura del sistema está compuesta por cinco módulos principales que permiten realizar el análisis de videos de plasma en formato .cih desde la carga de datos hasta la generación de resultados. Cada módulo ejecuta una etapa específica del proceso y entrega la información necesaria para la siguiente fase.

Los módulos que conforman el sistema son: adquisición de datos, preprocesamiento, entrenamiento, detección de anomalías y resultados.

Módulo de adquisición de datos

Este módulo permite cargar archivos en formato CIH y visualizar la información asociada al video seleccionado. Durante esta etapa se obtiene información como la resolución, cantidad de frames y frecuencia de captura. Además, el usuario puede definir el FPS que será utilizado durante el procesamiento.

Módulo de preprocesamiento

En este módulo se realiza la selección de las tres regiones de interés y la aplicación de filtros de imagen. Las características obtenidas son almacenadas en un conjunto de datos que posteriormente será utilizado para el entrenamiento del modelo y serán mostradas las características extraídas en la tabla, este módulo permite reiniciar la selección de regiones de interés en caso de requerir ajustes, así como visualizar de forma inmediata el resultado de los filtros

Módulo de entrenamiento

Este módulo utiliza el conjunto de datos generado durante el preprocesamiento para entrenar el modelo de clasificación usando el algoritmo de random forest, el cual permite identificar patrones asociados a comportamientos normales y anómalos del plasma. Una vez finalizado el entrenamiento se muestran las métricas de evaluación obtenidas del modelo.

Módulo de detección de anomalías

Este módulo utiliza el modelo previamente entrenado para analizar nuevas secuencias de plasma. Para cada ROI se calculan las características necesarias y se envían al modelo Random Forest, el cual determina si el comportamiento corresponde a una condición normal o a una anomalía. Los eventos detectados son registrados y presentados al usuario junto con su probabilidad. Adicionalmente, el sistema genera gráficas de energía absoluta, energía normalizada, z-score y probabilidad de anomalía para cada región de interés, permitiendo visualizar la evolución temporal del comportamiento del plasma durante el proceso de detección (ver [Anexo 10](#), [Anexo 11](#), [Anexo 12](#) y [Anexo 13](#)).

Módulo de resultados

Este módulo reúne la información generada durante el proceso de análisis y la presenta de forma organizada para facilitar su interpretación. Entre los resultados disponibles se incluyen los eventos detectados, las gráficas de comportamiento de cada ROI, las probabilidades de anomalía calculadas por el modelo y la generación de un reporte en formato PDF (ver [Anexo 15](#)).

La Figura 2 presenta la arquitectura general del sistema propuesto. En ella se observa la organización de los módulos que intervienen en el procesamiento de los videos de plasma, iniciando con la adquisición de los datos, seguida del preprocesamiento, el entrenamiento del modelo de clasificación, la detección de anomalías, la presentación de los resultados y generación de pdf.

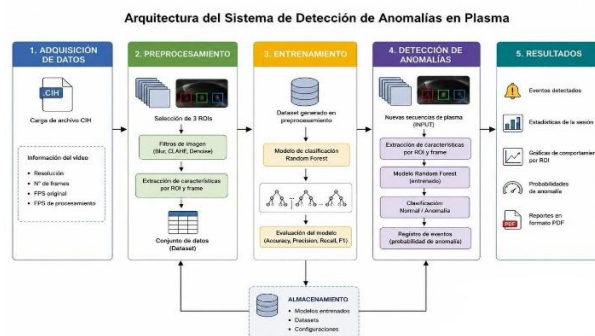


Figura 2: Arquitectura del sistema

3.2.2. Diagrama

Diagrama de casos de uso

La Figura 3 presenta el diagrama de casos de uso del sistema desarrollado. En este diagrama se observa que el usuario interactúa con el sistema mediante tres funcionalidades principales: entrenar el modelo, detectar anomalías y consultar los resultados obtenidos. Durante el proceso de detección de anomalías, el sistema ejecuta de forma automática las etapas de adquisición de datos, preprocesamiento, extracción de características y clasificación de las regiones de interés (ROI), las cuales forman parte del flujo interno del proceso. Finalmente, el usuario puede visualizar los resultados obtenidos y generar un reporte en formato PDF para su posterior análisis o almacenamiento.

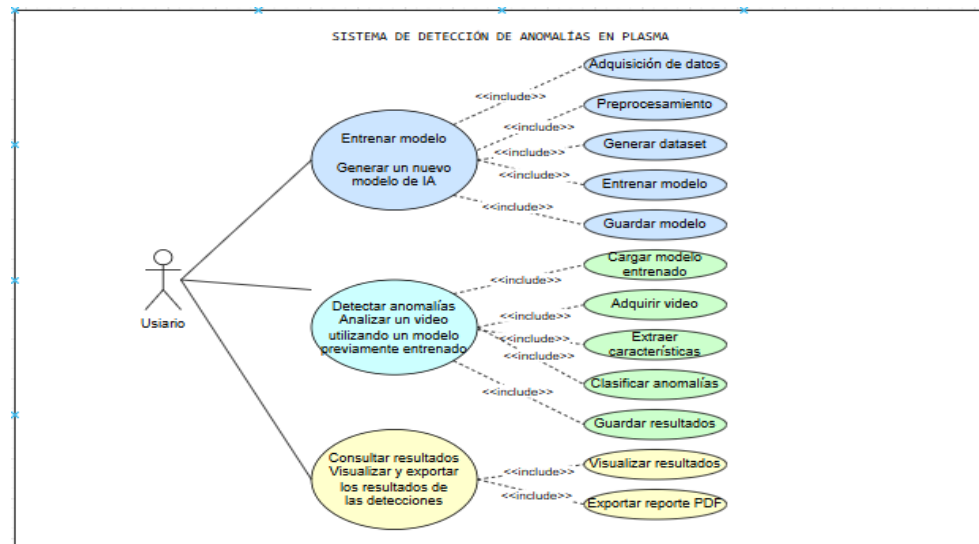


Figura 3: Caso de uso del sistema

Caso de Uso 01 – Entrenar modelo

Descripción

Este caso de uso permite al usuario entrenar un modelo de detección de anomalías a partir de un archivo de video CIH. El sistema realiza la adquisición de datos, el preprocesamiento mediante la selección de las regiones de interés (ROI), genera y etiqueta el conjunto de datos, entrena el modelo de aprendizaje automático y lo almacena para su posterior utilización en la detección de anomalías.

Actor	Usuario
Evento desencadenador	El usuario selecciona la opción Entrenar modelo.
Precondición	El usuario dispone de un archivo de video CIH válido. El sistema se encuentra operativo
Secuencia principal 1. El usuario accede a la opción Entrenar modelo. 2. El sistema solicita seleccionar un archivo de video CIH. 3. El usuario selecciona el archivo de video. 4. El sistema realiza la adquisición de los datos del video. 5. El sistema permite al usuario seleccionar las tres regiones de interés (ROI). 6. El sistema realiza el preprocesamiento de las imágenes utilizando las regiones seleccionadas. 7. El sistema genera y etiqueta automáticamente el conjunto de datos. 8. El usuario inicia el entrenamiento del modelo. 9. El sistema entrena el modelo de aprendizaje automático utilizando el conjunto de datos generado. 10. El sistema almacena el modelo entrenado. 11. El sistema informa que el entrenamiento ha finalizado correctamente.	
Errores / Alternativas ER-01. El archivo seleccionado no corresponde a un formato CIH válido. ER-02. No fue posible leer el archivo de video. ER-03. No se seleccionaron correctamente las tres regiones de interés (ROI). ER-04. Se produjo un error durante el preprocesamiento o la generación del conjunto de datos. ER-05. Ocurrió un error durante el entrenamiento del modelo de aprendizaje automático. ER-06. No fue posible almacenar el modelo entrenado.	
Postcondición El sistema genera y almacena un modelo de aprendizaje automático entrenado, el cual queda disponible para ser utilizado posteriormente en el proceso de detección de anomalías.	

Tabla 9: Entrenar modelo

Caso de Uso 02 – Detectar anomalías	
Descripción Este caso de uso permite al usuario detectar anomalías en un video CIH utilizando un modelo de aprendizaje automático basado en Random Forest previamente entrenado. El sistema carga el modelo, procesa el video, extrae las características de las regiones de interés (ROI), clasifica cada ROI como normal o con anomalía	
Actor	Usuario
Evento desencadenador	El usuario selecciona la opción Detectar anomalías, si lo requiere ajustar la sensibilidad del modelo o velocidad de detección
Precondición	Existe al menos un modelo previamente entrenado.
Secuencia principal • El usuario accede a la opción Detectar anomalías. • El sistema carga el modelo Random Forest previamente entrenado. • El sistema solicita seleccionar un archivo de video CIH. • El usuario selecciona el archivo y confirma la carga. • El sistema procesa el video y extrae las regiones de interés (ROI). • El sistema calcula las características estadísticas y dinámicas de cada ROI. • El sistema utiliza el modelo Random Forest para clasificar cada ROI como normal o con anomalía. • El sistema registra las anomalías detectadas. • El sistema almacena los resultados de la detección. • El sistema muestra los resultados al usuario.	
Errores / Alternativas ER-01. El archivo seleccionado no corresponde a un formato de video CIH válido. ER-02. El archivo de video está dañado o no puede ser leído. ER-03. No existe un modelo entrenado disponible para realizar la detección. ER-04. No fue posible almacenar los resultados de la detección.	
Postcondición Los resultados de la detección de anomalías quedan almacenados en el sistema y disponibles para su consulta en el módulo resultados.	

Tabla 10: Detectar anomalías

Caso de Uso 03 –Resultados	
Descripción	
Este caso de uso describe cómo el usuario visualiza los resultados obtenidos durante la detección de anomalías y si el usuario lo requiere genera un reporte en formato PDF.	
Actor	Usuario
Evento desencadenador	El usuario selecciona la opción resultados.
Precondición	Deben existir las sesiones guardadas por el proceso de detección de anomalías.
Secuencia principal 1. El usuario accede a la opción resultados. 2. El sistema recupera el historial de resultados almacenados. 3. El sistema muestra la lista de detecciones disponibles. 4. El usuario selecciona un resultado para visualizar su información. 5. El sistema presenta los detalles de la detección, incluyendo las anomalías identificadas y las gráficas de análisis (energía, z-score y demás métricas calculadas). 6. El usuario revisa los resultados obtenidos. 7. De forma opcional, el usuario solicita la generación del reporte. 8. El sistema genera el documento con el resumen de la detección, las gráficas y la información relevante. 9. El sistema permite exportar el reporte en formato PDF.	
Errores / Alternativas ER-01. No existen resultados almacenados para mostrar. ER-02. Se produjo un error al recuperar la información de los resultados. ER-03. Se produjo un error durante la generación del reporte PDF. ER-04. No fue posible guardar el archivo PDF en la ubicación seleccionada. ER-05. No fue posible cargar una o más imágenes o gráficas asociadas al resultado.	
Postcondición Los resultados son visualizados correctamente y si el usuario lo solicita exporta reporte en formato PDF.	

Tabla 11: Resultados

3.2.3. Modelado de datos

Creación del dataset

El conjunto de datos utilizado en este trabajo se construye a partir del procesamiento de videos de plasma, donde cada imagen es analizada frame por frame mediante regiones de interés (ROIs). A partir de cada una de estas regiones se extraen distintas características que permiten describir el comportamiento del sistema en diferentes instantes de tiempo.

Cada registro del dataset corresponde a una ROI en un frame específico, el cual se representa mediante un conjunto de variables numéricas que incluyen medidas de energía, estadísticas de intensidad y características relacionadas con la distribución espacial de la señal. Estas variables permiten capturar tanto la dinámica del plasma como sus variaciones en el tiempo.

Adicionalmente, se incorporan métricas derivadas del comportamiento temporal, como el cambio entre frames y la desviación respecto al comportamiento histórico. Finalmente, cada muestra es etiquetada como 45 normal o anomalía en función de las reglas establecidas en el sistema de detección.

Organización del dataset

El dataset se organiza en formato tabular, donde cada fila representa una observación correspondiente a una región de interés en un frame específico de un video, datos incluye información de identificación, como el video y el frame, así como las características extraídas durante el procesamiento. Entre ellas se encuentran variables relacionadas con energía, estadísticas de intensidad, actividad en la imagen, y medidas de segmentación como el número de clusters y el área de las regiones detectadas.

También se incluyen variables temporales que describen la evolución del comportamiento, como el z-score, el delta de energía y la velocidad de cambio. La última columna corresponde a la etiqueta de clase, que indica si la observación pertenece a un comportamiento normal o anómalo.

Variables del dataset

El dataset está conformado por variables de identificación, características extraídas de las regiones de interés, métricas temporales y la variable objetivo a utilizada durante el entrenamiento de modelo, cada grupo de variable aporta información específica sobre el comportamiento del plasma durante el procesamiento de los videos.

Variables de Identificación utilizadas para identificar el video, el frame procesado y la región de interés correspondiente.

Características	Descripción
video_id	Identificador del video procesado.
frame	Número del frame analizado dentro del video.
roi_id	Identificador de la región de interés (ROI).

Tabla 12: Variables de Identificación

Variables de energía: que describen el comportamiento energético del plasma y las variaciones entre frames.

Características	Descripción
energia_cambio	Energía calculada a partir de las diferencias entre frames consecutivos.
energia_absoluta	Energía total calculada en la región de interés.
energia_norm	Energía normalizada de la región de interés para facilitar su comparación.
densidad_energia	Relación entre la energía calculada y el área de la región de interés.

Tabla 13: Variables de energía

Variables estadísticas: obtenidas a partir de los valores de intensidad de cada ROI.

Características	Descripción
media	Intensidad promedio de los píxeles de la región de interés.
std	Desviación estándar de la intensidad de los píxeles.
minimo	Valor mínimo de intensidad registrado en la región de interés
maximo	Valor máximo de intensidad registrado en la región de interés
rango	Diferencia entre la intensidad máxima y mínima.
Varianza	Medida de la dispersión de las intensidades respecto a la media.
entropia	Medida de la complejidad o variabilidad de la distribución de intensidades.
contraste	Medida de la variación local de las intensidades dentro de la región de interés.

Tabla 14: Variables estadísticas

Variables de actividad: representan la cantidad y proporción de píxeles con actividad dentro de la región analizada

pixeles_activos	Cantidad de píxeles con actividad detectada en la región de interés.
porcentaje_activos	Proporción de píxeles activos respecto al total de píxeles de la región de interés.

Tabla 15: Variables de actividad

Variables de agrupamiento (clusters): describen el tamaño, cantidad y distribución de las regiones detectadas tanto en la imagen original como en el mapa de diferencias entre frames.

Características	Descripción
num_clusters_img	Número de regiones (clusters) detectadas en la imagen original.
area_cluster_total_img	Área total ocupada por los clusters detectados en la imagen original.
area_cluster_max_img	Área del cluster de mayor tamaño en la imagen original.
ratio_area_img	Relación entre el área ocupada por los clusters y el área total de la región de interés.
densidad_clusters_img	Densidad de clusters presentes en la imagen original.
num_clusters_cambio	Número de clusters detectados en el mapa de diferencias entre frames.
area_cluster_total_cambio	Área total ocupada por los clusters detectados en el mapa de diferencias entre frames.
area_cluster_max_cambio	Área del cluster de mayor tamaño detectado en el mapa de diferencias entre frames.

Tabla 16: Variables de agrupamiento

Variables temporales: utilizadas para describir la evolución temporal del comportamiento energético del plasma.

Características	Descripción
z_score	Desviación estandarizada de la energía respecto al comportamiento histórico de la región de interés.
delta_energia	Variación de la energía entre frames consecutivos.
velocidad_cambio	Tasa de cambio de la energía a lo largo del tiempo.

Tabla 17: Variables temporales

Variable objetivo: utilizada como etiqueta de clasificación, donde el valor 0 representa comportamiento normal y el valor 1 indica la presencia de una anomalía.

anomalía	Etiqueta de clasificación utilizada durante el entrenamiento del modelo, donde 0 representa un comportamiento normal y 1 indica la presencia de una anomalía.
----------	---

Tabla 18: Variable objetivo

Generación de la variable objetivo

Para entrenar el modelo de aprendizaje automático, a cada registro del conjunto de datos se le debía asignar una etiqueta. Dado que los videos de plasma no tenían una clasificación previa, la variable objetivo de anomalía se generó automáticamente utilizando un conjunto de reglas definidas para las características de cada región de interés (ROI).

Las reglas se desarrollaron teniendo en cuenta indicadores relacionados con el comportamiento del plasma, como cambios de energía, nivel de actividad, distribución de regiones expuestas y mediciones de tiempo. Los umbrales se seleccionaron mediante pruebas preliminares y un proceso de ajuste iterativo que analizó los valores obtenidos durante el procesamiento del video hasta obtener una configuración que permitiera una discriminación suficiente entre el comportamiento normal y anormal.

La Tabla 15 presenta las reglas y los umbrales utilizados para generar la variable objetivo. Cada registro es evaluado con base en estas condiciones; si al menos una de ellas se cumple, se asigna el valor 1 (anomalía), mientras que, si ninguna condición se cumple, se asigna el valor 0 (normal). Esta etiqueta es utilizada posteriormente durante el entrenamiento y la evaluación del modelo.

Regla	Condición	Descripción
z-score	$ z\text{-score} > 3$	Detecta desviaciones importantes respecto al comportamiento habitual de la región de interés.
Cambio de energía	$ \Delta E > 1\,000\,000$	Identifica cambios bruscos de energía entre dos frames consecutivos.
Clusters	≥ 15 clusters y área > 500	Detecta una cantidad elevada de regiones con cambios significativos entre imágenes consecutivas.
Clusters en imagen	≥ 8 clusters o área > 390	Evalúa la cantidad y el tamaño de las regiones activas presentes en la imagen.
Actividad	$< 3\%$ o $> 50\%$	Detecta niveles de actividad inferiores o superiores al comportamiento esperado.

Regla	Condición	Descripción
Energía Absoluta	$< 2\,000\,000$ o $> 80\,000\,000$	Verifica si la energía total de la región se encuentra fuera del rango considerado normal.

Tabla 19: Reglas utilizadas para la generación de la variable objetivo.

3.2.4. Entrenamiento y evaluación del modelo

El entrenamiento del modelo se realizó utilizando el dataset obtenido a partir del procesamiento de los videos de plasma. El conjunto de datos estuvo conformado por 126 000 registros, generados a partir de seis videos, donde cada registro corresponde a una región de interés analizada en un frame específico.

El modelo Random Forest fue entrenado utilizando una configuración de 200 árboles de decisión y una profundidad máxima de ocho niveles. Además, se ajustaron parámetros para limitar el crecimiento de los árboles y reducir el sobreajuste. Considerando el desbalance existente entre las clases, se empleó un ajuste de pesos balanceados durante el entrenamiento. Asimismo, se definió una semilla de ejecución (`random_state = 42`) para que el entrenamiento pudiera reproducirse bajo las mismas condiciones.

Métricas de evaluación

Para evaluar el desempeño del modelo se utilizaron las métricas Accuracy, Precision, Recall y F1-score, junto con la matriz de confusión, siguiendo los criterios estadísticos de evaluación de modelos de aprendizaje supervisado [42]. Estas métricas permitieron medir la capacidad del modelo para clasificar correctamente los registros del dataset e identificar las anomalías presentes en los videos de plasma.

Experimento 1 (17 características)

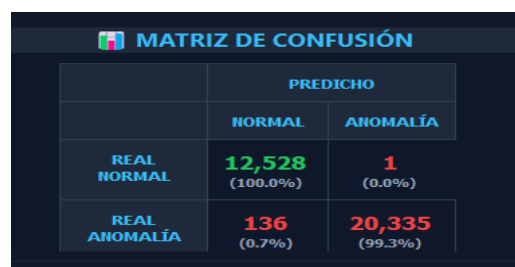
Muestras del experimento

En este experimento se utilizó el conjunto de datos descrito anteriormente, considerando únicamente 17 características seleccionadas del dataset. El objetivo

fue evaluar el desempeño del modelo con un conjunto reducido de variables y establecer una línea base para compararlo con el segundo experimento. Se buscó comprobar que el modelo no dependiera únicamente de las variables asociadas a la energía, sino que fuera capaz de identificar anomalías a partir de otras características del comportamiento del plasma.

Matriz de confusión

La Figura 4 presenta la matriz de confusión obtenida en la etapa de evaluación. Se observa que 12 528 registros normales y 20 335 registros anómalos fueron clasificados correctamente. Asimismo, se registró 1 falso positivo, correspondiente a un registro normal identificado como anomalía, y 136 falsos negativos, que representan anomalías clasificadas como normales. Estos resultados se reflejan en un Accuracy de 99,58 %, un Recall de 99,34 % y un F1-score de 99,66 %, lo que evidencia un buen comportamiento del modelo al distinguir entre ambas clases.



MATRIZ DE CONFUSIÓN		
	PREDICHO	
	NORMAL	ANOMALÍA
REAL NORMAL	12,528 (100.0%)	1 (0.0%)
REAL ANOMALÍA	136 (0.7%)	20,335 (99.3%)

Figura 4: Matriz de confusión del experimento 17 características

Características más importantes

La Figura 5 muestra las diez características con mayor importancia obtenidas durante el entrenamiento. Entre ellas destacan velocidad_cambio (0,4694) y energia_cambio (0,2793), seguidas por densidad_energia (0,0497) y energia_norm (0,0476). También aparecen variables como media, contraste, entropia, std, varianza y pixeles_activos, aunque con una contribución menor. En conjunto, estos resultados indican que las variables relacionadas con los cambios de energía y el comportamiento dinámico del plasma fueron las que tuvieron mayor influencia en la clasificación de los registros.

Características importantes	
Feature	Importancia
Velocidad de cambio	0.4694 (46.9%)
Tasa de cambio energético	0.2793 (27.9%)
Densidad de energía	0.0497 (5.0%)

Figura 5: Características más importantes del experimento 17 características

Experimento 2 (25 características)

Muestras del experimento

En el segundo experimento se utilizó el mismo conjunto de datos empleado en la prueba anterior, con la diferencia de que el modelo fue entrenado utilizando 25 características extraídas de cada registro. El propósito de esta prueba fue evaluar si la incorporación de nuevas variables permitía mejorar la capacidad del modelo para clasificar los registros normales y las anomalías.

Matriz de confusión

La Figura 6 muestra la matriz de confusión obtenida durante la evaluación del modelo. Se observa que 12 528 registros normales y 20 444 registros correspondientes a anomalías fueron clasificados correctamente. Además, se obtuvo 1 falso positivo y 27 falsos negativos, lo que representa una disminución en los errores de clasificación con respecto al experimento realizado con 17 características. Estos resultados se reflejan en un Accuracy de 99,92 %, un Recall de 99,87 % y un F1-score de 99,93 %, evidenciando una mejora en el desempeño del modelo.

MATRIZ DE CONFUSIÓN		
	PREDICHO	
	NORMAL	ANOMALÍA
REAL NORMAL	12,528 (100.0%)	1 (0.0%)
REAL ANOMALÍA	27 (0.1%)	20,444 (99.9%)

Figura 6: Matriz de confusión del experimento 25 características

Características más importantes

La Figura 7 presenta las diez características con mayor importancia obtenidas durante el entrenamiento. Las variables `velocidad_cambio`, `delta_energia` y `energia_cambio` fueron las que tuvieron mayor influencia en el proceso de clasificación. También destacaron `energia_norm`, `densidad_energia`, `energia_absoluta`, `media`, `z_score`, `contraste` y `entropia`, aunque con una contribución menor. En comparación con el primer experimento, la incorporación de estas características permitió obtener una clasificación más precisa y reducir el número de errores durante la evaluación.



The screenshot shows a dark-themed window titled "Características importantes" with a red star icon. It contains a table with two columns: "Feature" and "Importancia". The table lists three features with their respective importance scores and percentages.

Feature	Importancia
Velocidad de cambio	0.3177 (31.8%)
Delta de energía	0.2903 (29.0%)
Tasa de cambio energético	0.1699 (17.0%)

Figura 7: Características más importantes del experimento 25 características

3.3. Diseño de interfaz

Pantalla de importación de datos

En la Figura 8 se presenta la pantalla correspondiente a la carga del video en formato CIH que será utilizado como entrada para el sistema. Una vez seleccionado el archivo, la aplicación realiza la lectura de sus metadatos y muestra información relevante, como la ruta de almacenamiento, la resolución del video, el número total de fotogramas y la velocidad de captura expresada en fotogramas por segundo (FPS). Estos datos permiten al usuario verificar que el archivo cargado corresponde al video que será procesado. Adicionalmente, la interfaz permite configurar el número de FPS que se empleará durante el procesamiento, lo que brinda la posibilidad de adaptar el análisis a las necesidades del experimento y optimizar el rendimiento del sistema antes de continuar con las siguientes etapas del procesamiento y extracción de características.

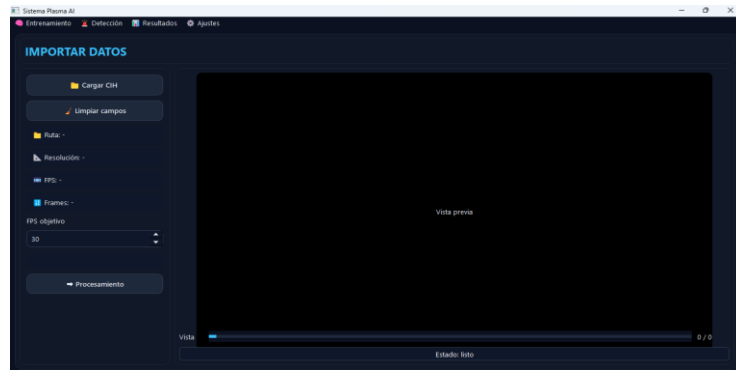


Figura 8: Pantalla de importación de datos

Pantalla de preprocesamiento

La Figura 9 muestra la pantalla donde se lleva a cabo el preprocesamiento del video. En esta etapa es posible seleccionar las regiones de interés (ROI), aplicar los filtros disponibles y ejecutar el proceso de extracción de características. Además, se presenta una tabla con los datos obtenidos de cada ROI, los cuales posteriormente son utilizados para el entrenamiento del modelo.

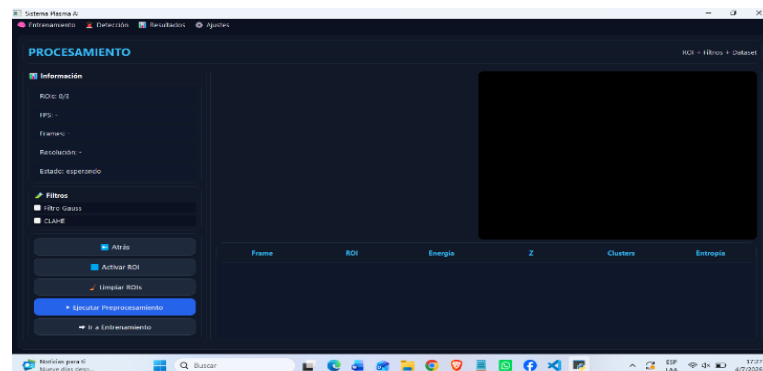


Figura 9: Pantalla de preprocesamiento

Pantalla de entrenamiento

La Figura 10 corresponde a la pantalla de entrenamiento del modelo. Desde esta interfaz se inicia el entrenamiento del algoritmo Random Forest y, una vez finalizado, se muestran las métricas obtenidas, la matriz de confusión y la importancia de las características utilizadas por el modelo. Esta información permite verificar el desempeño alcanzado antes de utilizar el modelo para la detección de anomalías.

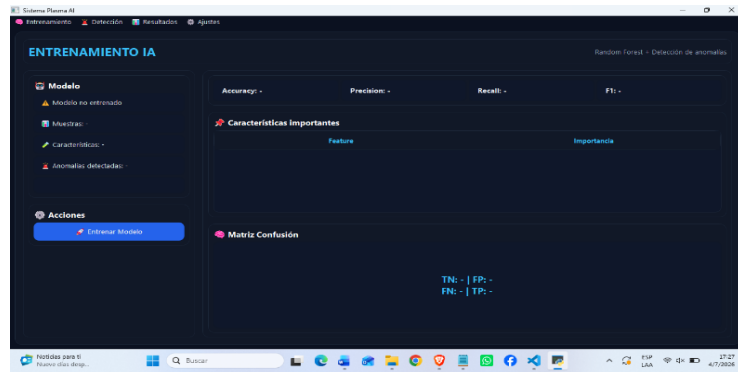


Figura 10: Pantalla de entrenamiento

Pantalla de detección de anomalías

La Figura 11 muestra la pantalla utilizada para analizar los videos mediante el modelo de detección de anomalías. Desde esta interfaz el usuario puede cargar el video, seleccionar la región de interés e iniciar el análisis. A medida que el video es procesado, el sistema muestra el estado obtenido para cada ROI, la probabilidad de anomalía y el de evento detectado. Se presenta un registro con los eventos encontrados y opciones para ajustar la sensibilidad del modelo y la velocidad de detección de acuerdo con los requerimientos del usuario., además al finalizar se muestra la sesión que se crea automáticamente (ver [Anexo 14](#)).

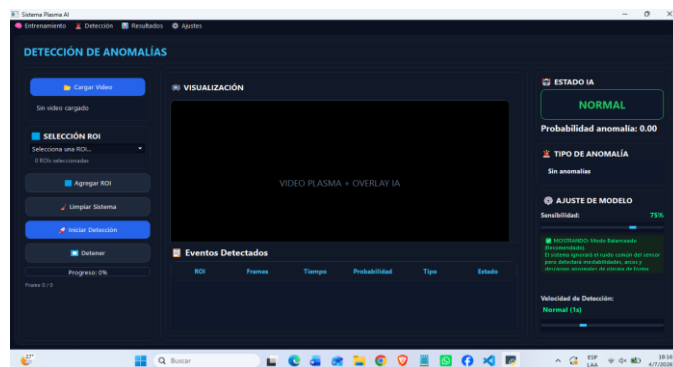


Figura 11: Pantalla de detección de anomalías

Pantalla Resultados

En la Figura 12 permite visualizar y gestionar los eventos detectados durante el análisis de las muestras de plasma. Presenta un listado de sesiones procesadas, una

tabla con los eventos detectados y un panel de visualización donde se muestran la imagen y la gráfica correspondientes al evento seleccionado. Además, permite visualizar las gráficas generadas durante el análisis para facilitar la interpretación de los resultados.

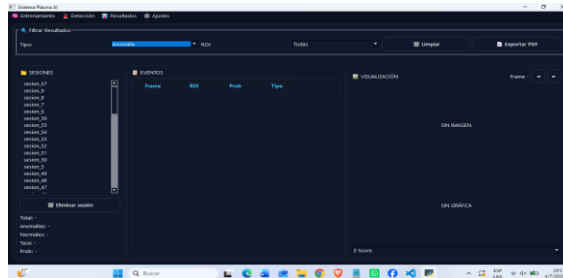


Figura 12: Pantalla Resultados

3.4. Pruebas

CASO DE PRUEBA	CP-001
Módulo	Importación del video
Objetivo	Verificar la correcta importación de un video CIH.
Condiciones: El sistema se encuentra operativo. El usuario dispone de un archivo de video CIH válido.	Procedimiento Seleccionar un archivo de video CIH. Cargar el archivo en el sistema. Verificar que se muestre la información del video.
RESULTADOS	
Resultado esperado	
El sistema carga el video correctamente y presenta la resolución, la velocidad de captura (FPS), el número de fotogramas y la ruta del archivo.	
Resultado obtenido	
El video fue cargado correctamente y el sistema mostró la información correspondiente, permitiendo continuar con la siguiente etapa del procesamiento como se muestra en la figura 13 (ver también Anexo 1, Anexo 2 y Anexo 3, referentes al proceso de carga de archivos).	

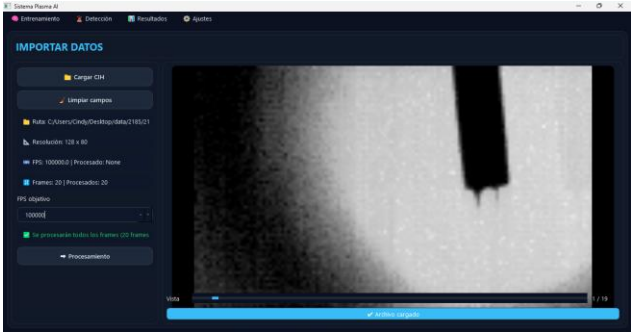
CASO DE PRUEBA	CP-002
 Figura 13: Resultado de la prueba de importación del video	
Evaluación:	☒ Exitoso ☐ Fallido

Table 20: Caso de prueba importación del video

CASO DE PRUEBA	CP-002
Módulo	Preprocesamiento
Objetivo	Verificar que el sistema realice correctamente el preprocesamiento del video y la extracción de características.
Condiciones: El video fue importado correctamente. Se seleccionaron las regiones de interés.	Procedimiento Seleccionar las regiones de interés. Ejecutar el preprocesamiento. Iniciar la extracción de características. Verificar la generación de los datos obtenidos.
RESULTADOS	
Resultado esperado	El sistema procesa las regiones de interés y genera correctamente las características que serán utilizadas durante el entrenamiento del modelo.
Resultado obtenido	El sistema ejecutó el preprocesamiento y generó las características correspondientes para cada región de interés, las cuales fueron almacenadas para la etapa de entrenamiento (ver Anexo 4 , extracción de características).

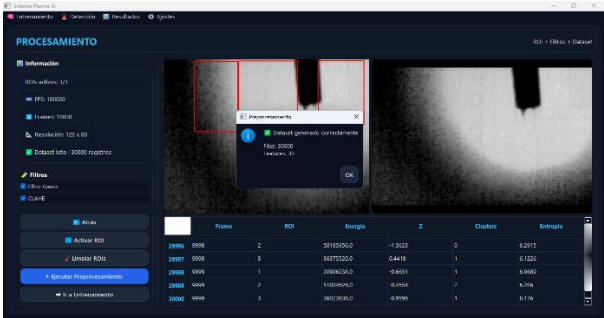
CASO DE PRUEBA	CP-002
 Figura 14: Resultado de la prueba de preprocesamiento	
Evaluación:	☒ Exitoso ☐ Fallido

Tabla 21: Caso de prueba preprocesamiento

CASO DE PRUEBA	CP-003
Módulo	Entrenamiento del modelo
Objetivo	Verificar que el sistema entrene correctamente el modelo de aprendizaje automático.
Condiciones: Existe un conjunto de datos generado durante el preprocesamiento.	Procedimiento Iniciar el entrenamiento del modelo. Esperar la finalización del proceso. Verificar las métricas de evaluación obtenidas.
RESULTADOS	
Resultado esperado	El sistema completa el entrenamiento y muestra las métricas de evaluación del modelo.
Resultado obtenido	El entrenamiento finalizó correctamente y el sistema presentó las métricas de evaluación, la matriz de confusión y la importancia de las características utilizadas por el modelo (ver Anexo 5).

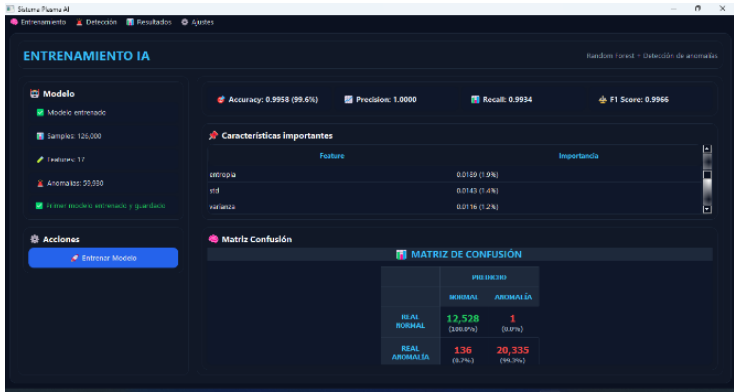
CASO DE PRUEBA	CP-003
Resultado obtenido	
 Figura 15: Resultado de la prueba entrenamiento del modelo	
Evaluación	☒ Exitoso ☐ Fallido

Tabla 22: Caso de prueba entrenamiento del modelo

CASO DE PRUEBA	CP-004
Módulo	Detección de anomalías
Objetivo	Verificar que el sistema identifique correctamente las anomalías presentes en un video utilizando el modelo entrenado.
Condiciones: Existe un modelo previamente entrenado. Se dispone de un video para el análisis.	Procedimiento Cargar el video. Seleccionar la región de interés. Ejecutar la detección. Verificar los eventos detectados.
RESULTADOS	
Resultado esperado El sistema clasifica las regiones de interés y registra las anomalías detectadas durante el análisis	
. Resultado obtenido El sistema procesó el video correctamente, identificó las anomalías presentes y registró los eventos detectados junto con su probabilidad (ver Anexo 6 , Anexo 7 , Anexo 8 y Anexo 9).	

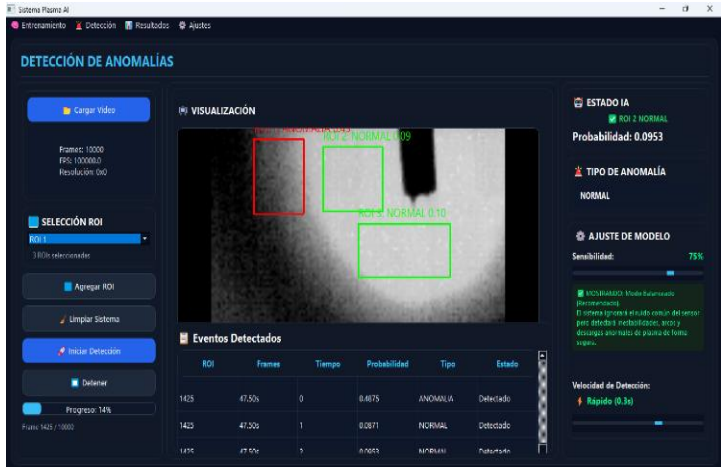
CASO DE PRUEBA	CP-004
Resultado obtenido	
 Figura 16: Resultado de la prueba detección de anomalías	
Evaluación:	☒ Exitoso ☐ Fallido

Tabla 23: Caso de prueba detección de anomalías

CASO DE PRUEBA	CP-005
Módulo	Generación del reporte
Objetivo	Verificar que el sistema genere correctamente el reporte en formato PDF con la información obtenida durante el análisis.
Condiciones: Existe una sesión con resultados almacenados.	Procedimiento Seleccionar una sesión registrada. Elegir la opción Exportar PDF. Verificar la creación del reporte.
RESULTADOS	
Resultado esperado El sistema genera correctamente el reporte en formato PDF con la información de la sesión seleccionada.	

Resultado obtenido

El sistema generó el reporte correctamente, incluyendo las estadísticas, los eventos detectados y las imágenes correspondientes al análisis realizado (ver [Anexo 16](#)).



Figura 17: Resultado de la prueba de Generación del reporte

Evaluación:

☒ Exitoso ☐ Fallido

Table 24: Caso de prueba de Generación del reporte

CONCLUSIONES

Se desarrolló una interfaz gráfica que integra técnicas de visión artificial, análisis estadístico y aprendizaje automático para el análisis de secuencias de video de plasma. El sistema permite procesar automáticamente los videos, extraer información de las regiones de interés y detectar anomalías mediante un modelo de aprendizaje automático, cumpliendo con el objetivo general planteado.

Se diseñó una interfaz gráfica que integra los principales módulos del sistema, permitiendo cargar videos, visualizar los cuadros originales y procesados, seleccionar las regiones de interés y controlar los parámetros necesarios para el procesamiento. La organización de estos elementos dentro de una misma aplicación facilitó la interacción del usuario con el sistema y permitió realizar el análisis de forma ordenada.

La integración de las librerías de procesamiento de imágenes y aprendizaje automático dentro de la interfaz gráfica permitió automatizar el flujo completo de trabajo, desde la carga del video hasta la generación de los resultados. Esta integración eliminó la necesidad de utilizar herramientas independientes y simplificó el proceso de análisis para el usuario.

La implementación del método de segmentación mediante regiones de interés permitió centrar el análisis únicamente en las zonas seleccionadas por el usuario. Esto facilitó la extracción de características representativas del comportamiento del plasma y contribuyó a que el modelo de aprendizaje automático dispusiera de información relevante para la clasificación de las anomalías.

Se generó un dataset a partir de las características extraídas de las regiones de interés y se utilizó para entrenar un modelo Random Forest capaz de clasificar registros normales y anómalos. Los resultados obtenidos durante la evaluación demostraron que el modelo alcanzó un alto nivel de desempeño, especialmente al utilizar las 25 características seleccionadas, reduciendo los errores de clasificación respecto al primer experimento.

RECOMENDACIONES

Es conveniente continuar evaluando el sistema con nuevos videos de plasma obtenidos en diferentes condiciones de operación. Esto permitirá verificar que el comportamiento observado durante las pruebas realizadas en este trabajo se mantenga al analizar nuevos datos y contribuirá a fortalecer la confiabilidad del sistema en futuras aplicaciones. Asimismo, resulta pertinente incorporar nuevas opciones de visualización que permitan presentar con mayor detalle la información generada durante el procesamiento, como estadísticas adicionales o gráficos.

Por otra parte, se sugiere realizar pruebas de funcionamiento cada vez que se actualicen las librerías o se incorporen nuevas funcionalidades al sistema. De esta manera, se puede garantizar que todos los módulos continúen trabajando de forma correcta y que las modificaciones no afecten el desempeño general de la aplicación.

También es aconsejable mantener criterios consistentes al momento de seleccionar las regiones de interés, procurando que estas representen adecuadamente las zonas donde se desea realizar el análisis. Una selección uniforme favorece la comparación de resultados entre diferentes videos y contribuye a mantener la calidad de la información utilizada por el sistema.

Adicionalmente, se recomienda realizar estudios comparativos del desempeño del sistema frente a otros modelos de análisis de plasma, como XGBoost y LIGHTGBM, con el fin de identificar fortalezas y posibles áreas de mejora respecto a enfoques alternativos. En esta misma línea, resulta de gran interés explorar la implementación de YOLOv8 para la detección y análisis de filamentos, dado su probado rendimiento en tareas de segmentación y detección de objetos.

Finalmente, sería deseable ampliar progresivamente el conjunto de datos incorporando nuevos videos de plasma obtenidos bajo diferentes condiciones de adquisición. Contar con una mayor diversidad de muestras permitirá seguir evaluando el comportamiento del modelo y disponer de una base de datos más representativa para futuras investigaciones o mejoras del sistema.

BIBLIOGRAFÍA

- [1] R. Szeliski, Computer Vision: Algorithms and Applications, London: Springer, 2010.
- [2] I. Goodfellow, Y. Bengio, y A. Courville, Deep Learning, Cambridge: MIT Press, 2016.
- [3] L. E. Gordillo Erazo y J. L. Yáñez Roca, «Aplicación de visión con LabVIEW para la detección de frascos con turbiedades,» Guayaquil, 2009.
- [4] S. D. Peñafiel Falcón y L. O. Espinosa Viteri, «Entrenamiento del modelo YOLO para detección de una placa vehicular previamente capturada en imagen o video y aplicación de OCR para obtención de sus caracteres,» Quito, 2022.
- [5] E. C. S. B. y D. A. A. Fernández-Delgado, «Do we need hundreds of classifiers to solve real world classification problems?,» *Journal of Machine Learning Research*, vol. 15, p. 3133–3181, 2014.
- [6] F. F. Chen, Introduction to Plasma Physics and Controlled Fusion, New York: Springer, 2016.
- [7] D. F. a. J. Ponce, Computer Vision: A Modern Approach, Upper Saddle River, New Jersey, USA: Pearson, 2012.
- [8] C. M. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2006.
- [9] O. J. M. S. O. Yilmaz, «Object Tracking: A Survey,» *ACM Computing Surveys*, vol. 38, nº 4, p. 1–45, 2006.

- [10] S. H. S. S. A. S. S. Heravi, «survey of deep learning-based object detection in video analysis,» *Journal of Visual Communication and Image Representation*, vol. 81, 2021.
- [11] S. N. d. P. d. Ecuador, «Plan Nacional de Desarrollo 2025–2029: Ecuador no se detiene,» Ecuador, Quito, 2025.
- [12] A. S. & B. H. Tanenbaum, *Modern Operating Systems*, Pearson, 2015.
- [13] B. e. a. Shneiderman, *Designing the User Interface*, Pearson, 2016.
- [14] M. (2024), «Visual Studio Code Documentation,» Microsoft. (2024), [En línea]. Available: <https://code.visualstudio.com/docs>.
- [15] P. S. Foundation, «Python Documentation,» 2025. [En línea]. Available: <https://docs.python.org/3/>.
- [16] G. B. a. A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, CA, USA: O'Reilly Media: Sebastopol, 2008.
- [17] C. R. H. e. al, «Array programming with NumPy,» *Nature*, vol. 585, p. 357–362, 2020.
- [18] W. McKinney, *Python for Data Analysis*, Sebastopol, California, USA: O'Reilly Media, 2017.
- [19] F. P. e. al, «Scikit-learn: Machine Learning in Python,» *Journal of Machine Learning Research*, vol. 12, p. 2825–2830, 2011.
- [20] J. D. Hunter, «Matplotlib: A 2D Graphics Environment,» *Computing in Science & Engineering*, vol. 9, n° 3, p. 90–95, 2007.

- [21] P. Documentation, «Riverbank Computing Limited,» Riverbank Computing Limited, Didcot, United Kingdom, 2023.
- [22] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, Sebastopol, CA, USA: O'Reilly Media, 2019.
- [23] Y. O. J. & Z. Y.-Q. Wang, Video Processing and Communications, New Jersey, USA: Prentice Hall, Upper Saddle River, 2002.
- [24] R. C. & W. R. E. Gonzalez, Digital Image Processing (4th ed.), New York, USA: Pearson, 2018.
- [25] D. A. & P. J. Forsyth, Computer Vision: A Modern Approach (2nd ed.), Boston, Massachusetts, USA: Pearson Education, 2011.
- [26] A. K. Jain, Fundamentals of Digital Image Processing, USA: Prentice-Hall: Englewood Cliffs, NJ,, 1989.
- [27] A. K. D. R. P. W. & M. J. Jain, «Statistical Pattern Recognition: A Review,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, n° 22, p. 4–37, 2000.
- [28] S. Few, Show Me the Numbers: Designing Tables and Graphs to Enlighten, Burlingame, California, USA: Analytics Press, 2012.
- [29] P. K. S. a. A. A. Mamun, Introduction to Dusty Plasma Physics, Bristol: UK: Institute of Physics Publishing, 2002.
- [30] D. A. Forsyth y J. Ponce, Computer Vision: A Modern Approach, Upper Saddle River, NJ: Pearson, 2012.

- [31] R. J. H. a. G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed, Melbourne: Australia: OTexts, 2021.
- [32] T. M. Mitchell, *Machine Learning*, New York, N: USA: McGraw-Hill, 1997.
- [33] L. Breiman, «Random Forests,» *Machine Learning*, vol. 45, n° 1, p. 5–32, 2001.
- [34] A. B. y. V. K. Varun Chandola, «Anomaly Detection: A Survey,» *Review article (Survey paper)*, p. 1–58, 2009.
- [35] C. C. Aggarwal, *Outlier Analysis*, Cham, Switzerland: Springer, 2017.
- [36] S. K. a. N. B. N. L. Johnson, *Continuous Univariate Distributions*, New York, NY: USA: Wiley, 1994.
- [37] J. W. M. a. A. T. C. R. V. Hogg, *Introduction to Mathematical Statistics*, 8th ed, Boston, MA,: USA: Pearson, 2019.
- [38] D. M. W. Powers, «Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,» *Journal of Machine Learning Technologies*, vol. 2, n° 1, p. 37–63, 2011.
- [39] N. D. a. B. Triggs, «Histograms of Oriented Gradients for Human Detection,» *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 886–893, 2005.
- [40] J. Canny, «A Computational Approach to Edge Detection,» *IEEE TPAMI*, vol. 8, n° 6, p. 679–698, 1986.

- [41] V. Chandola, A. Banerjee y V. Kumar, «Anomaly Detection: A Survey,» *ACM Computing Surveys*, vol. 41, nº 3, p. 1–58, 2009.
- [42] R. Hernández Sampieri, C. Fernández Collado y P. Baptista Lucio, *Metodología de la investigación*, México D. F: McGraw-Hill Education, 2014.
- [43] I. Sommerville, *Software Engineering*, Pearson: Boston, MA, USA, 2016.
- [44] A. B. a. V. K. V. Chandola, «Anomaly detection: A survey,» *ACM Computing Surveys*, vol. 41, nº 3, p. 1–58, 2009.
- [45] R. T. a. J. T. Hastie, Friedman, *The Elements of Statistical Learning*, 2nd ed, New York, NY: USA: Springer, 2009.

ANEXOS

Anexo 1 Pantalla importacion de datos error al cargar un archivo

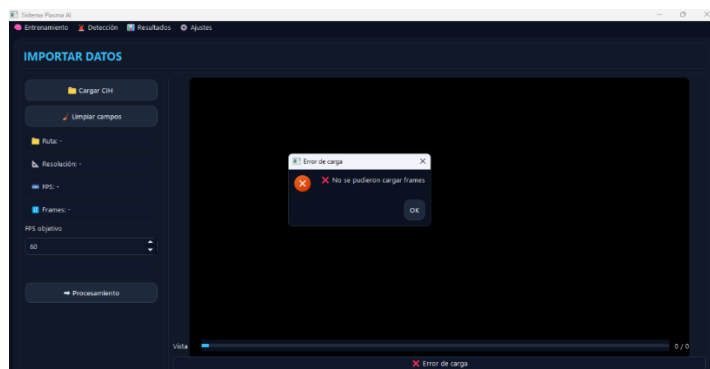


Figura 18: Mensaje de error al cargar un archivo ch

Anexo 2 Pantalla importación de datos selección del archivo

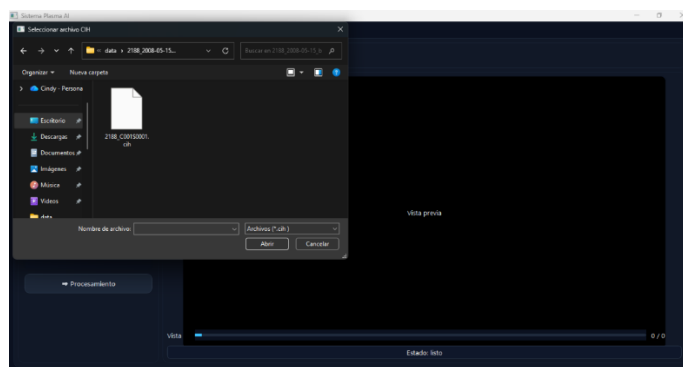


Figura 19: Selección de archivo

Anexo 3 Pantalla importación de datos archivo cargado correctamente

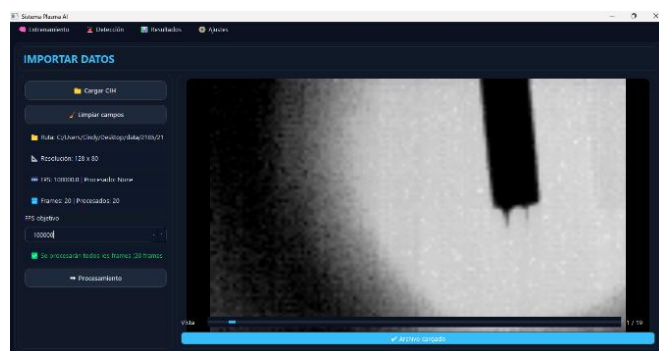


Figura 20: Archivo cargado correctamente

Anexo 4 Pantalla procesamiento extracción de característica

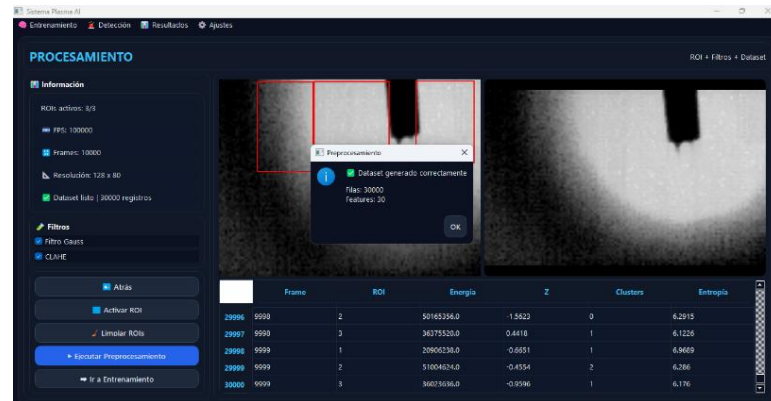


Figura 21: Extracción de característica

Anexo 5 Pantalla entrenamiento del modelo

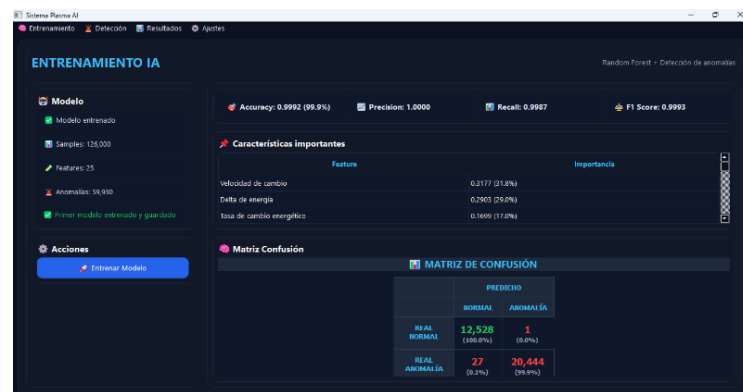


Figura 22: Entrenamiento del modelo

Anexo 6 Pantalla detección de anomalías selección de archivo

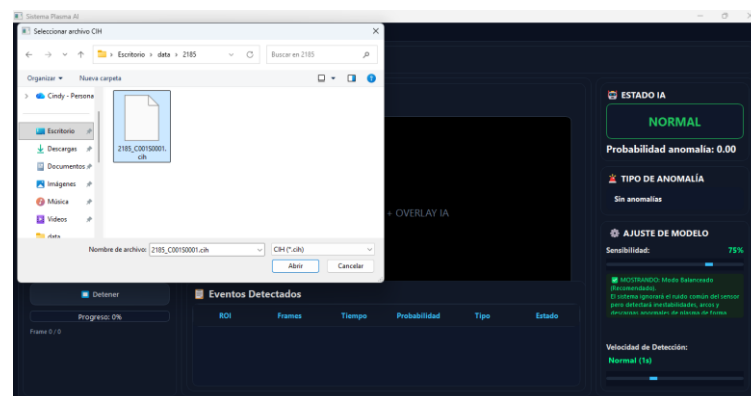


Figura 23: Selección de archivo pantalla anomalías

Anexo 7 Pantalla detección de anomalías archivo cargado correctamente

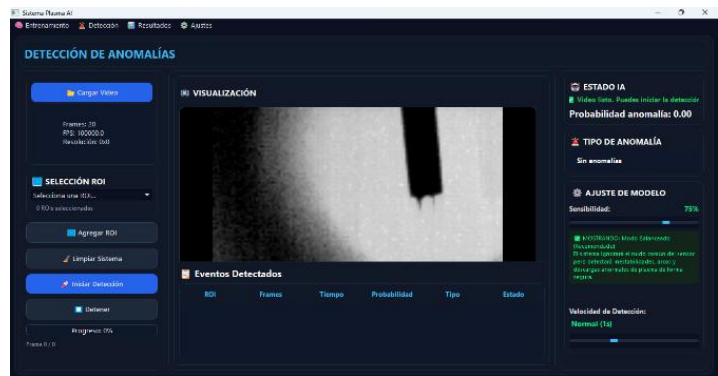


Figura 24: Archivo cargado correctamente

Anexo 8 Pantalla detección de anomalías error debe seleccionar al menos una roi

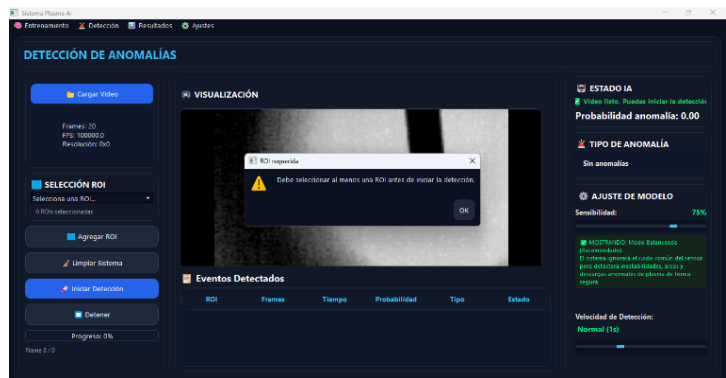


Figura 25: Seleccionar al menos una región de interés

Anexo 9 Pantalla detección probabilidad de anomalías con modelo entrenado

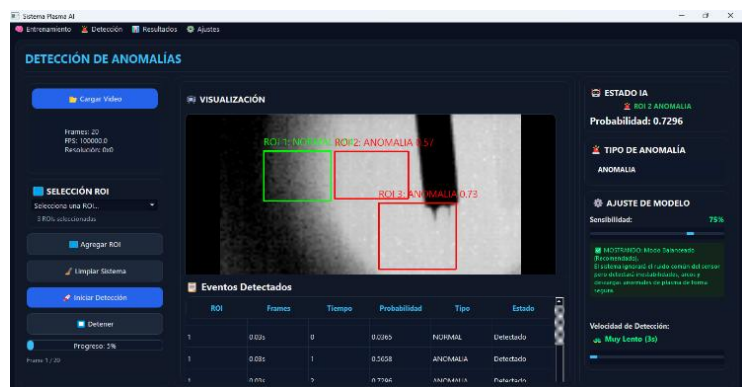


Figura 26: Probabilidad del modelo en las regiones de interés

Anexo 10 Gráfica de energía absoluta

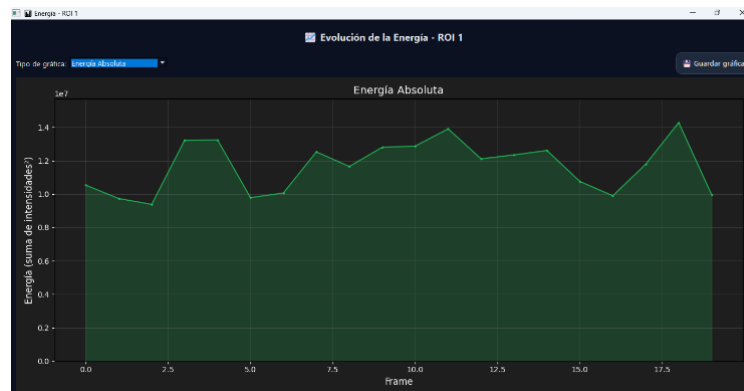


Figura 27: Energía absoluta de una región de interés

Anexo 11 Gráfica de energía normalizada

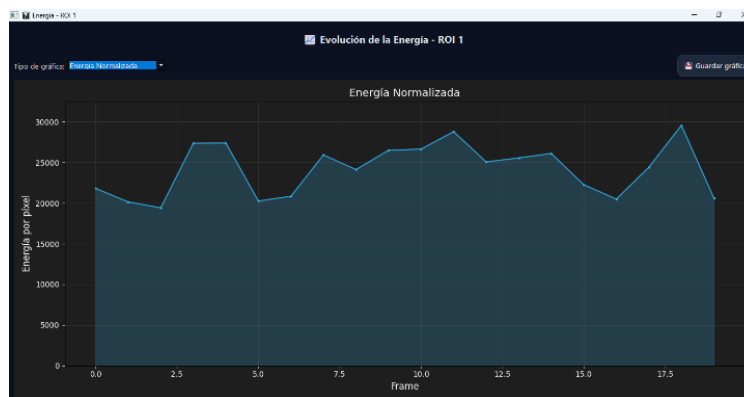


Figura 28: Energía normalizada de una región de interés

Anexo 12 Gráfica de z-score



Figura 29: Gráfica de z-score de una región de interés

Anexo 13 Gráfica de probabilidad del modelo

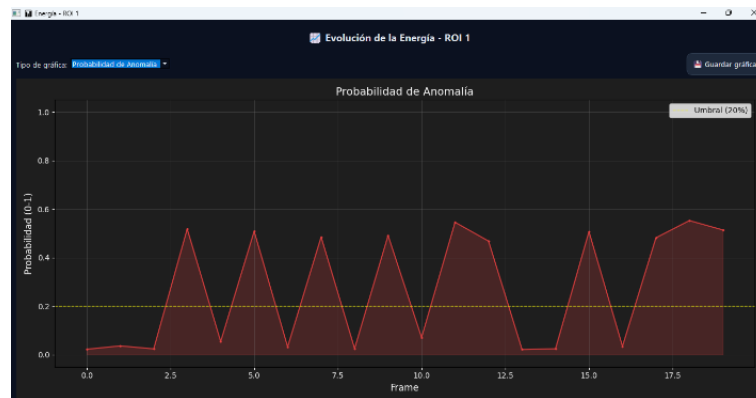


Figura 30: gráfica de probabilidad de modelo

Anexo 14 Guardado de sesiones autoaticamete

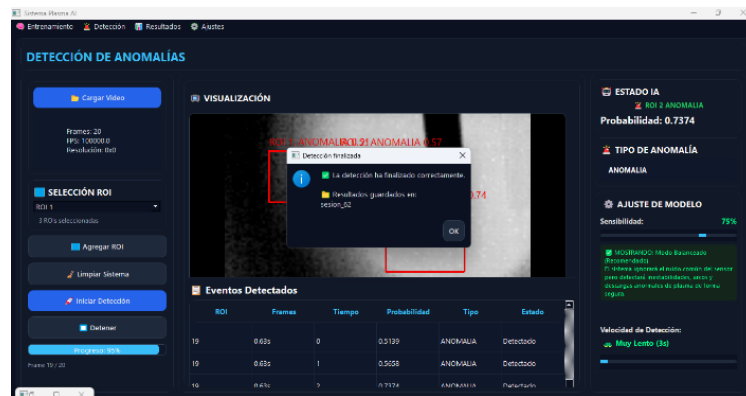


Figura 31: Guardado en análisis en su respectivo archivo

Anexo 15 Visualización pantalla resultado

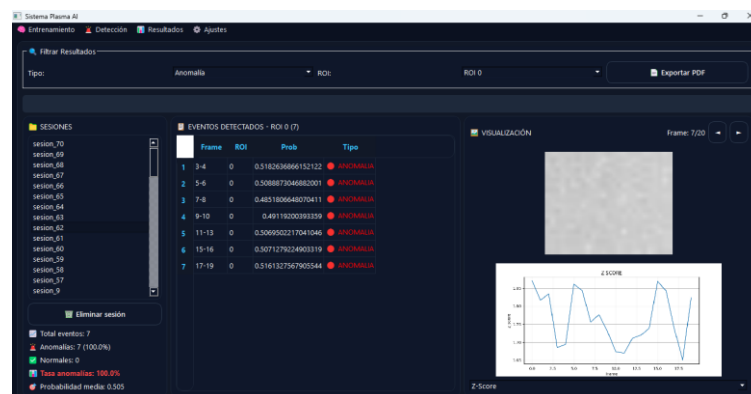


Figura 32: Visualización pantalla resultado

Anexo 16 Reporte de sesión

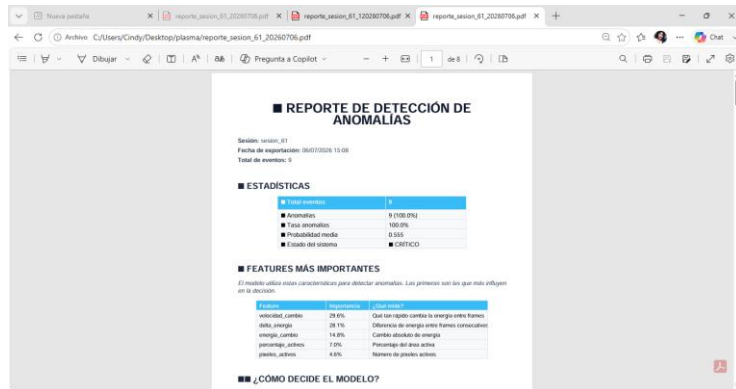


Figura 33: Reporte de detección