

Sistema de información basado en tecnología Web, Orientado al análisis de detección de cancer de mama a nivel mundial



**Alan Fernandez
Daniela Guzman
Josue Nisthaus
Jhulianna Tarqui**

**Sistemas de Información III - 2024
Ing. Yamil Cárdenas**

Dataset: <https://www.kaggle.com/datasets/adhyanmaji31/breast-cancer-prediction/code>

1. INTRODUCCIÓN

El cáncer de mama es una de las enfermedades más prevalentes y mortales entre las mujeres en todo el mundo, lo que subraya la importancia de su detección temprana. Un sistema de información web orientado al análisis de detección de cáncer de mama a nivel mundial, puede no solo facilitar la carga y análisis de datos clínicos, sino también ofrecer predicciones automatizadas y comparaciones con casos históricos, mejorando la precisión diagnóstica y contribuyendo a una atención más rápida y eficiente en la lucha contra el cáncer de mama.

Con el fin de alcanzar este propósito, se emplearán tecnologías avanzadas para el desarrollo de un sistema web como HTML, CSS, JavaScript, React, Cloud, asegurando una interfaz interactiva y fácil de utilizar. Además, se utilizarán modelos de análisis funcional como IDEF0 para el diseño estructurado del sistema y BPMN para la optimización de los procesos de flujo de información. En conjunto, estas tecnologías proporcionarán un sistema integral, eficiente y flexible que responda a las necesidades de accesibilidad urbana, mejorando así la calidad de vida de las personas con discapacidad.

2. IDEA BRAINSTORM

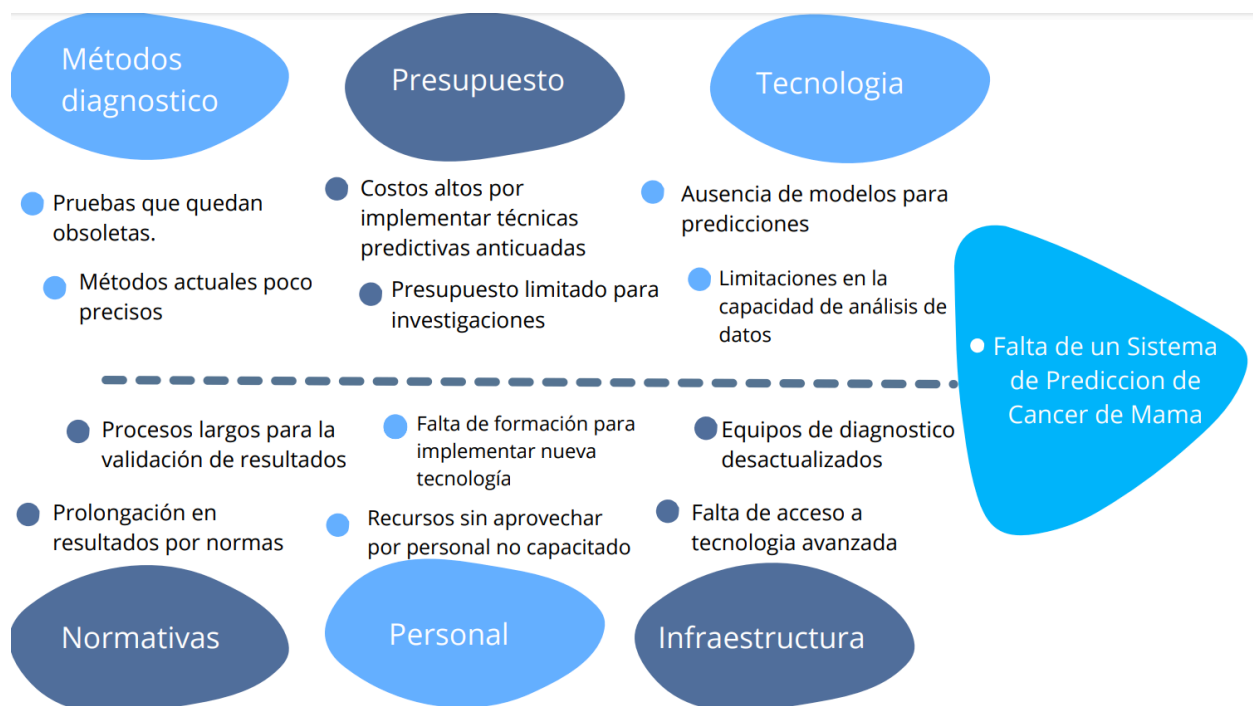


• CONTEXTO

El sistema de predicción para la detección de cáncer de mama se encuentra bajo el contexto de proporcionar una solución integral que permita diagnosticar esta enfermedad de manera precisa y eficiente, utilizando herramientas tecnológicas avanzadas. La plataforma ofrece una *interfaz intuitiva* y accesible para médicos y usuarios, permitiendo la *carga de datos* médicos, como imágenes de mamografías y otros *historiales clínicos*, para su análisis. Además, cuenta con funciones de *visualización de datos*, presentando los resultados de manera clara mediante gráficos y reportes que se generan en un *DashBoard interactivo*. A través de un modelo de predicción automática, se calcula la probabilidad de cáncer de mama, lo que facilita la evaluación del modelo y mejora la toma de decisiones. El sistema asegura un diagnóstico

seguro, generando alertas automáticas cuando se detectan posibles riesgos, y permite la exportación de *informes personalizados*, útiles en consultas y estudios médicos. También se pueden aplicar *filtros personalizados* para obtener análisis más específicos, mientras que un *dashboard interactivo* permite el monitoreo en tiempo real. Se incorporan *casos históricos* para mejorar la precisión de los *modelos predictivos*, y se ofrece la personalización de informes según las necesidades del usuario. Con la integración de tecnologías como React para una interfaz interactiva y adaptable, y la compatibilidad con Python para facilitar el trabajo de desarrolladores, el sistema está diseñado para ser accesible y eficiente, con el objetivo final de mejorar el diagnóstico y tratamiento del cáncer de mama.

3. ISHIKAWA



• DEFINICIÓN DEL PROBLEMA

La falta de un Sistema de Predicción de Cáncer de Mama genera un conflicto, ya que en la actualidad existe una gran variedad de herramientas y recursos que pueden ser aprovechados por instituciones afines a la salubridad, lo cual no se está realizando. Las pruebas quedan

obsoletas a medida que se generan nuevos descubrimientos en el área de medicina. Y también se debe considerar que los métodos actuales (aplicados en hospitales de La Paz) son poco precisos. Es importante tener en cuenta que existen costos altos por implementar técnicas predictivas de cáncer de mama, también que en la mayoría de hospitales, resulta costoso y no se cuenta con suficiente presupuesto para la realización de investigaciones. En cuanto a la tecnología, hoy en día existe un amplio campo del machine learning y minería de datos que pueden ser aprovechados para generar modelos de predicción, los cuales no se implementan y son ausentes. Por otro lado, también hay muchas limitaciones en la capacidad de análisis de datos por parte de los hospitales.

4. PREGUNTA DEL PROBLEMA

¿Cómo se puede desarrollar una solución tecnológica que permita mejorar la precisión y rapidez en la detección temprana del cáncer de mama a nivel mundial, considerando la necesidad de analizar grandes volúmenes de datos clínicos y la identificación de factores de riesgo clave para apoyar a los profesionales de la salud en la toma de decisiones?

5. OBJETIVO GENERAL Y ESPECÍFICO

5.1. Objetivo General:

Desarrollar un Sistema de información basado en tecnología Web, orientado al análisis y detección del cáncer de mama a nivel mundial, para ayudar a los profesionales de la salud a diagnosticar de manera precisa y rápida casos de cáncer de mama, facilitando la toma de decisiones clínicas con acceso global.

5.2. Objetivos Específicos:

- Desarrollar un sistema web accesible globalmente, que permita a los profesionales de la salud cargar características clínicas y obtener una predicción automática sobre la probabilidad de que un tumor sea benigno o maligno, usando algoritmos de aprendizaje automático.

- Implementar un análisis de casos clínicos históricos a nivel mundial, permitiendo a los usuarios del sistema comparar los casos de sus pacientes con registros previos similares y consultar el diagnóstico y tratamiento aplicado en esos casos.
- Identificar y visualizar los principales factores de riesgo clínicos que influyen en la clasificación de tumores malignos, destacando los aspectos más importantes para la toma de decisiones informadas en el diagnóstico.
- Generar reportes personalizados y automatizados en tiempo real, que incluyan los resultados de la predicción, análisis de similitud de casos y los factores clave que influyen en la detección de cáncer, accesibles a través de la plataforma web para su consulta y seguimiento.
- Asegurar la actualización constante del sistema, integrando bases de datos y registros médicos internacionales, manteniendo así el modelo entrenado con datos actuales y relevantes de múltiples regiones del mundo para mejorar su precisión y aplicabilidad global.

6. SMART

Specific (Específico):

Desarrollar un sistema web global que permita la carga y análisis de datos clínicos, como imágenes de mamografías, para proporcionar predicciones automatizadas sobre la probabilidad de cáncer de mama. El sistema también debe generar reportes personalizados y permitir la comparación con casos históricos para ayudar a los profesionales de la salud en la toma de decisiones clínicas informadas.

Measurable (Medible):

El éxito del proyecto puede medirse a través de:

- Precisión del modelo predictivo: Evaluar la tasa de aciertos de las predicciones (benigno/maligno) con respecto a los diagnósticos médicos reales, estableciendo una meta de al menos un 85% de precisión.
- Tiempo de procesamiento: Lograr que el análisis de datos y la predicción tomen menos de 15 segundos para cada caso clínico.

Achievable (Alcanzable):

El proyecto es alcanzable porque las tecnologías necesarias, como React, machine learning (con bibliotecas como TensorFlow o scikit-learn), y almacenamiento en la nube ya están bien desarrolladas y disponibles. El equipo tiene acceso a bases de datos históricas y algoritmos de predicción que se pueden entrenar con datos clínicos existentes. Además, el diseño de una interfaz amigable facilitará la adopción por parte de médicos sin necesidad de formación técnica avanzada.

Relevant (Relevante):

Este sistema es altamente relevante debido a la necesidad global de mejorar la detección temprana del cáncer de mama. La plataforma contribuirá directamente a reducir la mortalidad al permitir que los profesionales de la salud realicen diagnósticos más precisos y rápidos, lo que llevará a tratamientos más efectivos y oportunos. Además, con la actualización constante de los datos y el análisis predictivo, se mantiene alineado con los avances en medicina.

Time-bound (Limitado en el Tiempo):

El proyecto tiene un plazo estimado de:

- 3 semanas para el desarrollo inicial del sistema web, incluyendo la carga de datos, análisis predictivo y generación de reportes.
- 1 semana para la integración de bases de datos históricas y optimización de los modelos de predicción.
- 4 días para pruebas piloto.

Este marco temporal permitirá entregar una solución funcional dentro de un año, con actualizaciones continuas y soporte a largo plazo.

7. PART

Procesos (Process):

Ingreso al Sistema: El médico accede al sistema con sus credenciales.

Carga de Datos Clínicos: Se ingresan los datos clínicos del paciente, como imágenes de mamografías y otros resultados médicos.

Verificación y Procesamiento: El sistema valida los datos y los procesa, asegurándose de que sean compatibles para el análisis.

Análisis Predictivo: Utilizando algoritmos de aprendizaje automático (machine learning), el sistema realiza una predicción sobre la probabilidad de que un tumor sea benigno o maligno.

Visualización de Resultados: Los resultados se muestran en un dashboard interactivo para que el médico los revise.

Diagnóstico y Decisión: El médico compara los resultados con casos históricos, toma decisiones informadas y proporciona explicaciones al paciente.

Almacenamiento de Resultados: Los diagnósticos y resultados finales se almacenan en la base de datos, validando o refutando la predicción del sistema.

Actores (Actors):

Médico: Principal usuario del sistema, responsable de ingresar los datos clínicos y revisar las predicciones para tomar decisiones informadas.

Paciente: Aunque no interactúa directamente con el sistema, es el sujeto de los análisis y el beneficiario final del diagnóstico.

Sistema Web: Plataforma automatizada que realiza el análisis de datos, predicciones y visualización de resultados.

Administradores del Sistema: Encargados de la gestión y mantenimiento técnico del sistema.

Recursos (Resources):

Base de Datos Médica Internacional: Fuentes de datos clínicos históricos para mejorar la precisión de los modelos predictivos.

Modelos de Machine Learning: Algoritmos utilizados para realizar las predicciones automáticas sobre el cáncer de mama.

Informes Personalizados: Reportes automáticos y gráficos generados para facilitar la comprensión de los datos clínicos y predicciones.

Pacientes y Registros Clínicos: Datos médicos del paciente, como mamografías, exámenes de laboratorio y otros historiales.

Tecnologías (Technology):

Desarrollo Web: Tecnologías como HTML, CSS y JavaScript se emplean para construir la interfaz interactiva y adaptable.

Frameworks Frontend: Uso de React para desarrollar una interfaz de usuario dinámica y accesible.

Backend y Cloud Computing: Implementación en servidores en la nube para almacenamiento seguro y procesamiento de grandes volúmenes de datos.

Machine Learning (Python): Herramientas y bibliotecas como scikit-learn o TensorFlow para el análisis de datos y predicción automatizada.

IDEF0 y BPMN: Modelos utilizados para estructurar y optimizar los procesos de flujo de información y diseño del sistema.

8. ANÁLISIS

8.1. BPMN

Definición

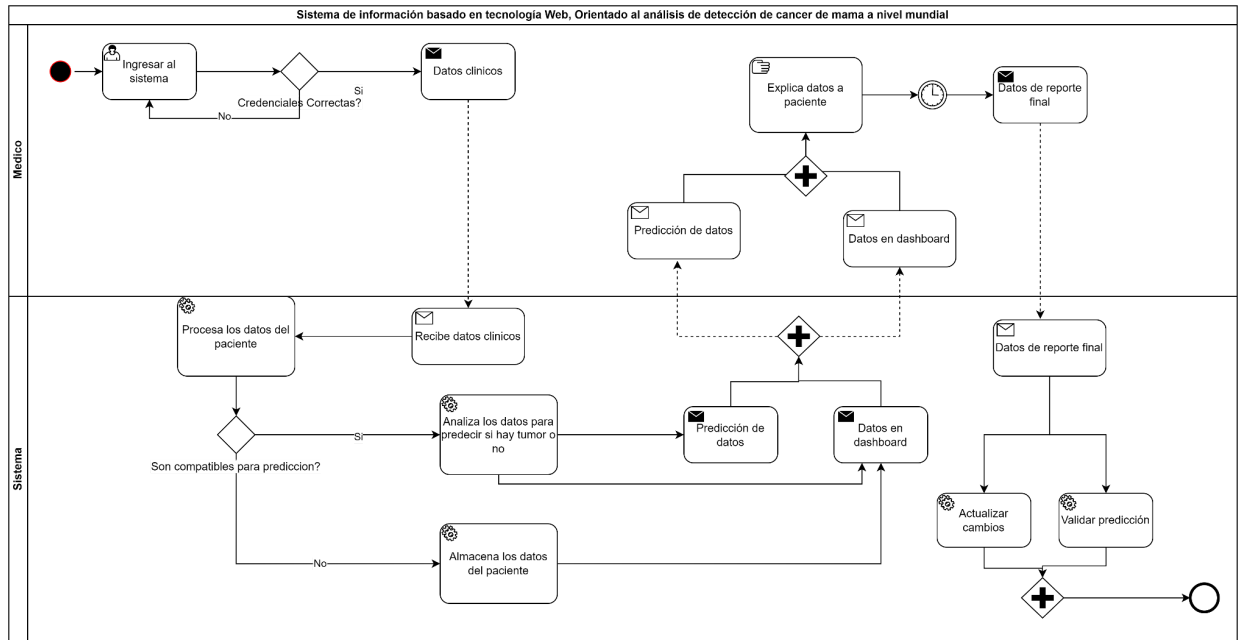
Es un estándar gráfico que funciona como un diagrama cuya función es presentar el procesos de negocios, se secciona en diferentes partes y sirve para hacer un análisis del negocio, un buen análisis sobre este diagrama puede ayudar a mejorar la eficiencia de procesos así también de implementar o reducir flujos de trabajo, además de definir de forma clara y comprensiva cada proceso que se genera en un negocio.

Un BPMN consta de eventos, actividades, discusiones y conectores.

- **Representación de eventos:** Los eventos marcan el inicio, intermedio y final de un proceso. Dependiendo de su forma, puede representar el inicio, el cambio, momento de espera y el final de cada serie de subprocesos. En los diagramas se representan mediante círculos.
- **Actividades:** Son las tareas a realizar en medio de los procesos, estos son como accionadores y su propósito es el de definir los pasos de forma denominativa. Su representación gráfica son rectángulos redondeados. Toda actividad debe hallarse conectado, sea bien a una actividad o a un evento, porque toda actividad debe contener un principio, un flujo y un fin.
- **Decisiones del proceso:** Es un elemento esencial, su principal característica es que bifurca un proceso en dos situaciones dependiendo de su uso o el contexto en el subproceso, gráficamente se representa con un rombo.
- **Conectores:** Este elemento conecta los eventos, actividades y decisiones del diagrama BPMN, su representación gráfica es la de una flecha. El inicio de la flecha podría considerarse como la causa y la punta de la flecha representaría el efecto que produce dicho evento o actividad.

BPMN del Proyecto

El presente proyecto tiene como BPMN el siguiente diagrama:

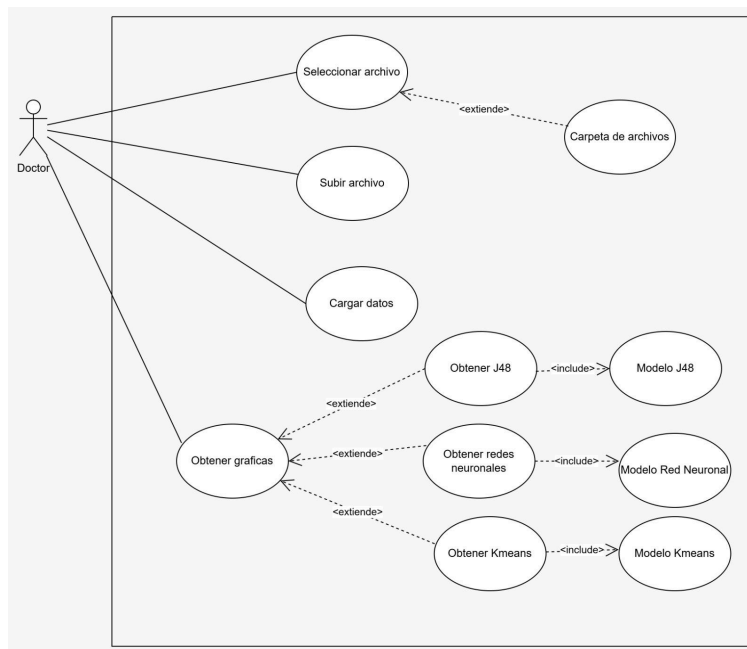


A continuación, se explica paso a paso el diagrama:

1. El proceso inicia cuando el médico ingresa al sistema con sus credenciales donde hay una pasarela de decisión que verifica si las credenciales son correctas. Si no lo son, el sistema rechaza el acceso. Si las credenciales son válidas, se continúa al siguiente paso.
2. Una vez dentro del sistema, el médico introduce los datos clínicos del paciente, como los resultados de pruebas y análisis relevantes para la predicción del cáncer de mama.
3. El sistema recibe estos datos y los procesa para verificar si son compatibles con el análisis predictivo. Si los datos son compatibles, el proceso continúa.
4. El sistema utiliza modelos de machine learning o algoritmos para analizar los datos y predecir si existe la presencia de un tumor maligno o benigno.
5. El sistema muestra la predicción de los datos tanto al médico como en un dashboard del sistema.

6. En este punto, el médico puede revisar los datos en el dashboard, con la predicción realizada, el médico puede explicar los resultados al paciente y tomar decisiones informadas sobre el tratamiento.
7. Después de realizar un diagnóstico final el médico ingresa este resultado al sistema, se almacena en la base de datos validando o no la predicción que se realizó con anterioridad.

Diagramas de casos de Uso



Cargar datos y Mostrar al usuario

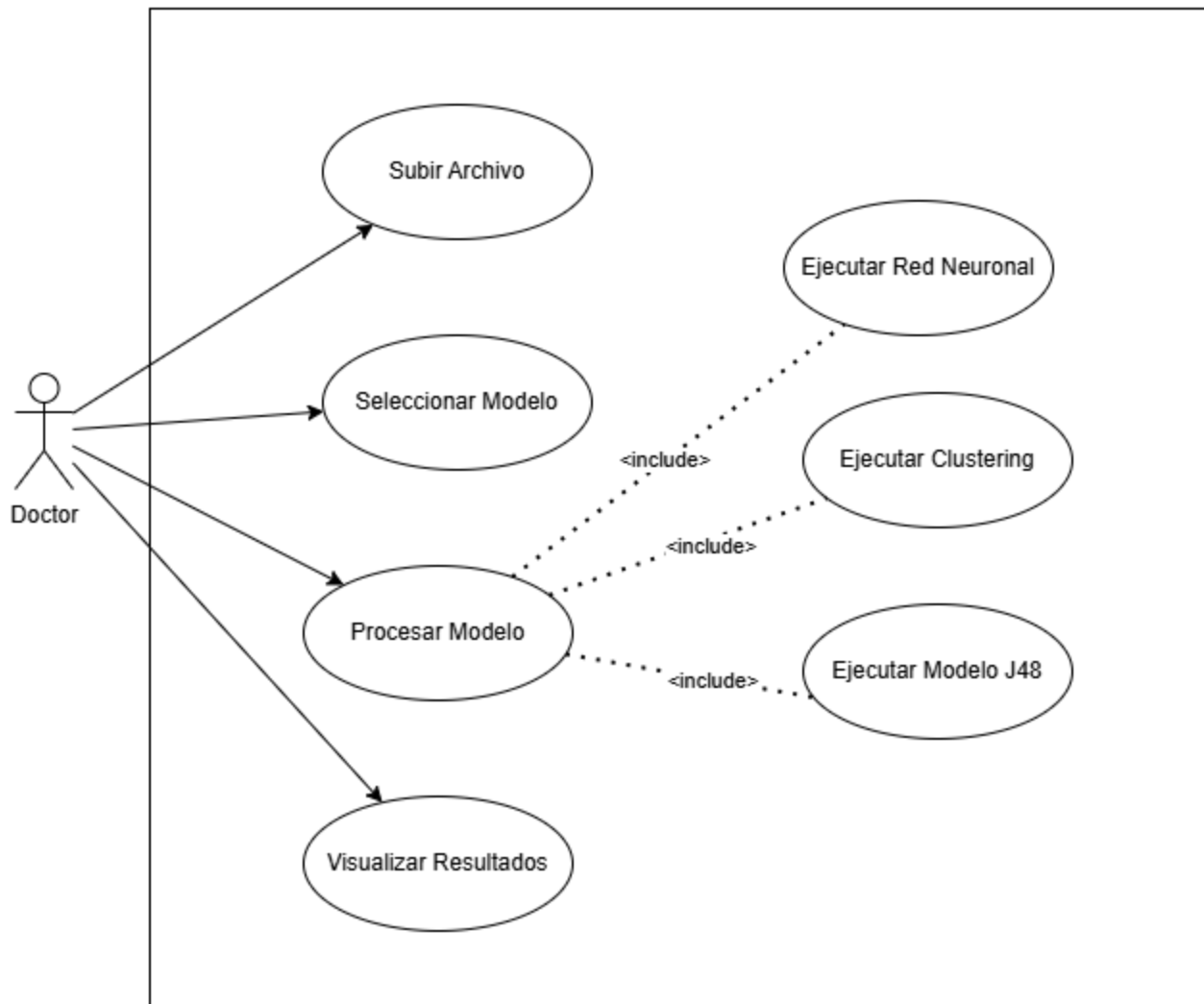
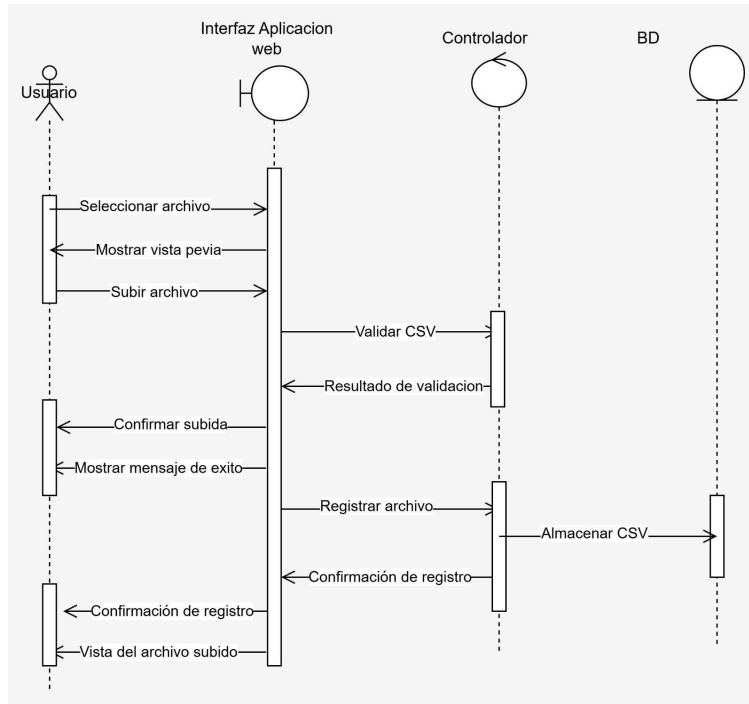


Diagrama de secuencia



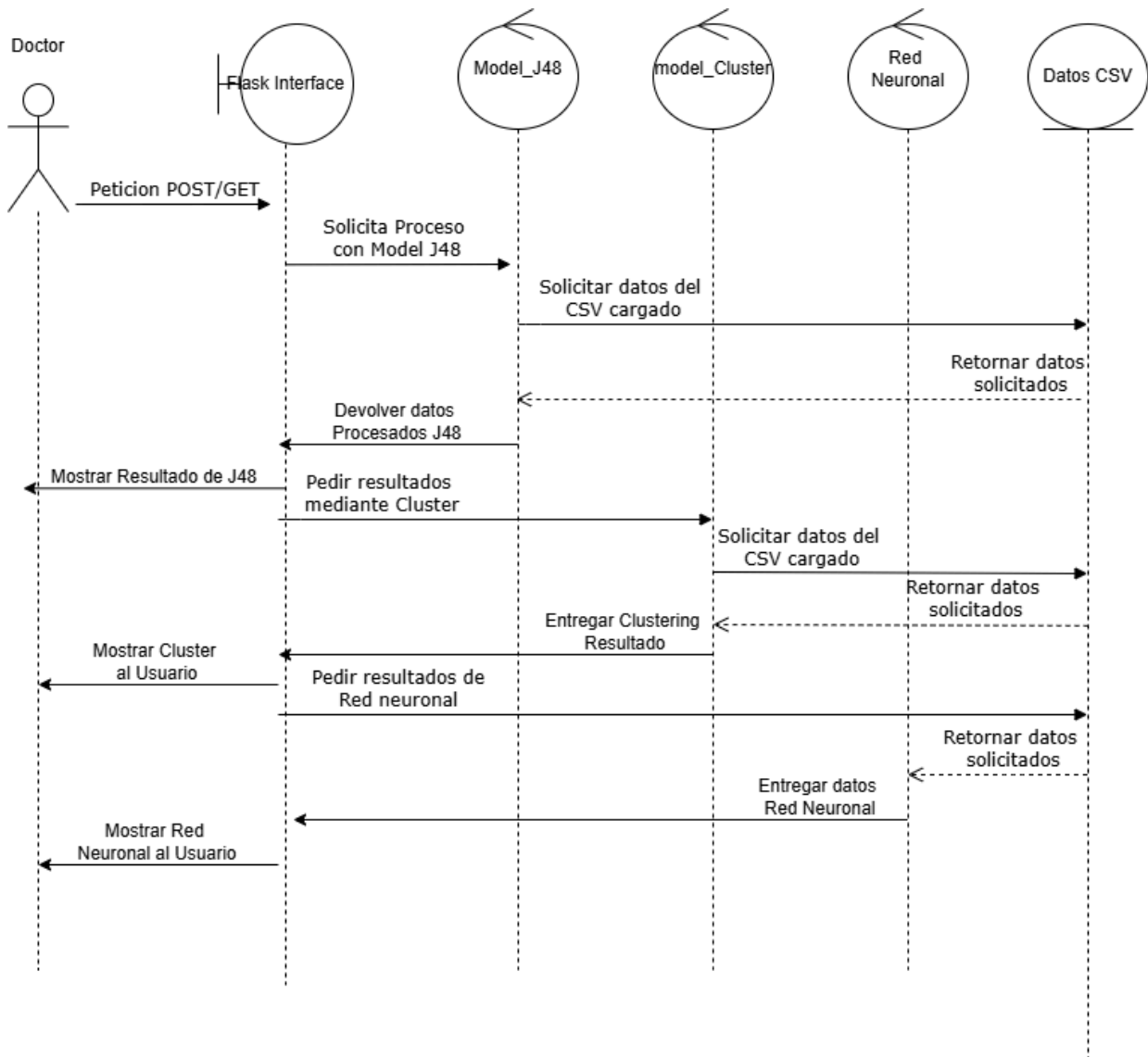


Diagrama de Comunicación

