Evaluación del riesgo alimentario mediante aprendizaje profundo y modelado molecular

Enol Junquera[1], Irene Díaz[1], Beatriz Remeseiro[1], Aida Zapico[2], Sonia González Solares[2], and Ferdinando Febbraio[3]

[1] Departamento de Informática, Universidad de Oviedo, Gijón, España
junqueraenol@uniovi.es
[2] Departamento de Biología Funcional, Universidad de Oviedo, Oviedo, España
zapicoaida@uniovi.es, soniagsolares@uniovi.es
[3] Consiglio Nazionale delle Ricerche (CNR), Nápoles, Italia
ferdinando.febbraio@cnr.it

Inicio del doctorado: septiembre de 2022, en régimen de dedicación a tiempo parcial

Resumen Esta propuesta doctoral aborda el uso de técnicas de inteligencia artificial para apoyar la evaluación del riesgo alimentario desde una perspectiva computacional, reproducible y alineada con las *New Approach Methodologies* (NAMs). La hipótesis central es que el riesgo asociado a la alimentación y a los compuestos químicos presentes en la cadena alimentaria no debe modelarse únicamente a partir de una modalidad de datos aislada, sino mediante representaciones que integren imagen, estructura química, interacción molecular y exposición dietética. La tesis se articula en torno a tres ejes conectados: el análisis automático de alimentos mediante visión artificial y aprendizaje multitarea; la predicción de toxicidad aguda, especialmente LD50, mediante arquitecturas profundas multimodales; y la construcción de representaciones exposición–molécula–proteína para estudiar señales asociadas al riesgo dietético y alteraciones colorrectales. Como resultado, la investigación propone un marco de inteligencia artificial para priorización y apoyo a la decisión en seguridad alimentaria, con potencial para reducir subjetividad en el seguimiento nutricional, orientar nuevos experimentos y contribuir a la reducción de experimentación animal.

Keywords: Inteligencia artificial · Aprendizaje profundo · Evaluación del riesgo alimentario · Docking molecular · Toxicidad · LD50 · NAMs · Visión artificial

1. Motivación y contexto de la tesis

La evaluación del riesgo alimentario requiere integrar información procedente de dominios muy distintos: composición nutricional, patrones de consumo, presencia de contaminantes o compuestos derivados del procesado, toxicidad química e interacción con sistemas biológicos. Este carácter multidimensional hace que la inteligencia artificial (IA), y en particular el aprendizaje profundo,

sean herramientas especialmente adecuadas para construir modelos capaces de extraer patrones a partir de datos heterogéneos.

El plan de investigación original partía de dos necesidades complementarias. Por un lado, mejorar el análisis automático de la dieta mediante imágenes de comida, reduciendo la subjetividad de los diarios alimentarios manuales y permitiendo estimar tipos de alimentos, cantidades y potenciales valores nutricionales. Por otro lado, incorporar un módulo de predicción de toxicidad, inicialmente centrado en pesticidas y LD50, como apoyo a la seguridad del consumidor y como vía indirecta para reducir la experimentación animal. Esta doble motivación ha evolucionado hacia una línea de investigación más amplia: el desarrollo de representaciones computacionales que conecten alimentación, química y biología mediante IA.

La relevancia de esta línea se justifica por tres motivos. Primero, las dietas no saludables y la exposición a compuestos asociados al procesado alimentario están relacionadas con enfermedades crónicas de alta prevalencia [1]. Segundo, la evaluación experimental de toxicidad sigue dependiendo en buena medida de ensayos animales, especialmente en indicadores clásicos como LD50, aunque existen trabajos previos de predicción de toxicidad mediante aprendizaje profundo [2]. Tercero, existe una separación metodológica entre los modelos de análisis dietético, los modelos de toxicidad química y los estudios mecanísticos de interacción molecular. La tesis propone conectar estos niveles mediante tres componentes complementarios: la caracterización del consumo y la exposición mediante imágenes alimentarias; la caracterización del peligro químico mediante predicción de toxicidad aguda; y la integración de exposición e interacción molecular a nivel de sujeto.

2. Investigaciones previas y trabajos derivados

La investigación se ha desarrollado progresivamente a través de varios trabajos conectados, que constituyen la base experimental y metodológica de la tesis. A continuación se describen dichos trabajos explicando su aportación específica, su relación con los objetivos generales y el modo en que cada uno contribuye a la construcción metodológica de la tesis.

2.1. NAMs para evaluación del riesgo mediante aprendizaje profundo

El primer trabajo publicado en esta línea fue *New approach methodologies for risk assessment using deep learning* [3]. Este artículo, derivado del programa EU-FORA de *European Food Safety Authority* (EFSA) y de la colaboración con el *Consiglio Nazionale delle Ricerche* (CNR) de Nápoles, planteó el uso de IA, docking molecular y estructuras 3D de proteínas como base para nuevas metodologías de evaluación de riesgo. En dicho trabajo se propuso explotar datos de toxicidad, como LD50, junto con información obtenida de la interacción entre compuestos químicos y proteínas humanas. Este artículo proporciona el marco

conceptual de la tesis: la IA no se plantea como sustituto directo del juicio experto, sino como herramienta para guiar experimentos, priorizar compuestos y reducir ensayos innecesarios.

2.2. NutriConv: aprendizaje multitarea para seguimiento dietético

El segundo eje se concreta en el artículo publicado *Nutriconv: multitask learning framework for digital dietary tracking trained on EFSA's pancake dataset* [4]. NutriConv es una red convolucional ligera diseñada para resolver simultáneamente dos tareas a partir de imágenes: clasificación de alimentos y estimación de su peso. Este trabajo actualiza la idea original del plan de investigación, centrada en segmentación de imágenes y clasificación multietiqueta, hacia una formulación multitarea directamente evaluable y transferible. El modelo combina clasificación y regresión mediante una función de pérdida híbrida y se entrena sobre una adaptación del dataset PANCAKE de EFSA [5], incluyendo máscaras de segmentación generadas con *Segment Anything Model* (SAM) [6] y validación externa con Nutrition5K [7].

2.3. ChenseNet121: predicción multimodal de LD50

El tercer eje corresponde al trabajo *Saving mice: ChenseNet121, a new deep learning architecture for LD50 Toxicity Estimation* [8]. En este artículo se propone ChenseNet121, una arquitectura profunda multimodal basada en DenseNet para la estimación de toxicidad aguda. ChenseNet121 integra imágenes moleculares 2D, descriptores moleculares derivados de la representación SMILES, puntuaciones procedentes del docking y representaciones 3D voxelizadas de complejos proteína–ligando. El modelo se evalúa mediante validación cruzada y se compara con otras arquitecturas basadas en ResNet, Inception y EfficientNet. Además de la regresión directa de LD50, el trabajo introduce una lectura de cribado conservador mediante un umbral de toxicidad de 2000 mg/kg, relevante para priorizar compuestos y reducir ensayos innecesarios cuando el modelo identifica sustancias con menor toxicidad aguda estimada según dicho umbral.

2.4. NutriDock-CRC: exposición dietética e interacción molecular

El cuarto eje se desarrolla en *NutriDock-CRC: integrating dietary exposure and docking-derived molecular interaction profiles for colorectal cancer screening* [9]. Este trabajo extiende la tesis desde el nivel de compuesto químico aislado al nivel de sujeto. A partir de cuestionarios de frecuencia alimentaria, se estiman exposiciones individuales a compuestos relacionados con la dieta y el procesado alimentario. Estos compuestos se vinculan con enzimas CYP relevantes en el metabolismo de xenobióticos mediante docking molecular, y las interacciones resultantes se agregan en función de la exposición de cada sujeto. La representación final combina dieta, propiedades fisico-químicas e interacción proteína–ligando para evaluar señales asociadas a alteraciones colorrectales.

2.5. Docking molecular como motor de datasets

Finalmente, el trabajo *Molecular Docking as a Dataset Engine for Chemical Prediction: From Acute Toxicity Modelling to Diet-Related Cancer Risk Stratification* [10] formaliza una idea transversal de la tesis: el docking molecular no debe entenderse únicamente como una herramienta de cribado virtual, sino también como un motor reproducible de generación de conjunto de datos. Esta perspectiva permite convertir hipótesis bioquímicas, como la selección de proteínas, sitios activos o supuestos de flexibilidad, en descriptores estructurados que pueden fusionarse con imágenes 2D, grafos moleculares, SMILES o variables de exposición.

3. Hipótesis de partida

La tesis se apoya en las siguientes hipótesis:

H1. Las representaciones multimodales que combinan imágenes, estructuras químicas, descriptores fisicoquímicos e información de docking contienen una señal predictiva más rica que las representaciones unimodales para tareas de toxicidad y riesgo alimentario.

H2. El docking molecular puede actuar como puente entre estructura química y mecanismo biológico, siempre que sus salidas se interpreten como proxies computacionales y no como evidencia experimental directa.

H3. El aprendizaje multitarea permite formular de manera más eficiente problemas alimentarios en los que varias salidas están relacionadas, como la clasificación de alimentos y la estimación de cantidades.

H4. La agregación de interacciones molécula–proteína ponderadas por exposición dietética puede capturar patrones a nivel de sujeto que no son visibles cuando la dieta, la química y la biología se analizan por separado.

Como principio transversal, los modelos de IA desarrollados se conciben como herramientas de priorización y apoyo a la decisión, no como sustitutos del juicio experto ni de la validación experimental.

4. Objetivos

4.1. Objetivo principal

El objetivo principal de la tesis es diseñar, implementar y evaluar metodologías basadas en inteligencia artificial que integren análisis de imágenes alimentarias, modelado molecular, datos toxicológicos y datos de exposición dietética para apoyar la evaluación del riesgo alimentario, con especial atención a la predicción de toxicidad, el seguimiento dietético y la reducción de experimentación animal.

4.2. Subobjetivos

O1. Revisar el estado del arte en aprendizaje profundo aplicado al análisis de imágenes de comida, predicción de toxicidad, QSAR, docking molecular y NAMs.

O2. Diseñar modelos de visión artificial para seguimiento dietético capaces de estimar simultáneamente categoría alimentaria y cantidad aproximada de alimento.

O3. Construir y adaptar conjuntos de datos alimentarios con información de imagen, clase, peso, máscaras y validación externa, priorizando la reproducibilidad.

O4. Generar conjuntos de datos químicos multimodales que integren descriptores moleculares derivados de SMILES, imágenes 2D, afinidades de docking y representaciones 3D de complejos proteína–ligando.

O5. Diseñar arquitecturas profundas multimodales para la estimación de LD50 y evaluar su utilidad como herramienta de cribado conservador.

O6. Explorar el uso de docking molecular como fuente de descriptores mecanísticos para modelos de toxicidad y riesgo dietético.

O7. Construir representaciones a nivel de sujeto que agreguen exposición dietética, propiedades químicas e interacción con proteínas metabólicas relevantes.

O8. Evaluar la utilidad de estas representaciones en escenarios realistas con conjuntos de datos pequeños, desequilibrio entre clases y objetivos de priorización.

O9. Integrar los resultados obtenidos en un marco metodológico reproducible, interpretable y transferible a futuras tareas de seguridad alimentaria.

5. Metodología y plan de trabajo

La metodología de la tesis se organiza en cinco paquetes de trabajo interrelacionados, que se describen a continuación.

5.1. PT1. Revisión, contextualización y definición de problemas

Este paquete incluye la revisión de literatura sobre análisis automático de alimentos, redes convolucionales, aprendizaje multitarea, predicción de toxicidad, docking molecular, QSAR, modelos 3D proteína–ligando y NAMs, tomando como referencia trabajos previos en segmentación de imágenes alimentarias y toxicidad computacional [2] [11].

5.2. PT2. Análisis automático de dieta mediante imágenes

Este paquete desarrolla modelos de visión artificial para seguimiento dietético. La metodología incluye adaptación del dataset PANCAKE, generación o incorporación de máscaras de segmentación, diseño de una arquitectura ligera

multitarea, definición de una función de pérdida híbrida para clasificación y regresión, ajuste del parámetro de ponderación entre tareas, comparación frente a modelos monotarea y validación externa con un conjunto de datos independiente para comprobar su capacidad de generalización. El resultado principal de este paquete es NutriConv.

5.3. PT3. Modelado molecular y predicción de toxicidad LD50

Este paquete construye el módulo de toxicidad de la tesis. Se parte de compuestos químicos con valores LD50 y se generan varias modalidades de entrada: imágenes 2D, descriptores SMILES, afinidades de docking y tensores 3D voxelizados derivados de complejos proteína–ligando. Las arquitecturas se evalúan mediante validación cruzada, métricas de regresión y análisis de cribado binario. El resultado principal de este paquete es ChenseNet121, inspirado parcialmente en trabajos que explotan campos de interacción 3D y energías moleculares para predicción computacional [12], cuya arquitectura se resume en la Fig. 1. La figura muestra cómo el modelo integra una rama convolucional 2D para imágenes moleculares, una rama convolucional 3D para volúmenes voxelizados derivados de docking y una rama densa para descriptores moleculares y afinidad.

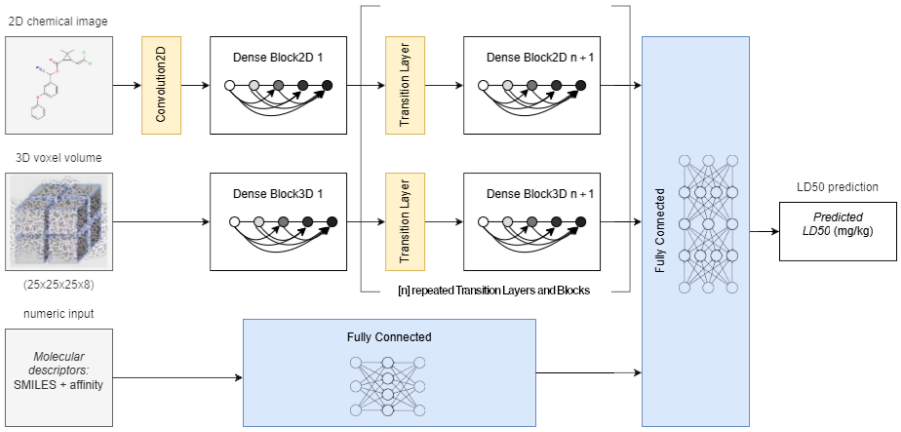


Figura 1. Arquitectura general de ChenseNet121 para predicción de LD50. El modelo integra imágenes químicas 2D, volúmenes 3D voxelizados de complejos proteína–ligando y entradas numéricas basadas en descriptores moleculares y afinidad de docking. Figura adaptada de [8].

5.4. PT4. Docking molecular como generador de representaciones

Este paquete define el papel metodológico del docking molecular dentro de la tesis. Las proteínas se seleccionan a partir de hipótesis biológicas o toxicológicas, habitualmente consultando repositorios estructurales como el Protein

Data Bank [13], mientras que los ligandos pueden obtenerse de bases químicas como PubChem [14]. Posteriormente, los ligandos se preparan y estandarizan, y las simulaciones se ejecutan de forma reproducible mediante herramientas de docking como AutoDock Vina [15]. Las salidas del docking se transforman en puntuaciones, mapas de interacción, tensores voxelizados o resúmenes espaciales. Este paquete proporciona la conexión entre mecanismo molecular y aprendizaje automático.

5.5. PT5. Modelado exposición–molécula–proteína a nivel de sujeto

Este paquete integra datos dietéticos y moleculares. Las exposiciones individuales a compuestos relacionados con la dieta se combinan con descriptores de ligandos e interacciones con proteínas, especialmente enzimas CYP relevantes en metabolismo de xenobióticos. La agregación ponderada por exposición genera representaciones a nivel de sujeto, evaluadas mediante modelos tabulares, validación cruzada, análisis de ablación y formulaciones de clasificación multiclase o binaria. El resultado principal de este paquete es NutriDock-CRC. El flujo metodológico completo se resume en la Fig. 2, donde se observa cómo los datos dietéticos y de cohorte se transforman en bloques de características (Diet, MolAgg, ProtAgg y Combined) para su posterior evaluación predictiva.

6. Planificación temporal e hitos

La tesis se desarrolla a tiempo parcial desde septiembre de 2022. La planificación se interpreta, por tanto, en un horizonte de cinco cursos académicos, con solapamiento entre revisión, desarrollo metodológico, experimentación y escritura, tal como se resume en la Tabla 1.

7. Resultados preliminares y artículos de la tesis

Los resultados obtenidos hasta el momento muestran que la tesis ha evolucionado desde un planteamiento inicial centrado en análisis de imágenes de comida hacia una línea integrada de IA para seguridad alimentaria. La Tabla 2 resume los trabajos principales.

En la línea de toxicidad, ChenseNet121 obtuvo los menores errores entre las configuraciones profundas evaluadas. Además de la regresión directa de LD50, el análisis mediante un umbral de 2000 mg/kg permite estudiar su utilidad como herramienta de priorización conservadora. Este uso no pretende sustituir la validación experimental, sino apoyar la selección de compuestos para estudios posteriores y contribuir a reducir ensayos innecesarios.

En la línea de seguimiento dietético, NutriConv demuestra que una arquitectura compacta puede aprender simultáneamente clasificación y estimación de peso. Aunque la clasificación puede ser inferior a modelos especializados, el enfoque multitarea ofrece un compromiso práctico entre precisión, eficiencia y posible despliegue en dispositivos móviles o wearables.

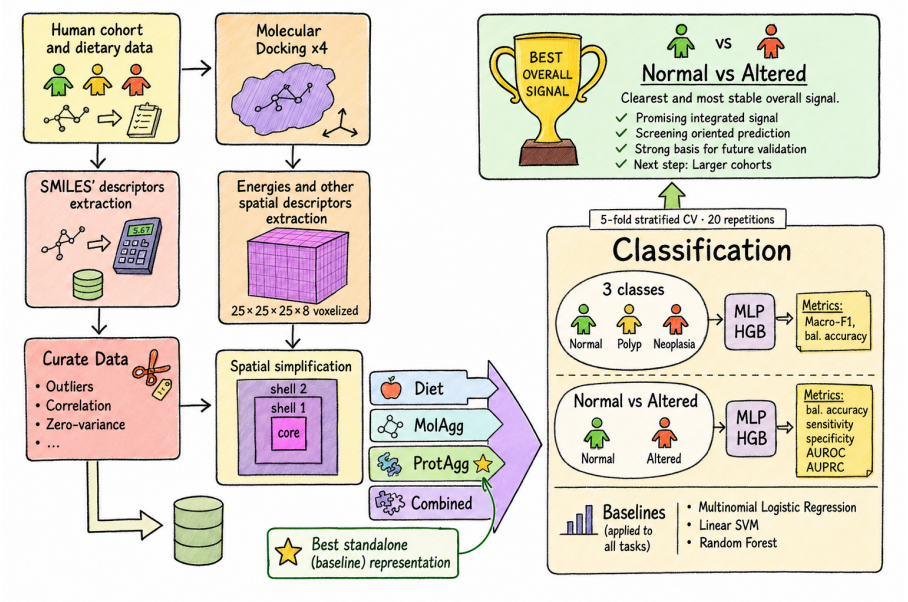


Figura 2. Flujo general de NutriDock-CRC. La representación propuesta combina datos de cohorte y exposición dietética, extracción de descriptores SMILES, docking molecular frente a cuatro proteínas, simplificación espacial de los descriptores de interacción y construcción de bloques de características para clasificación multiclase y cribado *Normal* frente a *Alterado*. Figura adaptada de [9].

Tabla 1. Planificación temporal de la tesis doctoral.

Periodo	Tareas e hitos principales
2022–2023	Revisión del estado del arte; definición del plan de investigación; prototipos iniciales en análisis de imágenes alimentarias, clasificación, segmentación y predicción de toxicidad.
2023–2024	Desarrollo de la línea NAMs durante el programa EU-FORA; primeros experimentos con docking, LD50, arquitecturas convolucionales y combinación de descriptores; publicación del artículo EFSA Journal.
2024–2025	Desarrollo de NutriConv y consolidación del eje de seguimiento dietético; adaptación de conjunto de datos, aprendizaje multitarea y validación externa; desarrollo avanzado de ChenseNet121 para predicción multimodal de LD50.
2025–2026	Desarrollo de NutriDock-CRC; integración de exposición dietética, descriptores moleculares y docking con CYP; análisis de ablación, evaluación en escenarios desbalanceados e interpretación de resultados; formalización del docking como motor de conjunto de datos.
2026–2027	Integración final de los resultados; ampliación de validaciones cuando sea posible; análisis de interpretabilidad; cierre de manuscritos; redacción y defensa de la tesis doctoral.

En la línea exposición–molécula–proteína, NutriDock-CRC relaciona la exposición dietética individual a compuestos químicos con información molecular y de interacción proteína–ligando para explorar su utilidad en la priorización de alteraciones colorrectales. El estado de cada sujeto tras un año de seguimiento se clasifica como Normal, Pólipo o Neoplasia. El problema se formuló inicialmente como una tarea de clasificación multiclase entre estas tres categorías. Sin embargo, el reducido número de casos de Neoplasia limita la estimación robusta de las fronteras de decisión. Por ello, se evaluó también un escenario binario, Normal frente a Alterado, donde Alterado agrupa los casos de Pólipo y Neoplasia. Esta reformulación resulta más coherente con un uso preliminar de priorización o cribado, cuyo objetivo no es realizar un diagnóstico clínico, sino identificar perfiles con mayor probabilidad de presentar alteraciones. En este contexto, los resultados corregidos muestran que la representación basada en interacciones proteína–ligando ponderadas por exposición (ProtAgg) contiene una señal predictiva relevante y obtiene el mejor comportamiento entre los bloques individuales en las principales métricas sensibles al desequilibrio, mientras que la integración completa de características ofrece una mejora más modesta y dependiente del modelo.

8. Relevancia científica y transferencia

La relevancia científica de la tesis reside en la integración de niveles que normalmente se estudian por separado. El análisis de imágenes alimentarias permi-

Publicación o manuscrito	**Estado**	**Contribución a la tesis**
New approach methodologies for risk assessment using deep learning [3]	Publicado en revista JCR	NAMs: marco conceptual para el uso de aprendizaje profundo, docking molecular y estructuras 3D en evaluación de riesgo, con el objetivo de orientar la priorización de compuestos y reducir experimentación animal.
NutriConv: multitask learning framework for digital dietary tracking trained on EFSA's pancake dataset [4]	Publicado en revista JCR	NutriConv: modelo multitarea ligero para clasificación de alimentos y estimación de peso a partir de imágenes, dentro del eje de seguimiento dietético automatizado.
Saving mice: ChenseNet121, a new deep learning architecture for LD50 toxicity estimation [8]	Enviado a revista JCR, en revisión	ChenseNet121: arquitectura multimodal para estimación de LD50 mediante la combinación de imágenes 2D, descriptores moleculares, docking y representaciones 3D.
NutriDock-CRC: integrating dietary exposure and docking-derived molecular interaction profiles for colorectal cancer screening [9]	Enviado a revista JCR, en revisión	NutriDock-CRC: pipeline exposición–molécula–proteína para integrar dieta individual, descriptores moleculares e interacciones con CYP en un escenario de cribado colorrectal.
Molecular docking as a dataset engine for chemical prediction: from acute toxicity modelling to diet-related cancer risk stratification [10]	Aceptado en congreso nacional	Docking as Dataset Engine: formalización metodológica del docking molecular como generador reproducible de descriptores para modelos predictivos.

Tabla 2. Trabajos que articulan la tesis doctoral.

te estimar qué se consume y en qué cantidad; el modelado molecular permite caracterizar cómo determinados compuestos pueden interactuar con proteínas relevantes; y el aprendizaje automático permite transformar estas fuentes en representaciones predictivas. La originalidad del trabajo se sitúa en la conexión entre estos niveles, especialmente en el uso del docking como fuente de datos para modelos de IA y en la agregación de interacciones moleculares en función de la exposición dietética individual.

Desde el punto de vista de transferencia, la tesis puede contribuir en tres direcciones. La primera es el apoyo al seguimiento dietético mediante herramientas automáticas más objetivas que los diarios manuales. La segunda es la priorización de compuestos químicos en estudios de toxicidad, reduciendo costes y ayudando a seleccionar experimentos más informativos. La tercera es la cons-

trucción de modelos de cribado o estratificación de riesgo que puedan orientar futuras investigaciones en nutrición, toxicología y salud pública.

La tesis mantiene una interpretación prudente de los modelos. Las predicciones de IA y los scores de docking no se presentan como verdades biológicas ni como diagnósticos clínicos, sino como señales computacionales útiles para priorizar, generar hipótesis y apoyar la toma de decisiones bajo incertidumbre.

9. Limitaciones y trabajo pendiente

Las principales limitaciones actuales son el tamaño de los conjuntos de datos, la heterogeneidad de las fuentes de datos y la necesidad de validación externa. En toxicidad, los valores LD50 proceden de fuentes históricas y pueden contener variabilidad experimental. En docking, las afinidades y poses dependen de supuestos de preparación molecular, selección de proteínas, definición de cajas y flexibilidad. En NutriDock-CRC, el bajo número de casos de neoplasia limita la robustez de los modelos multiclase. En NutriConv, el tamaño y diversidad del conjunto de datos condicionan la generalización a escenarios reales.

El trabajo pendiente se centrará en reforzar la reproducibilidad, ampliar validaciones, mejorar la interpretabilidad y cerrar la integración narrativa de la tesis. En particular, se explorarán análisis de importancia de variables para identificar compuestos, proteínas o grupos de descriptores relevantes, así como estrategias para presentar los modelos como herramientas de apoyo y no como sistemas automáticos de decisión.

10. Conclusiones

Esta tesis propone un marco de inteligencia artificial para la evaluación del riesgo alimentario que combina aprendizaje profundo, visión artificial, docking molecular y modelado de exposición. Su contribución principal es construir representaciones que conecten la alimentación observada, los compuestos químicos implicados y su posible interacción con proteínas relevantes. Los trabajos desarrollados hasta la fecha muestran que esta integración es viable y que puede aportar valor en tres escenarios: seguimiento dietético, predicción de toxicidad aguda y priorización de riesgo dietético. En conjunto, la tesis aspira a contribuir a una evaluación del riesgo más reproducible, con mayor fundamento mecanístico y alineada con la reducción de la experimentación animal.

Referencias

1. Afshin, A., Sur, P.J., Fay, K.A., Cornaby, L., Ferrara, G., Salama, J.S., et al.: Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 393(10184), 1958–1972 (2019).
2. Mayr, A., Klambauer, G., Unterthiner, T., Hochreiter, S.: DeepTox: toxicity prediction using deep learning. Frontiers in Environmental Science 3, 80 (2016).

3. Junquera, E., Díaz, I., Montes, S., Febbraio, F.: New approach methodologies for risk assessment using deep learning. EFSA Journal 22(S1), e221105 (2024).
4. Junquera, E., Rico, N., Díaz, I., González, S., Remeseiro, B.: Nutriconv: multitask learning framework for digital dietary tracking trained on EFSA's pancake dataset. Neural Computing and Applications 37, 23833–23862 (2025).
5. Ocké, M., de Boer, E., Brants, H., van der Laan, J., van Niekerk, M., van Rossum, C., Temme, L., Freisling, H., Nicolas, G., Casagrande, C., et al.: PANCAKE–Pilot study for the Assessment of Nutrient intake and food Consumption Among Kids in Europe. EFSA Supporting Publications 9(9), EN-339 (2012). https://doi.org/10.2903/sp.efsa.2012.EN-339
6. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al.: Segment Anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4015–4026 (2023).
7. Thames, Q., Karpur, A., Norris, W., Xia, F., Panait, L., Weyand, T., Sim, J.: Nutrition5K: Towards automatic nutritional understanding of generic food. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8903–8911 (2021).
8. Junquera, E., Díaz, I., Remeseiro, B., Febbraio, F.: Saving mice: ChenseNet121, a new deep learning architecture for LD50 toxicity estimation. Manuscrito en revisión en Journal of Cheminformatics (2026).
9. Junquera, E., Zapico, A., González, S., Remeseiro, B., Díaz, I.: NutriDock-CRC: integrating dietary exposure and docking-derived molecular interaction profiles for colorectal cancer screening. Manuscrito en revisión en Journal of Cheminformatics (2026).
10. Junquera, E., Díaz, I., Remeseiro, B.: Molecular Docking as a Dataset Engine for Chemical Prediction: From Acute Toxicity Modelling to Diet-Related Cancer Risk Stratification. Aceptado en SEIO (2026).
11. Aguilar, E., Nagarajan, B., Remeseiro, B., Radeva, P.: Bayesian deep learning for semantic segmentation of food images. Computers and Electrical Engineering 103, 108380 (2022).
12. Hassan-Harrirou, H., Zhang, C., Lemmin, T.: RosENet: improving binding affinity prediction by leveraging molecular mechanics energies with an ensemble of 3D convolutional neural networks. Journal of Chemical Information and Modeling 60(6), 2791–2802 (2020).
13. Burley, S.K., Bhikadiya, C., Bi, C., et al.: RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Research 49(D1), D437–D451 (2021). https://doi.org/10.1093/nar/gkaa1038
14. Kim, S., Chen, J., Cheng, T., et al.: PubChem 2023 update. Nucleic Acids Research 51(D1), D1373–D1380 (2023).
15. Trott, O., Olson, A.J.: AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31(2), 455–461 (2010).