Desarrollo de una plataforma integral modular para un ecosistema predictivo de gestión hospitalaria, la cual está orientada al análisis de datos. Como resultado de la aplicación de dichos modelos, se generarán reportes y dashboards interactivos que facilitarán la visualización e interpretación de la información, apoyando la toma de decisiones del instituto.

ALCANCE FUNCIONAL

El proyecto abarca el desarrollo de una plataforma de modelos predictivos, que integra los siguientes módulos:

- Módulo-00 Configuración.
- Módulo-01 Riesgo predictivo para desarrollar cáncer de mama en pacientes con resultado de BI-RADS 4,5 Y 6 en mastografías.
- Módulo-02 Alteraciones Electroencefalográficas en pacientes con diagnóstico de TEA
- Módulo-03 Modelo predictivo de riesgo de Cáncer Cervicouterino.
- Módulo-04 Dashboard integral

MÓDULO-00 CONFIGURACIÓN

- Login
    - Autenticación 2F
    - Cambio contraseña
    - Bloqueo de usuario por intentos erróneos
- Accesos
- Perfiles
- Usuarios
- Parametrización
- Catálogos Generales

M01 Riesgo predictivo para desarrollar cáncer de mama en pacientes con resultado de BI-RADS 4,5 Y 6 en mastografías.

Riesgo predictivo para desarrollar cáncer de mama en pacientes con resultado de BI-RADS 4,5 Y 6 en mastografías en Centro Médico ISSSTELEON.

Determinar el riesgo predictivo para desarrollar cáncer de mama en pacientes con resultado de BI-RADS 4,5 Y 6 en mastografías en el periodo de enero a diciembre de 2025.

- Correlacionar los resultados BI-RADS 4, 5 Y 6 con las variables de antecedentes personales y datos sociodemográficos.
- Establecer el perfil epidemiológico de las pacientes con resultados de BIRADS 4,5 y 6 en mastografías.
- Diseñar un modelo de riesgo predictivo que facilite la toma de decisiones y la mejora en los procesos.

Alcance

CONFIDENCIAL

- Implementar un módulo de procesamiento de lenguaje natural (NLP) para la extracción automatizada de información relevante a partir de reportes clínicos en texto libre.
- Preproceso de datos en tablas y datos de registros clínicos en texto libre.
- Integrar la información demográfica, personal y de imágenes en pacientes con BIRADS 4, 5 Y 6, en una base de datos depurada.
- Determinación de variables y evaluación estadística de indicadores clínicos (clasificación tumoral, distribución por estadio, tamaño tumoral y proporción de diagnósticos en etapas tempranas vs avanzadas).
- Clasificación de indicadores de diagnóstico que conducen al modelo predictivo.
- Diseñar modelo predictivo para estimar el riesgo de desarrollar la enfermedad.
- Desarrollar e implementar una herramienta digital que permitirá la visualización de indicadores clínicos, epidemiológicos, de riesgo a desarrollar la enfermedad, heatmap con grupo de indicadores clínicos relevantes y consulta de perfiles globales e individuales de pacientes a través de reportes (PDF, XLS), tablas y/o gráficas, mismos que se describen en la sección de **Indicadores**.

Fuentes de información y orígenes de datos

En esta sección se describen las fuentes de información que servirán como insumo para el desarrollo del modelo predictivo, incluyendo la base de datos, sistemas orígenes y repositorios relevantes proporcionados por el ISSSTELEON.

- Datos sociodemográficos, etapa reproductiva y antecedentes familiares.
- Datos clínicos, reportes de interpretación (clasificación BI-RADS) e imágenes (mastografía).

Indicadores

En la presente sección se describen los indicadores definidos para el análisis de la información, los cuales se detallan de manera individual para facilitar su interpretación y seguimiento.

- Porcentaje de pacientes con resultados de BI-RADS 4,5 Y 6 en población ISSSTELEON.
- Porcentaje individual de mastografías con resultado en la clasificación de BI-RADS 4, 5 Y 6.
- Calculadora de índices de riesgo; clasificación de pacientes en niveles de riesgo, probabilidad individual de recurrencia y supervivencia e identificación de factores pronósticos relevantes.

M02 Alteraciones Electroencefalográficas en pacientes con diagnóstico de TEA.

Identificación de alteraciones electroencefalográficas en pacientes de 16 a 36 meses con diagnóstico clínico positivo de TEA en el periodo de enero a diciembre de 2025.

Alcance.

- Implementación de un módulo de procesamiento de lenguaje natural (NLP) para la extracción automatizada de información relevante a partir de reportes clínicos en texto libre.
- Preproceso de las señales de EEG y los datos extraídos de los registros clínicos.
- Integrar la información de reportes clínicos de neurología pediátrica y registros de EEG de pacientes con diagnóstico positivo de TEA, en una base de datos depurada.
- Modelo para la extracción de alteraciones neurofisiológicas a partir de señales de EEG en pacientes pediátricos con diagnóstico positivo de TEA.
- Evaluación estadística y determinación de correlaciones entre alteraciones neurofisiológicas y los hallazgos clínicos, grado de severidad y patrones conductuales.

CONFIDENCIAL

- Modelo de estimación predictiva diagnóstica para el TEA basado en alteraciones neurofisiológicas.
- Desarrollar e implementar una herramienta digital (dashboard) que permitirá la visualización de indicadores a través de reportes (PDF, XLS), tablas y/o gráficas, mismos que se describen en la sección de **Indicadores**.

Fuentes de información y orígenes de datos.

En esta sección se describen las fuentes de información que servirán como insumo para el desarrollo de la plataforma digital, incluyendo la base de datos, software y repositorios relevantes proporcionados por el ISSSTELEON.

- Datos de reportes clínicos correspondientes al periodo de enero a diciembre de 2025 de pacientes pediátricos con diagnóstico de TEA.
- Archivos de EEG (de ser posible con acceso a software NeuroWorks EEG) y reportes de EEG en población pediátrica con diagnóstico de TEA.
- Criterios de inclusión; pacientes con diagnóstico positivo de TEA y con registro de EEG en el periodo de enero a diciembre de 2025.
- Criterios de exclusión; pacientes que no tengan registro de EEG y/o sin diagnóstico de TEA.

Indicadores

En la presente sección se describen los indicadores definidos para el análisis de la información, los cuales se detallan de manera individual para facilitar su interpretación y seguimiento.

Indicadores de las alteraciones neurofisiológicas

- Correlación entre las desviaciones electroencefalográficas y el grado de Autismo.
- Porcentaje de EEG normal y anormal de pacientes diagnosticados con TEA
- Porcentaje de Pacientes con diagnóstico de TEA con EEG
- Porcentaje de Pacientes con diagnóstico de TEA sin EEG.

Validación del modelo basado en alteraciones neurofisiológicas relacionadas con el TEA

La presente propuesta incorpora un marco de validación clínica, estadística y metodológica alineada con estándares internacionales, con el objetivo de garantizar la robustez, reproducibilidad y aplicabilidad al diagnóstico temprano del TEA.

M03 Modelo predictivo de riesgo de Cáncer Cervicouterino.

Desarrollar y validar un modelo de predicción de riesgo para la progresión a lesiones neoplásicas intraepiteliales cervicales de alto grado (NIC2, NIC3) y cáncer cervicouterino, a partir de resultados de tamizaje

CONFIDENCIAL

(citología y VPH), hallazgos colposcópicos, antecedentes clínicos y factores de riesgo, con el propósito de asignar una calificación de riesgo a cada paciente y optimizar la priorización en la atención clínica.

Alcance

- Integrar la información en una base de datos depurada y longitudinal que contemple todas las etapas del proceso de atención, desde el tamizaje inicial (citología y/o prueba de VPH), evaluación colposcópica, confirmación histopatológica, hasta el seguimiento clínico y tratamiento de las pacientes.
- Desarrollar, validar e implementar indicadores de calidad y modelos de riesgo de progresión individual a lesiones intraepiteliales cervicales de alto grado (NIC2, NIC3) y cáncer cervicouterino y supervivencia.
- Desarrollar un modelo predictivo que estimen el riesgo individual de progresión, integrando variables clínicas, resultados de laboratorio y antecedentes. Este modelo permitirá asignar una calificación de riesgo para la atención clínica y la toma de decisiones terapéuticas.
- Desarrollar e implementar una herramienta digital que permitirá visualizar indicadores clínicos y epidemiológicos, riesgo de progresión a CaCu y la probabilidad de supervivencia en pacientes y consultar perfiles individuales de riesgo, a través de reportes (PDF, XLS), tablas y/o gráficas, mismos que se describen en la sección de **Indicadores**.

Fuentes de información y orígenes de datos

En esta sección se describen las fuentes de información que servirán como insumo para el desarrollo del modelo predictivo, incluyendo la base de datos, sistemas orígenes y repositorios relevantes proporcionados por el ISSSTELEON.

- **Tamizaje**
  - Resultado de citología (ASC-US, LSIL, HSIL, etc.)
  - Prueba de VPH por PCR (positivo/negativo, genotipo **16 y 18**)
- **Colposcopía: Texto libre no estandarizado;** Índice Colposcópico de Reid (RCI), Índice de Swede (Swede Score)
  - Hallazgos colposcópicos
  - Índices de severidad
  - Nota medica de colposcopía
- **Histopatología**
  - NIC1, NIC2, NIC3
  - Resultados de citología cervical
- **Seguimiento y tratamientos**
  - Confirmación diagnóstica
- Diagnóstico médico para dar trazabilidad al proceso y atención oportuna.

Metodología

La metodología propuesta se basa en un enfoque estructurado y progresivo.

- Integrar una base de datos longitudinal que incluya información de tamizaje (citología, VPH), colposcopía, histopatología, antecedentes clínicos y seguimiento de pacientes.
- Preproceso de las bases de datos; limpieza y depuración de datos, registros duplicados, detección de inconsistencias, valores faltantes, estandarización de clasificaciones clínicas, codificación de variables categóricas, integración temporal de eventos clínicos.
- Se analizará la trazabilidad del proceso clínico; desde el tamizaje hasta el diagnóstico (biopsia / histopatología, clasificación, inicio de tratamiento, seguimiento y control) identificando tiempos de atención.

CONFIDENCIAL

- Se determinará la proporción de mujeres con diagnóstico de CaCu (casos existentes y casos nuevos por mes o año) por grupo etario, estadio y/o dependencia. Además, se establecerán análisis sobre bases de datos oficiales, reportadas por el sistema de salud estatal que servirán como referencia para estimar tasas de incidencia y prevalencia.
- Se identificarán factores de riesgo asociados a la progresión hacia NIC2, NIC3 o cáncer cervicouterino mediante análisis estadístico y modelos multivariados.
- Se desarrollará un modelo predictivo que estimen el riesgo individual de progresión, integrando variables clínicas, resultados de laboratorio y antecedentes. Definir puntos de corte de riesgo clínicamente relevantes que permitan clasificar a las pacientes en niveles de prioridad (atención inmediata, prioritaria o programada). Estos modelos permitirán asignar una calificación de riesgo para la atención clínica y la toma de decisiones terapéuticas.
- Se aplicarán un método bioestadístico para evaluar la supervivencia global (OS) y supervivencia libre de enfermedad (DFS). Se determinará probabilidad de supervivencia a 1, 3 y 5 años (por niveles de riesgo). Los resultados podrán ser observados en curvas de Kaplan-Meier y modelos para identificar variables asociadas a mayor riesgo de recurrencia o mortalidad.
- Implementar una plataforma digital que permita visualizar indicadores clínicos y epidemiológicos, riesgo de progresión a CaCu y la probabilidad de supervivencia en pacientes (por edad, estadio, entre otras variables significativas) y consultar perfiles individuales de riesgo. Esta herramienta digital permita la visualización del riesgo y facilite la toma de decisiones en el seguimiento y manejo clínico.

Indicadores

En la presente sección se describen los indicadores definidos para el análisis de la información, los cuales se detallan de manera individual para facilitar su interpretación y seguimiento.

- **Indicadores clínicos**
  - Distribución de resultados citológicos
  - Riesgo de desarrollo de Ca Cervicouterino
  - Tasa de confirmación histopatológica
  - Estimación de la supervivencia
- **Indicadores de oportunidad**
  - Tiempo de tamizaje a diagnóstico
  - Tiempo de diagnóstico a tratamiento

M04 Dashboard integral.

Este módulo integrará la visualización de indicadores definidos en cada módulo a través de reportes (PDF , XLS), tablas y/o gráficas, mismos que se describen en la sección de **Indicadores** de cada uno.

CONFIDENCIAL

La solución para cada propuesta se articula en tres componentes principales:

• Gestión de Datos: componente encargado de la recopilación, integración, curación y almacenamiento de datos provenientes de fuentes operativas y el historial médico de los derechohabientes, así como su almacenamiento en repositorios optimizados y separados de los sistemas transaccionales para no afectar su rendimiento.

• Analítica y modelado: modelos predictivos basados en analítica avanzada, orientados a la estimación de riesgos, proyección de escenarios y generación de indicadores clave para la toma de decisiones.

• Dashboard Analítico Interactivo: interfaz visual que presenta los indicadores claves, facilitando la interpretación de la información.

La arquitectura propuesta está basada en tecnologías modernas de alto rendimiento: Angular Material en el frontend, ASP.NET Core (C#) en el backend, y SQL Server como motor de base de datos principal, complementado con servicios de machine learning en Python/ML.NET.

Nota: La arquitectura técnica de diseño se confirmará entre el Equipo ISSSTELEON y USIT durante el desarrollo del proyecto para asegurar la compatibilidad de tecnologías inicialmente propuestas en este documento.

METODOLOGÍA PARA EL DESARROLLO DE MODELOS PREDICTIVOS.

En el área médica el desarrollo de modelos predictivos requiere una metodología integral que garantice la calidad de los datos, la confiabilidad de los resultados y el cumplimiento de criterios éticos y científicos. La presente propuesta contempla el desarrollo de modelos predictivos orientados al análisis y aprovechamiento de información clínica, mediante el uso de herramientas de analítica de datos e inteligencia artificial, que facilite la interpretación y uso de los resultados obtenidos.

Cada una de las etapas planteadas tiene como finalidad fortalecer la capacidad de análisis, identificación de patrones, factores de riesgo y variables de relevancia médica que apoyen la toma de decisiones en entornos médicos y de investigación.

Para ello, se consideran los siguientes componentes estratégicos:

- Definición de las fuentes de información

  Se delimitarán las poblaciones de estudio correspondientes a cada eje: pacientes con resultados de tamizaje cervical (citología y detección de VPH), pacientes pediátricos de 16 a 36 meses con diagnóstico clínico de TEA y registros electroencefalográficos asociados, y pacientes con hallazgos de mastografías clasificados como BI-RADS 4 5 y 6. Para cada población se identificarán las fuentes primarias de datos, que podrán incluir registros clínicos institucionales, sistemas de información en salud, expedientes electrónicos y bases de datos de estudios de gabinete, según corresponda.

- Integración y depuración de bases de datos

  Se diseñará una base de datos estructurada para cada eje que integre los datos clínicos, sociodemográficas, de imagen, neurofisiológicas y de laboratorio pertinentes. Se aplicarán herramientas de procesamiento de lenguaje natural (NLP) para extraer la información necesaria de

CONFIDENCIAL

los reportes clínicos en texto libre. También ser realizarán procedimientos estandarizados de depuración que incluirán la verificación de consistencia, el tratamiento de datos faltantes, la detección de valores atípicos, la codificación homogénea de variables y la integración temporal de eventos clínicos. Este proceso garantizará la calidad y la estandarización de la información antes de su análisis.

- Exploración y selección de biomarcadores clínicos

  Se realizará un análisis exploratorio de datos para caracterizar estadísticamente los datos, identificar correlaciones relevantes y detectar patrones preliminares en cada eje. Mediante herramientas de análisis (métodos estadísticos de aprendizaje supervisado y no supervisado) se evaluará la representación de los datos y la reducción de dimensionalidad o bien el manejo de datos completo. Se determinarán los datos que aporten la mayor capacidad discriminativa para cada resultado esperado: progresión a lesiones de alto grado o cáncer cervicouterino, parámetros de señales electroencefalográficas asociadas con TEA y a la progresión en lesiones asociadas al cáncer de mama. A partir de la identificación de los patrones asociados a estos ejes, se evaluarán las propiedades cuantificables relacionadas a diagnóstico y pronóstico para su posible uso como biomarcadores clínicos.

- Desarrollo y validación de modelos predictivos

  Se construirán modelos predictivos empleando técnicas de bioestadística, análisis multivariable, aprendizaje automático e inteligencia artificial, seleccionadas en función de la naturaleza de los datos y el desenlace de cada eje. Entre las técnicas candidatas se considerarán regresión logística con bootstrapping, árboles de decisión, bosques aleatorios, máquinas de soporte vectorial, redes neuronales y modelos de ensamble, entre otros. Cada modelo será sometido a procesos de validación interna mediante partición de datos (entrenamiento-prueba) y validación cruzada. El desempeño se evaluará con métricas estandarizadas como sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, área bajo la curva ROC (AUC) y calibración del modelo.

- Plataforma analítica

  Los modelos con mejor desempeño se integrarán en una plataforma analítica que permita la estimación automatizada del riesgo en cada uno de los tres ejes de interés. De acuerdo con los requerimientos planteados por el área médica especializada en cada eje de análisis, la plataforma facilitará la generación de reportes clínicos, indicadores epidemiológicos y de desempeño institucional, orientados a apoyar la priorización diagnóstica, el seguimiento individualizado y la evaluación de los procesos de atención. Esta plataforma analítica tendrá como fuente a la base de datos estructurada, integrada y depurada, que funcionará como repositorio intermedio para la explotación de la información. Una capa de integración de datos mediante servicios web (APIs), orientada a la actualización y sincronización de la información desde los sistemas fuente hacia la base de datos intermedia. También se integrará el código fuente de la plataforma tecnológica, que incluye: módulos de procesamiento de lenguaje natural (NLP), reportes analíticos y dashboards interactivos. Se entregará el manual de usuario, que describe la operación funcional del sistema y un manual técnico, que documenta la arquitectura de la solución, incluyendo los procesos de integración, modelos analíticos y flujos de información.

CONFIDENCIAL

- Consideraciones éticas

  El manejo de la información se realizará conforme a la normatividad vigente en materia de protección de datos personales y de investigación en salud. Se podrá garantizar la confidencialidad de los datos y tener la aprobación de comités de ética e investigación de la institución, en caso de ser requerido.

CALIDAD DE LOS DATOS DISPONIBLES.

Los resultados generados por la plataforma dependen directamente de la calidad, integridad, consistencia y representatividad de los datos disponibles. En aquellos casos donde la información sea insuficiente, incompleta, heterogénea o no estandarizada los modelos pueden presentar limitaciones en su capacidad predictiva y diagnóstica. La plataforma constituye un apoyo analítico cuyo desempeño requiere validación clínica continua y depende de la adecuada calidad de los datos para generar resultados confiables.

ENTREGABLES.

Como resultado del desarrollo del proyecto, se entregará una solución de software orientada al análisis y aprovechamiento de la información, que incorpora técnicas de modelado predictivo, conformada por los siguientes componentes:

- Base de datos estructurada, integrada, depurada y preparada para el análisis, que funcionará como repositorio intermedio para la explotación de la información.
- Capa de integración de datos mediante servicios web (APIs), orientada a la actualización y sincronización de la información desde los sistemas fuente hacia la base de datos intermedia.
- Código fuente de la plataforma tecnológica, que incluye:
  - Módulos de procesamiento de lenguaje natural (NLP)
  - Modelo predictivo para estimar el riesgo de desarrollar cáncer de mama
  - Modelo de estimación predictiva diagnóstica para el TEA basado en alteraciones neurofisiológicas.
  - Modelos de riesgo para la progresión a lesiones neoplásicas y cáncer cervicouterino.
  - Modelo de supervivencia asociado a la clasificación de los pacientes por el riesgo para la progresión a lesiones neoplásicas y cáncer cervicouterino
  - Dashboards interactivos, que considera los indicadores mencionados en cada módulo.
- Manual de usuario, que describe la operación funcional del sistema.
- Manual técnico, que documenta la arquitectura de la solución, incluyendo los procesos de integración, modelos analíticos, y flujos de información.
- Carta de Cierre de Proyecto.

REQUERIMIENTOS ADICIONALES.

- Contar con el equipo de trabajo para establecer las definiciones de Reglas de Negocio por parte de ISSSTELEON.

- Se requiere acceso autorizado a la estructura de las bases de datos y a la información necesaria para el análisis y desarrollo de los modelos predictivo del proyecto, procurando el cumplimiento de políticas de seguridad, confidencialidad y protección de datos.

CONFIDENCIAL