

Modelos de *Deep Learning* en la predicción de propiedades ADME/Tox

Deep Learning models in ADME/Tox property prediction

Carlos Alfredo Hernández Álvarez^{1*}, Anabel Abreu Llanes².

Recibido: 12/2023 | Aceptado: 04/2024 | Publicado: 12/2024

Resumen

El presente trabajo aborda el desarrollo de un modelo de *Deep Learning* para la predicción de propiedades ADME/Tox en péptidos cíclicos modificados desde la perspectiva de la ingeniería de software. El objetivo fue establecer una metodología sólida y replicable que guíe el diseño y la implementación del modelo, asegurando la calidad y la trazabilidad del proceso. Para ello, se definieron los requisitos funcionales y no funcionales mediante un análisis detallado del dominio y se diseñó una arquitectura modular basada en principios de escalabilidad y mantenibilidad. Se utilizaron herramientas de modelado como diagramas de casos de uso, diagramas de componentes y diagramas de secuencia, para documentar las interacciones del sistema. La metodología empleada incluyó un enfoque iterativo e incremental, soportado por prácticas ágiles que permitieron validar continuamente los avances. Los resultados alcanzados incluyen un prototipo funcional, cuyo diseño garantiza la extensibilidad para incorporar nuevos métodos de codificación molecular y métricas de evaluación. Este enfoque demostró ser eficiente para estructurar procesos complejos y mejorar la predictibilidad del sistema.

1* Universidad de Ciencias informáticas/ UCI, Facultad de Ciencias y Tecnologías Computacionales.

2 Universidad de Ciencias informáticas/ UCI, Facultad de Ciencias y Tecnologías Computacionales.

La integración de principios de ingeniería de software en proyectos de bioinformática no solo optimiza el desarrollo técnico, sino que también favorece la reproducibilidad y aplicabilidad en escenarios de investigación y desarrollo farmacológico.

Palabras clave: ingeniería de software; *Deep Learning*; ADME/Tox; diseño de fármacos; metodología ágil.

Abstract

This paper deals with the development of a Deep Learning model for the prediction of ADME/Tox properties in modified cyclic peptides from a software engineering perspective. The aim was to establish a robust and reproducible methodology to guide the design and implementation of the model, ensuring the quality and traceability of the process. To this end, functional and non-functional requirements were defined through a detailed domain analysis, and a modular architecture was designed based on scalability and maintainability principles. Modelling tools such as use case diagrams, component diagrams and sequence diagrams were used to document system interactions. The methodology adopted included an iterative and incremental approach, supported by agile practices that allowed continuous validation of progress. The results achieved include a working prototype whose design ensures extensibility to incorporate new molecular encoding methods and evaluation metrics. This approach proved effective in structuring complex processes and improving system predictability. In conclusion, the integration of software engineering principles in bioinformatics projects not only optimizes technical development, but also improves reproducibility and applicability in pharmacological research and development scenarios.

Keywords: Software engineering; *Deep Learning*; ADME/Tox; drug design; agile methodology.

Introducción

El diseño de modelos de *Deep Learning* para predecir propiedades ADME/Tox en el contexto farmacológico plantea múltiples desafíos, tanto desde el punto de vista computacional como del diseño metodológico. Estos modelos deben garantizar la precisión predictiva, la reproducibilidad y la escalabilidad,

aspectos que pueden abordarse eficazmente mediante una sólida aplicación de principios de ingeniería de software (Patel, Thakor y Patel 2024). La integración de prácticas de desarrollo estructurado permite enfrentar la complejidad del problema, optimizar recursos y asegurar la calidad de los resultados (Shao et al. 2024).

En este trabajo, se propone un enfoque metodológico para el desarrollo de un modelo de *Deep Learning* aplicado a la predicción de propiedades ADME/Tox en péptidos cíclicos modificados, fundamentado en técnicas de ingeniería de software. Diversos estudios han destacado la relevancia de las metodologías ágiles en proyectos científicos por su capacidad para adaptarse a los cambios en los requisitos y facilitar el desarrollo incremental de sistemas complejos (Maltarollo et al. 2015; Cavasotto y Scardino 2022). En el ámbito del modelado computacional, el empleo de arquitecturas modulares y el diseño basado en patrones se ha mostrado efectivo para mejorar la extensibilidad y mantenibilidad de los sistemas predictivos (Orosz, Héberger y Rácz 2022; Wenzel, Matter y Schmidt 2019).

A pesar de los avances en predicción computacional, los modelos tradicionales de QSAR —Modelos de Relación Cuantitativa entre Estructura y Actividad— presentan limitaciones al abordar la complejidad de los péptidos modificados, lo que subraya la necesidad de adoptar enfoques más robustos, como el *Deep Learning* (Huang, Baber y Bahmanyar 2021; De Sá et al. 2022). Este estudio justifica su enfoque en la ingeniería de software al considerar que una metodología estructurada no solo incrementa la efectividad técnica, sino que también permite validar iterativamente los componentes del modelo, reduciendo errores y maximizando la confiabilidad.

El objetivo principal de esta investigación fue desarrollar un modelo predictivo utilizando principios de ingeniería de software, documentando detalladamente los requisitos, diseño y validación del sistema. Este enfoque no solo proporciona una solución técnica al problema planteado, sino que también establece una base metodológica replicable para futuros proyectos en el campo de la bioinformática y el diseño de fármacos (Alqahtani 2017; Zhang et al. 2018).

Materiales y métodos

Para este proyecto se ha seleccionado la metodología *Scrum*. Esta tiene un marco ágil debido a su capacidad de proporcionar una estructura organizada, al mismo tiempo, permitir flexibilidad y adaptabilidad, características esenciales para el desarrollo de un modelo de *Deep Learning*. Ofrece un enfoque iterativo e incremental a través de *sprints*, lo que facilita la gestión del proyecto en fases manejables y permite realizar ajustes y mejoras continuas basadas en los resultados y *feedback* obtenidos.

En este proyecto la metodología *Scrum* se basa en varios aspectos clave (figura 1):

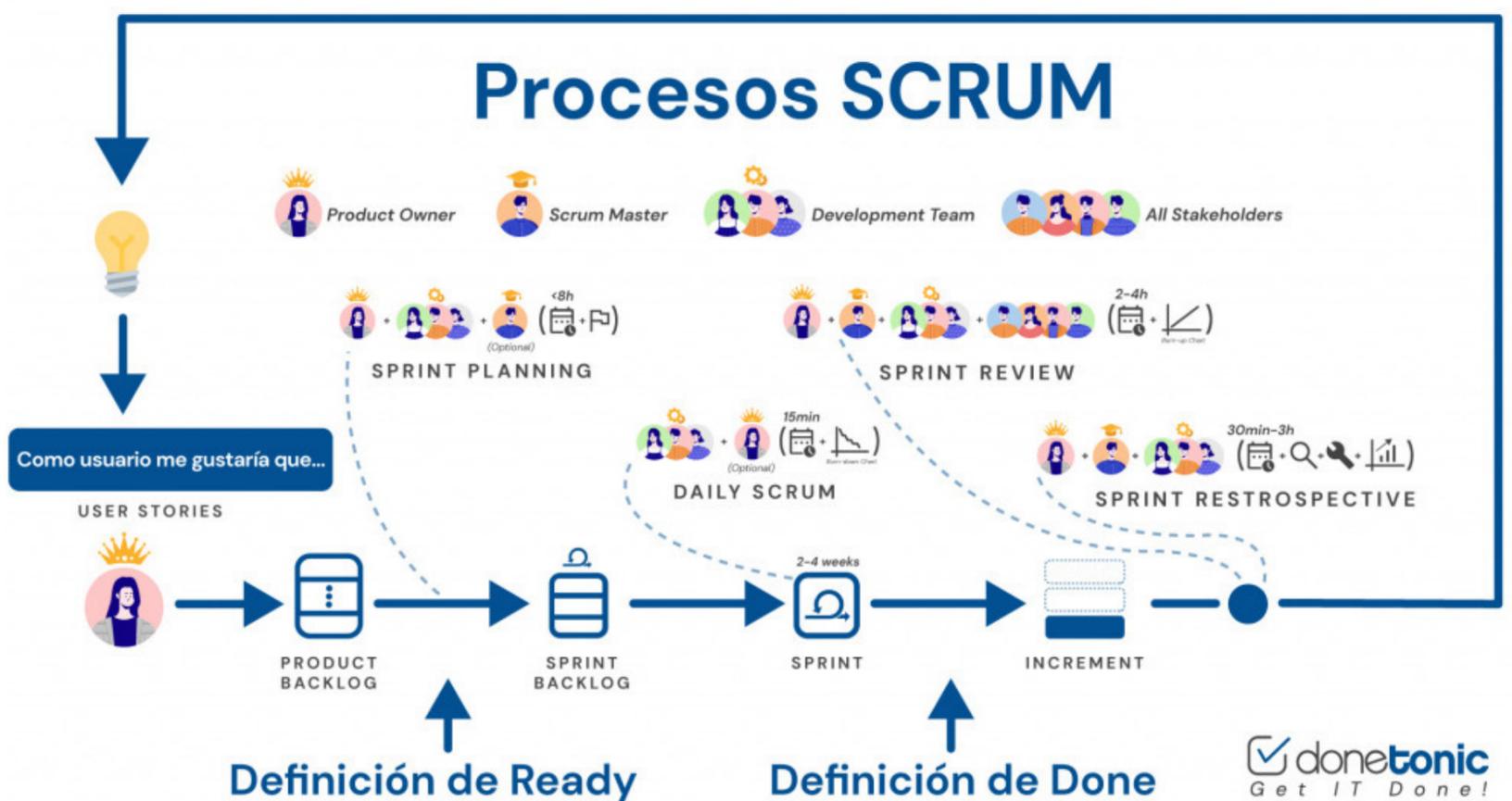


Figura 1. Componentes de la metodología *Scrum*

Composición del equipo: con tres integrantes, la comunicación cercana y la toma de decisiones rápida y eficiente son posibles gracias a la estructura de esta metodología, lo que es crucial para maximizar la productividad y minimizar los malentendidos en un equipo pequeño.

Experiencia del personal: la combinación de un desarrollador junior y dos *senior* en el equipo se ve reforzada por la estructura colaborativa de *Scrum*. Esto permite que ambos miembros del equipo trabajen en estrecha colaboración, aprovechando la experiencia del *senior* para guiar y mentorizar al junior, mientras que este último aporta una perspectiva fresca y nuevas ideas al equipo.

Criticalidad: la mayor consecuencia sería la interrupción del desarrollo del modelo, la pérdida de progreso significativo y la necesidad de repetir el trabajo ya realizado. Esto no solo consumiría recursos valiosos, como tiempo y dinero, sino que también podría retrasar significativamente el desarrollo del fármaco.

Dinamismo: es especialmente adecuado para proyectos como este, donde se requiere adaptabilidad y flexibilidad para responder con rapidez a cambios en los requisitos y descubrimientos durante el proceso de desarrollo del modelo.

Cultura: capaz de enfrentar y gestionar la complejidad inherente a la investigación y desarrollo de nuevos compuestos farmacéuticos. Esto implica una mentalidad abierta hacia el cambio, la disposición para experimentar y aprender de los fracasos y la capacidad para adaptarse rápidamente a nuevas circunstancias y descubrimientos.

A partir de los criterios de selección del proyecto se presenta como resultado la representación de los criterios para la selección del enfoque (ver figura 1).

Resultados y discusión

Descripción detallada de las funciones y características que se esperan del sistema, así como de los criterios de rendimiento y calidad que debe cumplir para satisfacer las necesidades del proyecto.

Requisitos Funcionales (RF):

RF1: Recibir solicitudes de análisis de péptidos

Debe permitir a los usuarios enviar solicitudes de análisis de péptidos cíclicos modificados a través de una interfaz intuitiva y fácil de usar.

RF2: Procesar solicitudes de análisis

Validar y procesar las solicitudes recibidas, asegurando la integridad de los datos ingresados.

RF3: Analizar detalladamente la estructura molecular

Realizar un análisis exhaustivo de la estructura molecular identificando características clave como enlaces, grupos funcionales y conformaciones.

RF4: Calcular los descriptores moleculares

Calcular una amplia gama de descriptores moleculares relevantes para la predicción de propiedades utilizando métodos

avanzados de cálculo y modelado molecular para asegurar la precisión y la fiabilidad de los descriptores calculados.

RF5: Generar informes detallados y comprensibles

Generar informes completos y comprensibles que resuman los resultados del análisis.

Requisitos No Funcionales (RNF):

RNF1: Interfaz de usuario adaptable

Proporcionar una interfaz de usuario adaptable que se ajuste a las necesidades y preferencias de los usuarios, permitiendo una fácil navegación y personalización de la experiencia.

RNF2: Escalabilidad y rendimiento óptimos

El sistema debe ser altamente escalable y capaz de manejar un gran volumen de solicitudes de análisis de manera simultánea sin comprometer su rendimiento.

RNF3: Seguridad y protección de datos

Garantizar la seguridad y la protección de los datos confidenciales y sensibles proporcionados por los usuarios durante el proceso de análisis.

RNF4: Mantenibilidad y extensibilidad

Ser fácilmente mantenible y extensible, permitiendo la incorporación de nuevas funcionalidades y la realización de actualizaciones periódicas sin interrupciones significativas en el servicio.

En la figura 2 se propone el diagrama de casos de uso y en la tabla 1 la descripción textual de cada caso de uso.

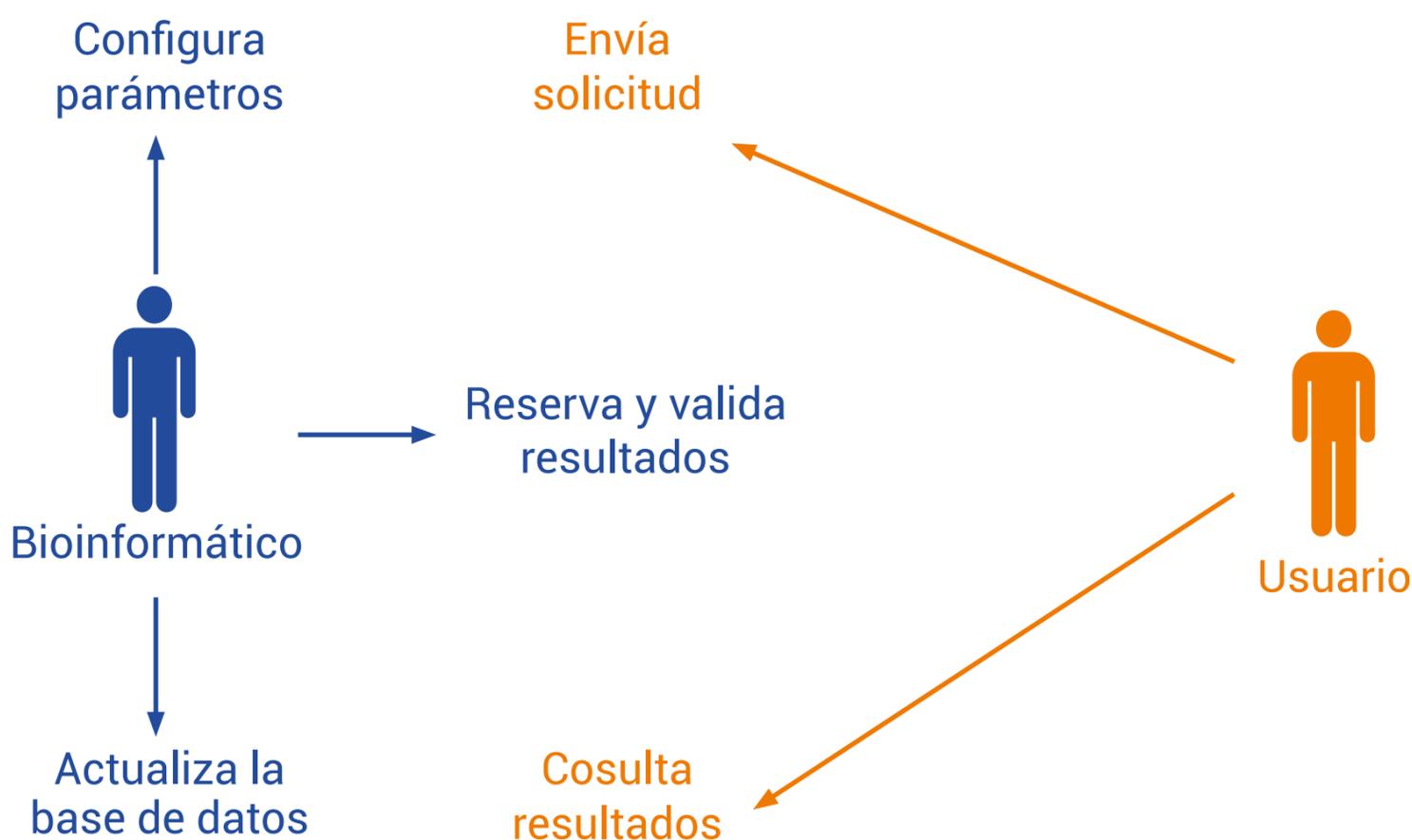


Figura 2. Diagrama de casos de uso

Actor	Caso de uso	Descripción
Usuario	Enviar la solicitud	Enviar una solicitud para analizar péptidos a través de la interfaz del sistema
	Consultar resultados	Consultar los resultados proporcionados
	Configurar parámetros del análisis	Ajusta los parámetros del análisis según las necesidades específicas del proyecto
	Revisar y validar resultados	Revisa y valida los resultados obtenidos por el sistema para garantizar su precisión
Bioinformático	Actualizar base de datos de péptidos	Actualiza y gestiona la base de datos para asegurar que la información esté actualizada
	<p>El usuario inicia el proceso enviando una solicitud de análisis. Se valida la integridad de los datos. Luego, el bioinformático configura los parámetros del análisis. A continuación, se analiza la estructura molecular y calcula descriptores moleculares. El bioinformático revisa y valida los resultados. Finalmente, se genera un informe detallado y el usuario puede consultarlo.</p>	

Tabla 1. Descripción textual de cada caso de uso

En la figura 3 se representa el modelo conceptual.

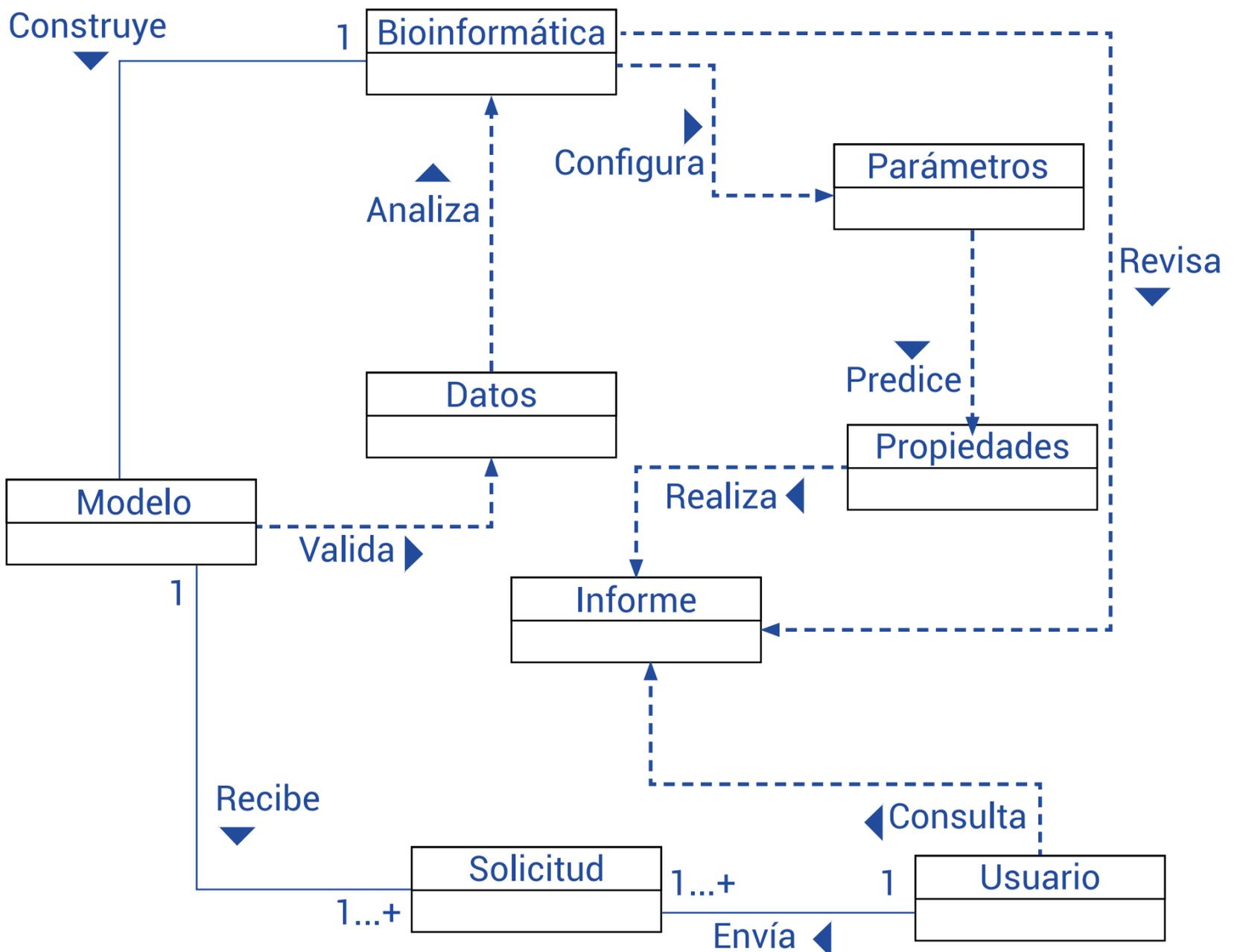


Figura 3. Modelo conceptual

Este modelo conceptual ilustra el proceso de predicción de propiedades:

Usuario: Persona que envía la solicitud para predecir y al final del proceso consulta los resultados.

Solicitud: Petición enviada por el usuario para predecir las propiedades.

Bioinformático: Profesional que construyó el modelo, analiza los datos, ajusta los parámetros necesarios para una buena predicción y se encarga de revisar el informe.

Datos: Información suministrada por el usuario en la solicitud y validados por el modelo, luego los analiza el bioinformático. En la tabla 2 se representa una descripción detallada del caso de uso “registrar solicitud”.

Parámetros: Configuraciones y ajustes específicos necesarios para ejecutar y asegurar la precisión de las predicciones.

Propiedades: Características físicas y químicas de los péptidos.

Modelo: Algoritmo capaz de predecir las propiedades de los fármacos.

Informe: Documento generado que contiene los resultados predichos.

En la figura 4 se representan los casos de usos de sistema y en la tabla 2 se representa una descripción detallada del caso de uso registrada en la solicitud.

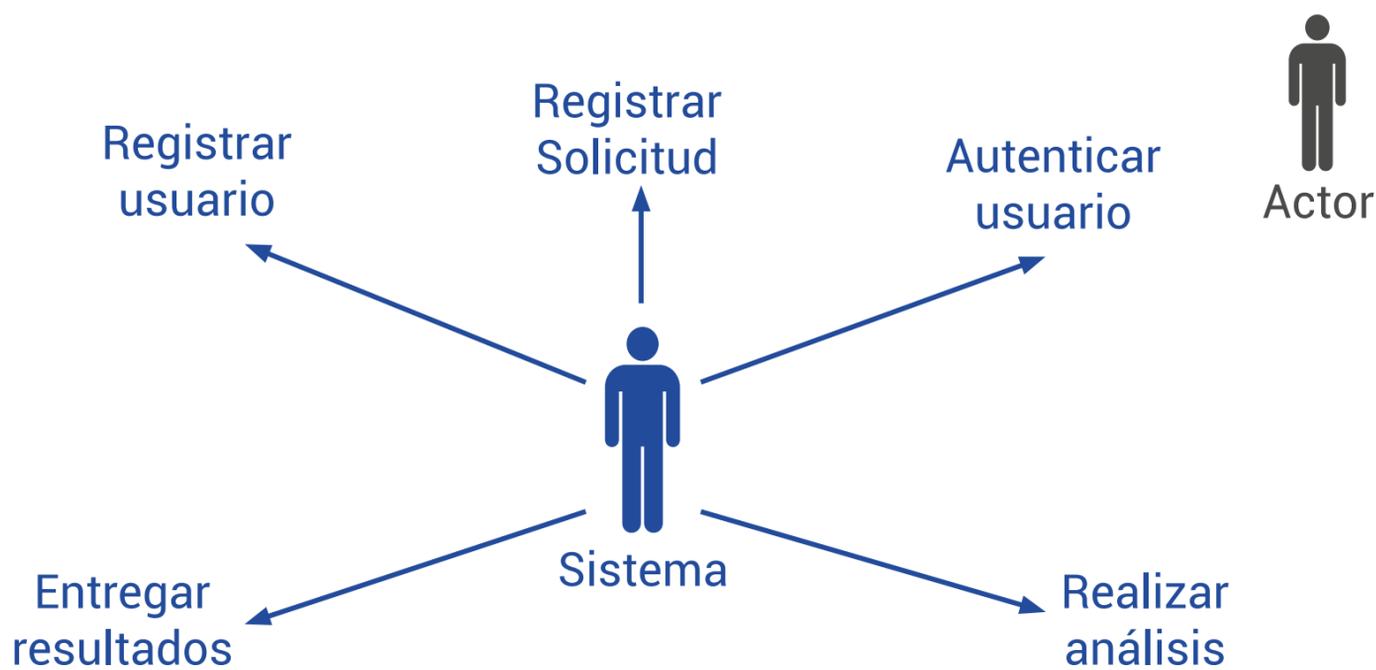


Figura 4. Casos de uso del sistema.

Objetivo	Permitir al usuario realizar una solicitud de análisis.
Actores	Usuario del sistema.
Resumen	El caso de uso se inicia cuando el usuario desea realizar un análisis.

Complejidad	Baja.
Prioridad	Alta.
Precondiciones	El sistema debe estar activo.
Postcondiciones	En usuario hizo una petición de análisis.
Flujo de eventos	
Flujo básico "Registrar solicitud"	
Actor	Sistema
1.	Introduce la URL para acceder.
2.	Introduce la secuencia y presiona "Analyze Peptide".
	<p>Muestra una pantalla con "en campo de texto" para introducir la secuencia y un botón para analizarla, además de mostrar los reportes recientes del usuario.</p> <p>a. Valida que se haya introducido una secuencia.</p> <p>b. Valida el formato de la secuencia.</p> <p>c. Verifica la autenticidad del usuario.</p> <p>d. Realiza el análisis.</p> <p>e. Agrega los resultados del análisis a los reportes recientes. Finalizando el caso de uso.</p>

Prototipo elemental de interfaz

The screenshot displays the Peptide Analyzer web application. The header includes the logo and navigation links: Home, Analyze Peptide, Reports, and Documentation. The main content area is divided into three sections:

- Analyze Your Peptide:** A form with a text input field for the peptide sequence and an "Analyze Peptide" button.
- Recent Reports:** A list of reports, with one highlighted: "Cyclic Peptide XYZ" submitted on 2023-06-01. Key metrics shown are Oral Bioavailability (45%), Clearance (2.5 L/h), and Toxicity (Low). A "View Report" link is provided.
- ADME/Tox Analysis:** A section showing horizontal progress bars for Pharmacokinetics (Oral Bioavailability, Clearance, Volume of Distribution) and Toxicology (Hepatotoxicity, Cardiotoxicity, Genotoxicity).

Below the screenshot is a 3D ball-and-stick model of a peptide structure, showing the backbone and side chains in various colors (grey, blue, red, white) against a dark blue background.

Figura 5. Prototipo elemental de la interfaz para el caso de uso del sistema

Flujos Alternos.		
2a. Campos vacíos.		
Actor.	Sistema.	
1.	Presiona Analyze Peptide	Muestra un mensaje: "Introduce una secuencia".
2b. Formato incorrecto		
Actor.	Sistema.	
1.	Presiona Analyze Peptide	Muestra un mensaje: "Formato incorrecto".
2c. Usuario inexistente		
Actor.	Sistema.	
1.	Presiona Analyze Peptide	Muestra un mensaje indicado: "Acceso denegado".
Relaciones.	CU Incluidos.	CU Autenticar usuario. CU Realizar análisis.
	CU Extendidos.	CU Entregar resultados.
Requisitos no funcionales.	No Aplica.	
Asuntos Pendientes.	No Aplica.	

Tabla 2. Representa una descripción detallada del caso de uso registrar solicitud.

Métodos de prueba de Caja Blanca:

- Se verificó que cada línea de código, incluyendo funciones de procesamiento de datos como la validación de solicitudes y el análisis de estructuras moleculares, se ejecutara al menos una vez durante las pruebas.
- Se aseguró que todas las bifurcaciones en el código, como las declaraciones *if-else* y los bucles, fueran probadas para verificar que todas las rutas posibles en la validación de la solicitud se manejen correctamente.

- Se realizaron pruebas para asegurar que todas las condiciones lógicas, como las configuraciones de parámetros del análisis, se evaluaran adecuadamente con diferentes configuraciones.
- Se ejecutaron pruebas para validar que todos los caminos posibles a través del código, en especial en la generación de informes detallados y comprensibles, fueran probados exhaustivamente.
- Se aseguró que las variables se iniciaran y usaran correctamente a lo largo del programa, en particular en el cálculo y uso correcto de descriptores moleculares en la predicción de propiedades.

Métodos de prueba de Caja Negra:

- Se enviaron solicitudes de análisis de péptidos cíclicos modificados para verificar que el sistema procesara las solicitudes y generara los informes de manera correcta.
- Se evaluó la interfaz para asegurar que fuera intuitiva y fácil de usar, permitiendo a los usuarios enviar solicitudes y consultar resultados de manera efectiva.
- Se realizaron pruebas de carga para garantizar que el sistema pudiera manejar múltiples solicitudes de análisis simultáneamente sin comprometer el rendimiento.
- Se verificó que los datos confidenciales de los usuarios estuvieran protegidos y que el sistema fuera resistente a posibles ataques de seguridad.
- Se realizaron pruebas con usuarios reales para evaluar la intuición de la interfaz y la comprensibilidad de los informes generados.
- Se probó la interfaz de usuario en diferentes navegadores y dispositivos para garantizar una experiencia consistente para todos los usuarios.
- Se aseguró que todos los componentes del sistema, como el módulo de análisis molecular y el de generación de informes, funcionaran correctamente juntos.

En la tabla 3 se brinda una representación del diseño de pruebas funcionales y en la tabla 4 se hace una descripción de variables.

Esce- nario	Descripción	Secuencia	Respuesta del sistema	Flujo cen- tral
EC 1.1. Formato correcto		V > Peptide XMLALP	El sistema confirma la re- cepción de la solicitud y muestra un mensaje de éxito.	
EC 1.2. Campo sin lle- nar	El usuario debe intro- ducir una secuencia, enviarla presionan- do "Analyze Peptide".	I (vacío)	El sistema procesa la so- licitud y muestra un men- saje de que hay campos vacíos.	Enviar la secuencia
EC 1.3. Formato in correc- to		I XMLALP	El sistema procesa la so- licitud y muestra un men- saje de que hay campos incorrectos.	

Tabla 3. Representación del diseño de pruebas funcionales.

No.	Nombre del campo	Clasificación	Valor nulo	Descripción
1	Secuencia	Fichero FAS- TA/texto	No	Almacena la solicitud del cliente para el análisis de una molécula, debe ser una secuencia con un ta- maño máximo de cincuen- ta caracteres o un archivo formato fasta.

Tabla 4. Descripción de las variables

Conclusiones

El estudio propone una metodología integral que combina técnicas de *Deep Learning* con el marco ágil de *Scrum* para abordar los desafíos en el diseño de fármacos, específicamente en la predicción de propiedades de péptidos cíclicos modificados. Se destaca la necesidad de superar las limitaciones de los modelos QSAR tradicionales debido a la complejidad estructural y diversidad química de estos compuestos. Mediante el desarrollo de un modelo de *Deep Learning* y la implementación de la metodología, se busca optimizar el proceso de diseño de fármacos y ampliar el conocimiento sobre la interacción entre la estructura molecular y las propiedades biológicas de los péptidos.

Referencias bibliográficas

- ALQAHTANI, S., 2017. *In silico ADME-Tox modeling: progress and prospects*. *Expert Opinion on Drug Metabolism & Toxicology* [en línea], [consulta: 2 julio 2024]. ISSN 1742-5255. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/17425255.2017.1389897>.
- CAVASOTTO, C. y SCARDINO, V., 2022. *Machine Learning Toxicity Prediction: Latest Advances by Toxicity End Point | ACS Omega*. [en línea]. [consulta: 2 julio 2024]. Disponible en: <https://pubs.acs.org/doi/full/10.1021/acsomega.2c05693>.
- DE SÁ, A.G.C., LONG, Y., PORTELLI, S., PIRES, D.E.V. y ASCHER, D.B., 2022. *toxCSM: comprehensive prediction of small molecule toxicity profiles*. *Briefings in Bioinformatics*, vol. 23, no. 5, ISSN 1467-5463, 1477-4054. DOI 10.1093/bib/bbac337.
- HUANG, D.Z., BABER, J.C. y BAHMANYAR, S.S., 2021. “The challenges of generalizability in artificial intelligence for ADME/Tox endpoint and activity prediction”. *Expert Opinion on Drug Discovery*, vol. 16, no. 9, ISSN 1746-0441. DOI 10.1080/17460441.2021.1901685.
- MALTAROLLO, V.G., GERTRUDES, J.C., OLIVEIRA, P.R. y HONORIO, K.M., 2015. “Applying machine learning techniques for ADME-Tox prediction: a review”. *Expert Opinion on Drug Metabolism & Toxicology*, vol. 11, no. 2, ISSN 1742-5255. DOI 10.1517/17425255.2015.980814.
- OROSZ, Á., HÉBERGER, K. y RÁCZ, A., 2022. “Comparison of Descriptor- and Fingerprint Sets in Machine Learning Models for ADME-Tox Targets”. *Frontiers in Chemistry* [en línea], vol. 10, [consulta: 2 julio 2024]. ISSN 2296-2646. DOI 10.3389/fchem.2022.852893. Disponible en: <https://www.frontiersin.org/journals/chemistry/articles/10.3389/fchem.2022.852893/full>.
- PATEL, A.S., THAKOR, K.B. y PATEL, M.K., 2024. *Predictive Modeling for ATME-TOX Properties of Drug Using Machine Learning: A Review*. , no. 01,
- SHAO, C., SHAO, F., HUANG, S., SUN, R. y ZHANG, T., 2024. An Evolved Transformer Model for ADME/Tox Prediction. [en línea]. [consulta: 2 julio 2024]. Disponible en: <https://www.mdpi.com/2079-9292/13/3/624>.
- WENZEL, J., MATTER, H. y SCHMIDT, F., 2019. “Predictive Multitask Deep Neural Network Models for ADME-Tox Properties: Learning from Large Data Sets”. *Journal of Chemical Information and Modeling*, vol. 59, no. 3, ISSN 1549-960X. DOI 10.1021/acs.jcim.8b00785.
- ZHANG, L., ZHANG, H., AI, H., HU, H., LI, S., ZHAO, J. y LIU, H., 2018. *Applications of Machine Learning Methods in Drug Toxicity Predict...: Ingenta Connect*. [en línea]. [consulta: 2 julio 2024]. Disponible en: <https://www.ingentaconnect.com/content/ben/ctmc/2018/00000018/00000012/art00006>.