EATITALL: An Artificial Intelligence Platform for Designing and Developing New Healthy Products in the Agri-Food Sector

Yoan Gutiérrez[1], Rafael Muñoz[1], Andrés Montoyo[1], Miquel Canal-Esteve[1], Mar Bonora[1] and Kamila Cheikh-Moussa[1,2]

[1]University of Alicante, Spain. Crta. San Vicentte del Raspeig s/n, Alicante, Spain
[2]Dr Balmis General University Hospital-ISABIAL, Alicante, Spain

Abstract

This project proposes industrial research leveraging Artificial Intelligence technologies, particularly in the field of Natural Language Processing, to optimise the development of ingredients, formulations, and functional foods. It will do so through a platform designed to identify consumers' nutritional needs by learning from clinical Gold Standard data. The EATITALL tool will apply this clinically validated knowledge to help agri-food companies better understand the nutritional requirements of the general population. The insights gained will support the development and marketing of product lines aimed at promoting healthier diets and preventing cardiometabolic diseases. The monitoring of physiological parameters and personalised follow-up during the clinical trial phase will enable the validation of newly developed formulations. Based on the results, these products may enter a cycle of continuous improvement, serving as a reference model for other companies in the agri-food sector.

Keywords

healthy products, clinical and nutritional profile, pattern extraction, Artificial Intelligence, Natural Language Procesing, EHRs,

1. Introduction

The recent demographic evolution in Spain shows positive aspects such as the increase in life expectancy or the generalised population growth, but also negative ones, such as the increase in health needs and, consequently, the increase in the costs and sustainability of the public health system.

A longer life expectancy means that health problems develop over more years. In addition, unhealthy food, sedentary lifestyles, stress and other negative habits that have become established in the population of all ages, is leading people to be less healthy than before; therefore, it is essential to seek innovative solutions to this situation. According to the INE study "Population projections between 2022 and 2072", it is estimated that the population in Spain will be 52.9 million people in 2072, of which a high percentage will be elderly people. According to the study, the population between 20 and 64 years of age, which currently accounts for 60.7% of the total, would represent 53.1% in 2052 and the percentage of the population aged 65 and over, which currently stands at 20.1% of the total, would reach 30.4% around 2050.

Another relevant fact of the study is that the centenarian population (those aged 100 years or more) would increase from 14,287 people today to 226,932 in 2072. On the other hand, healthy eating must be one of the milestones of the future, changing people's awareness to preserve their health, the CSIC study of 2021 on "New Trends in food production and consumption" points out that: sustainability and healthy eating are the axes around which food consumption will revolve in the coming decades, and that, for this, more efficient production must be promoted in which fewer resources are used and

SEPLN 2025: 41st International Conference of the Spanish Society for Natural Language Processing, Zaragoza, Spain, 23-26 September 2025.

ygutierrez@dlsi.ua.es (Y. Gutiérrez); rafael@dlsi.ua.es (R. Muñoz); montoyo@dlsi.ua.es (A. Montoyo); mikel.canal@ua.es (M. Canal-Esteve); mar.bonora@ua.es (M. Bonora); kamila.cheikh@ua.es (K. Cheikh-Moussa)
0000-0002-4052-7427 (Y. Gutiérrez); 0000-0001-8127-9012 (R. Muñoz); 0000-0002-3076-0890 (A. Montoyo); 0009-0006-8022-5534 (M. Canal-Esteve); 0000-0002-1561-6903 (K. Cheikh-Moussa)

processes with less environmental impact are developed. They also point out the need to take advantage of by-products (secondary materials generated unintentionally during food production that can still be reused) and waste and to develop more sustainable packaging systems, in line with the circular economy. Lastly, they suggest changing consumption habits and working towards greater integration of the food chain. For all of this, they consider precision nutrition, the development of foods aimed at populations with special needs, biotechnology, the use of Big Data and the support of artificial intelligence to be fundamental.

These trends in food consumption show us that it is necessary to work on 'precision nutrition'. This is the case of the proposed project, which aims to develop a platform for advanced analysis of text or unstructured content based on Natural Language Processing (NLP) techniques, for the extraction of clinical indicators and markers of healthy consumption in the context of the main cardiometabolic diseases of the general population and as a support tool for the design and launch of new products in the field of healthy food, adapted and personalised to the end customer, using the food vehicle as a preventive and/or curative element in the context of different pathologies.

2. Description of the State of Technology in Spain and Abroad

Cardiometabolic diseases are the main drivers of hospital care demand. Sedentary lifestyles, low intake of fresh produce, and high consumption of ultra-processed foods contribute to the rising prevalence of conditions such as hypertension, endocrine disorders, obesity, and diabetes. This trend is evident in the Valencian Community, where diabetes morbidity (64 per 100,000 inhabitants) and mortality (21.2%) exceed national averages. The province of Alicante follows a similar pattern, with above-average hospital admissions for hypertension and diabetes.

This project seeks to apply Natural Language Processing (NLP) and Artificial Intelligence (AI) to clinical data from patients with metabolic diseases at the General University Hospital Dr. Balmis in Alicante. The goal is to identify nutritional needs using additional data sources, enabling the design of tailored nutritional formulations and healthier food products.

The integration of clinical diagnoses with patients' genetic predispositions, dietary habits, and socio-economic factors provides a framework for developing both broad-spectrum and personalized nutritional solutions. Broad formulations would address the general needs of patients at cardiometabolic risk, while personalized products would cater to those with specific clinical conditions requiring targeted interventions. Nutrigenomics—nutrition based on genetic profiles—offers a promising path to improving health outcomes and quality of life through individualized dietary approaches.

This field is rapidly expanding, gaining traction in both research and industry. As an innovative and disruptive company, Juan y Juan Industrial SL (VICKY FOODS) embraces this strategic direction, aiming to develop personalized nutritional products aligned with increasing consumer demand for healthier options. These preparations will serve as targeted dietary complements to support clinical treatment and address specific deficiencies or health risks identified in medical diagnoses.

AI is at the forefront of healthcare digitization. One of its most promising applications is the analysis of electronic health records (EHRs), which offers new opportunities to enhance patient care through data-driven insights. Some of the applications of NLP to EHRs include identifying diagnoses and procedures, extracting information from laboratory and radiology test reports, classifying patients according to their disease risk, and detecting patterns of medication use. However, it is extremely difficult to digitise habits: most healthcare professionals still report their consultations, follow-ups and diagnoses by writing free text, i.e. unstructured. And unstructured data are data that are difficult to process. NLP technologies are used to make use of this extremely valuable information hidden in patients' medical records. Among other actions, they are algorithms capable of filtering out words that do not provide meaning, identifying words with meaning, compound expressions, labelling relevant terms, either by frequency of use or by contrast with an external knowledge model, etc. Applied to a

medical report, and supported by medical standards and ontologies such as SNOMED-CT[1], ICD-10[2], etc., it allows problems, symptoms, diagnoses, medications, specialities, etc. to be extracted from the free text notes contained in these reports. In this way, there are already many initiatives such as [1, 2, 3, 4] to automatically extract structured information from each EHR by applying NLP, with enormous value for healthcare centres as it allows them to optimise their routine clinical activity and improve clinical management in general.

Among these initiatives we can highlight those carried out in the hospitals themselves through research projects such as DeepEMR(TIN2017-87548-C2-1-R) [5], carried out in the Community of Madrid and whose objective is the development of NLP techniques and deep learning methods for the analysis of unstructured information from EHRs, with the ultimate aim of reducing the cost, in terms of time and resources, of epidemiological studies; the development of personalised models of care for multi-pathological patients supported by artificial intelligence in the Basque Country [3]; or in the Valencian Community, with projects such as the one carried out at ISABIAL[4] for the analysis of clinical data from patients affected by primary brain tumours, together with the implementation of algorithms based on Artificial Intelligence. At the national level, all these initiatives would have a place in the Infrastructure for Precision Medicine associated with Science and Technology (IMPaCT)[6] , through which the government aims to establish the pillars to facilitate the effective deployment of Precision Medicine in the National Health System through validated clinical data management tools (IMPaCT-Data Strategic Axis).

Recent studies show that artificial intelligence can enhance dietary assessment and personalization. [7] introduced a system using wearable data (glucose, heart rate, motion) with machine learning to estimate intake. [8] used explainable AI to predict dietary habits in Spain based on lifestyle and demographics. These works highlight the potential of AI in precision nutrition.

On the other hand, from the industry, companies such as the Spanish startup Savana[5], whose product is a platform to support medical decision-making based on the extraction of unstructured information from EHRs using AI and Big Data techniques, in an automatic and anonymised manner. This technology converts clinical information into scientific knowledge, to make it available to the medical and research community. However, it does not address the extraction of information relating to the health and nutrition of the population, on the one hand, nor the development of a tool trained on the basis of nutritional information from EHRs capable of learning to perform the task from other sources of information with an impact on the development of products in the agri-food industry.

With regard to the development of tools for optimising the design of food products, there are APIs such as FoodDatabaseAPI[6] that allow information on nutrition and diet to be searched for using different criteria (nutritional analysis of food, search for ingredients or calculation of health parameters) from the company's own food database. Recent research has explored the integration of AI in personalized nutrition and food formulation (e.g., [9, 10]). However, few initiatives combine NLP and clinical EHR analysis to inform functional food development, which highlights the novelty of our approach.

This project proposes to go a step further and use NLP techniques to generate a tool that, through Deep Learning, correlates clinical data ("gold standard") and builds rules that allow it to learn to "know" the nutritional needs of the population in order to subsequently validate its use by the agri-food industry in the region.

3. Expected Impact and Project Objetives

This project proposes an industrial research based on technologies developed within Artificial Intelligence, especially in the field of NLP, with the aim of optimising the development of ingredients,

[1] https://www.snomed.org/
[2] https://ais.paho.org/classifications/chapters/
[3] https://basquehealthcluster.org/en/ai-for-digital-health-and-personalized-medicine/
[4] https://isabial.es/investigacion/proyectos-competitivos/
[5] https://savanamed.com/
[6] https://www.nubentos.com/es/api/fooddatabaseapi/

formulations and functional foods, through a platform that allows the identification of consumers' nutritional needs based on learning from the clinical Gold Standard. The EATITALL tool will apply this learning based on clinical evidence in order for agri-food companies to discover the nutritional needs of the general population. The knowledge extracted will enable the marketing of product lines that favour the prevention of chronic diseases and a healthier diet. The monitoring of physiological parameters and the individualised follow-up during the clinical trial will allow the validation of the new formulated developments which, depending on the results, can enter a cycle of continuous improvement, so that they can serve as a model for other companies in the agri-food sector.

Main objective:

- To develop an AI-based platform, powered by NLP, that identifies consumers' nutritional needs by extracting clinically relevant patterns from Electronic Health Records (EHRs). These insights will drive the design and development of personalized functional food products, which will be validated through clinical trials and tailored to specific health profiles.

Specific objectives:

- To analyse the nutritional needs of people with cardiometabolic disease (CMD) treated in the department of the Hospital General Universitario de Alicante through the analysis of EHR.
- Research and develop advanced text analysis algorithms using NLP techniques, which allow the extraction of knowledge and the semantic exploration of nutritional recommendations from EHRs.
- To build a platform for storing clinical results and monitoring the detected nutritional needs of diabetics with cardiometabolic disorders (D-CMD).
- Detect clinical profiles based on the identified nutritional needs and develop an ad-hoc food product.
- To develop an intervention study adjusted to the detected profiles and their impact on the clinical evolution of people with (D-CMD).
- To disseminate the platform developed in the agri-food sector to promote the personalisation of food products according to consumption trends and the nutritional needs of the population.

4. Consortium and Workflow

This project is being carried out by a consortium involving several entities such as:

- as clinical data provider and tester: ISABIAL (Institute of Health and Biomedical Research of Alicante);
- as Software developer: the company DIFUSION S.L.
- as Food Manufacturer: Juan & Juan S.L.(Vicky Foods enterprise);
- as NLP, Machine Learning and Data Analitic expert: the University of Alicante.

Depending on the role played by each entity, these appear in different modules, as shown in Figure 1.

The project workflow follows the structure shown in Figure 1, with the following stages:

1. **EHRs:** EHRs are provided by health centres.
2. **AI:** Artificial Intelligence technologies are applied to the EHRs to extract relevant features and create clinical-nutritional profiles.
3. **MPP:** Manufacturing of products aimed at solving nutritional deficiencies of patients analysed in the EHR.
   - **3.1 HVAL:** Health experts monitor and validate clinical profiles in order to establish nutritional characteristics for product manufacturing.

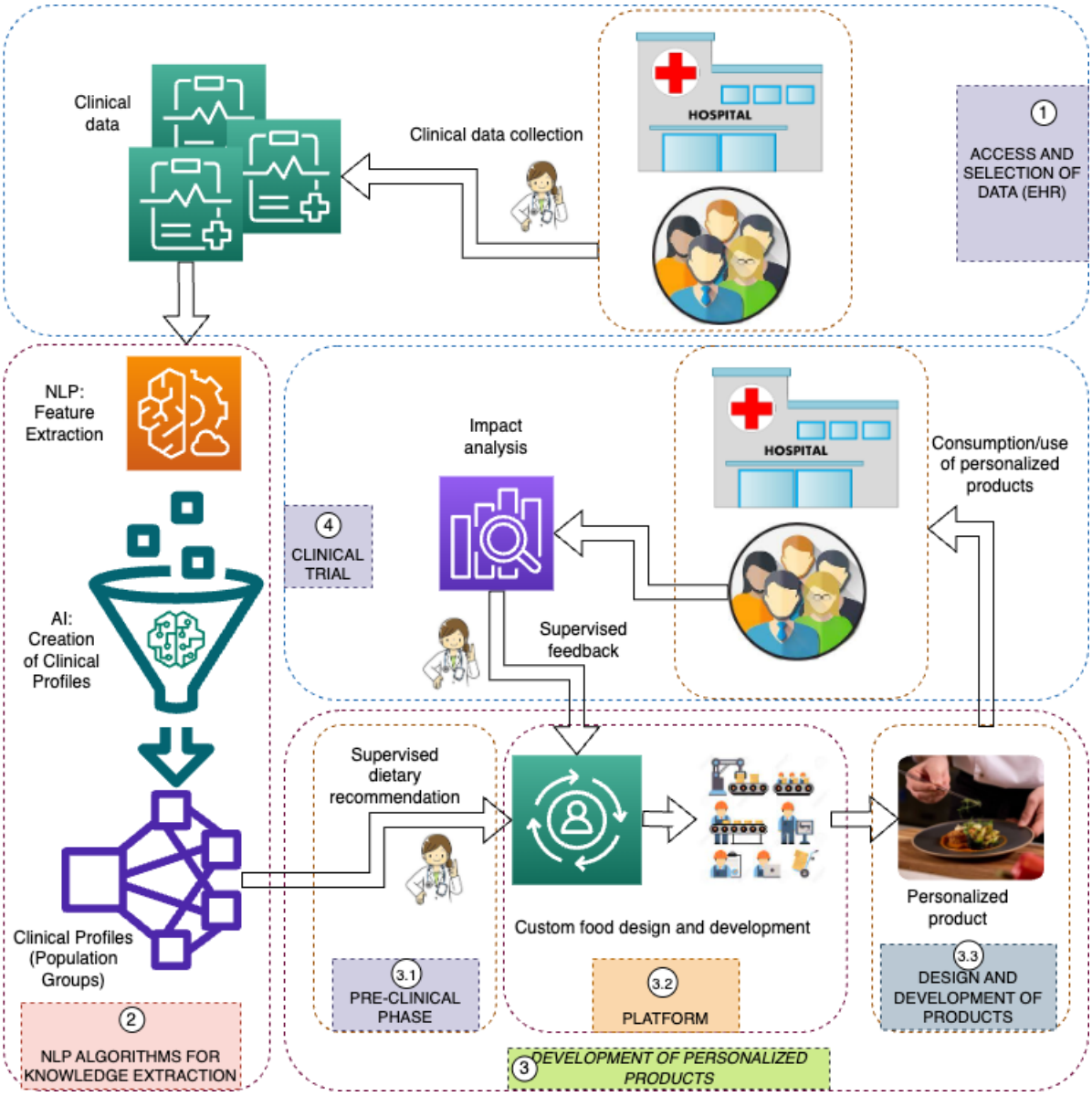

Figure 1: Workflow for the development of personalized nutrition products based on clinical data: from data extraction to product design and impact evaluation.

3.2 DP: Design of patient-tailored food products.

3.3 MP: Manufacturing of tailor-made food products.

4. **PS:** Food products are provided to patients in hospitals and their impact is analysed.

5. **IMP:** The result of the impact study is reused to repeat step 3.1 whenever necessary.

5. Preliminar Results

The project is currently in progress; however, some preliminary results can be reported concerning the planned stages above mentioned:

1. EHRs: 783 EHRs are supplied by health centers, each consisting of 111 core variables (demografic data, diagnosis, nutritional recommendations...). These clinical records have been anonymized and formatted to ensure proper processing by the platform.

2. AI: To achieve this, a data processing procedure is first applied to the core data, incorporating NLP and nutritional rule-based algorithms to enrich it with enhanced characteristics. The rules and algorithms have been developed based on clinic guidelines. The data transformation is as figure 2 shows.

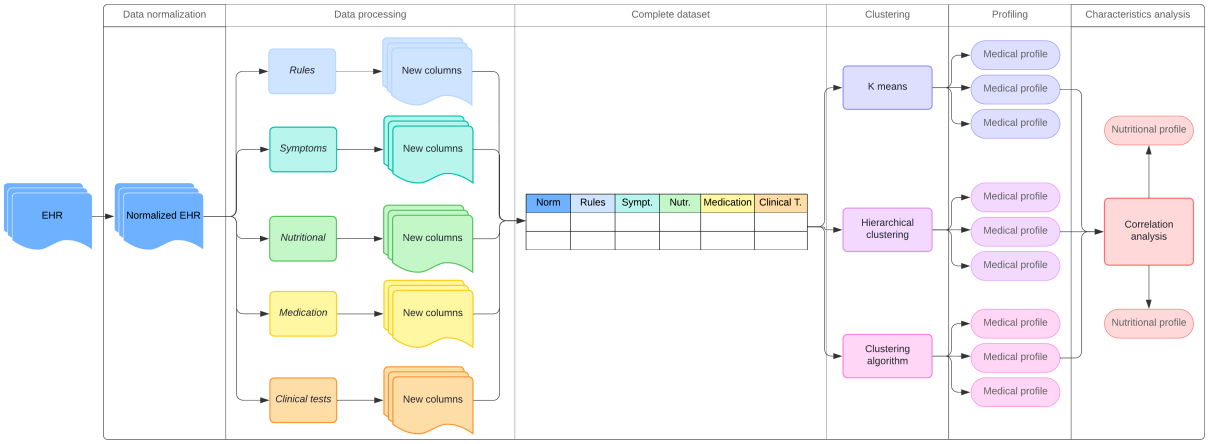

Figure 2: From EHR to Profiles.

Then, different algoritms are applied to create clinical-nutricional profiles (CNP), ie. group of patients grouped by clinical and nutritional similarities. Those clinical groups are:

- CAREME (Clinical Assessment and Risk Evaluation in Metabolic Diseases). Applying rules.
- CKM (Clinical Knowledge Model) for the clinical-nutritional profile in diabetes. Applying rules.
- and CP (Clustering Profiles), a third group freely created by AI clustering algorithms. Appling K-means.

Within each of these three sets of clinical-nutritional profiles, both statistical techniques are applied to assess the relevance of the characteristics and clustering methods are used to group patients with similar traits. By leveraging both traditional AI algorithms and non-traditional approaches like LLMs, clinical-statistical reports are generated from these profiles. These reports provide insights into the clinical status of the patient group (including diagnoses, medication, etc.), nutritional recommendations (such as dietary supplements), dietary advice based on a taxonomy developed by the health partner, and dietetic suggestions that serve as support for the creation of a food product. Table 1 presents an example of how AI language models can simplify the key characteristics of a clinical profile, presenting them as a textual report.

Additionally, based on prompt engineering techniques, functionalities for the automatic recommendation of dietary product traits have been developed, integrating the following:

- Relevant data from the patient group, including primary diagnoses, nutritional recommendations, and dietary advice, all aligned with labels of actual nutritional products.
- Examples of dietary products and a set of alignment criteria between nutritional advice.

The resulting product proposal will be initially evaluated by the nutritional health partner and subsequently by the manufacturing entity to ensure it meets the specific needs of each group and is feasible in terms of its formulation.

3. MPP Currently, this task is ongoing. Health experts are monitoring and validating clinical profiles to establish the nutritional characteristics required for product manufacturing. This process is supported by the technologies developed in the AI stage, which help select the most suitable food components for each profile group and subgroup, ensuring highly personalized dietary food products.

The next sub-stage will focus on the design of patient-tailored food products and the manufacturing of these customized products. Once completed, we expect to achieve the manufacturing of products specifically aimed at addressing the nutritional deficiencies of patients analyzed in the EHR.

Table 1
AI-generated textual example of a clinical profile.

Summary of the Clinical Profile of the Group:
The patient group consists of 432 individuals with primary diagnoses related to non-insulin-dependent diabetes mellitus (NIDDM) and its complications, including chronic kidney disease, chronic cardiovascular disease, and hypertension. Most patients have uncomplicated NIDDM (20.7%), followed by those with NIDDM and ophthalmic complications (20.5%). The most common associated diagnoses include essential hypertension, chronic kidney disease, and chronic cardiovascular disease.
Nutritional Recommendation:
Based on the provided information, a diet rich in fiber, with complex carbohydrates and low in saturated fats, is recommended. It is also important to include sources of lean protein and to avoid excessive salt consumption. The administration of nutritional supplements such as cyanocobalamin, iron, and folic acid may be necessary for patients with deficiencies or at risk of deficiency.
Dietary Advice:
• **High-Fiber Diet**: It is recommended to include fiber-rich foods such as fruits, vegetables, and whole grains. • **Lean Proteins**: The consumption of lean proteins is important for maintaining muscle mass and kidney health. • **Complex Carbohydrates**: It is advisable to include complex carbohydrates such as those found in whole grains and fruits. • **Low in Saturated Fats**: The consumption of saturated fats should be limited to reduce the risk of heart disease. • **Avoid Salt**: Excessive salt intake may worsen hypertension and increase the risk of chronic kidney disease.

4. PS Once the personalized products are produced, they will be supplied to patients in hospitals, and their impact will be analyzed.

5. IMP The results of the impact study will be used to refine the food products as needed, until the most suitable products for each patient profile are achieved.

6. Conclusions

This project leverages AI technologies to enhance the development of ingredients, formulations, and functional foods within the agri-food industry. AI is employed to identify consumers' nutritional needs through an advanced analysis of EHRs, helping agri-food companies understand the broader requirements of the population.

The validation of new formulations is ensured through the monitoring of physiological parameters and personalized follow-up during clinical trials, allowing for continuous improvements. This initiative fosters the creation of new product lines that address nutritional needs, contribute to preventing chronic diseases, and promote healthier diets. The evidence-based scientific approach drives innovation within the agri-food industry.

Currently, the project is in an intermediate stage. Initially, it focused on capturing both technical and functional requirements, as well as collecting EHRs. Additionally, the Key Performance Indicators (KPIs) essential were defined, leading to the collection of patient EHRs, annotating nutritional labels based on each patient's needs, and developing a statistical and AI system to support healthcare personnel in easily identifying patient profile groups, key features, and obtaining nutritional recommendations and traits for them.

In short, the food products are being developed, and once produced, they will be supplied to patients in hospitals. Their impact will be analyzed, and the products will be refined until the most suitable formulations for each patient profile are achieved.

Acknowledgments

This work is funded by the Comunitat Valenciana Programme European Regional Development Fund (ERDF) 2021-2027 and the Agència Valenciana de la Innovació (AVI) through the project EATI-TALL(INNEST/2023/10), and partially by: HEART-NLP (PID2024-156263OB-C22); and COOLANG (PID2021-122263OB-C22) funded by MCIN/AEI/10.13039/501100011033 and, as appropriate, by ERDF A way of making Europe, by the European Union or by the European Union NextGenerationEU/PRTR; the Generalitat Valenciana (Conselleria d'Educacio, Investigacio, Cultura i Esport), FEDER and AVI granted funding for NL4DISMIS (CIPROM/2021/21).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] M. Ghassemi, T. Naumann, P. Schulam, A. L. Beam, I. Y. Chen, R. Ranganath, Practical guidance on artificial intelligence for health-care data, The Lancet Digital Health 1 (2019) e157–e159.

[2] A. N. Jagannatha, H. Yu, Structured prediction models for rnn based sequence labeling in clinical text, in: Proceedings of the conference on empirical methods in natural language processing., volume 2016, NIH Public Access, 2016, pp. 856–865.

[3] E. Soysal, J. Wang, M. Jiang, Y. Wu, S. Pakhomov, H. Liu, H. Xu, CLAMP–a toolkit for efficiently building customized clinical natural language processing pipelines, Journal of the American Medical Informatics Association 25 (2018) 331–336.

[4] T. Wang, Y. Zhang, Y. Zhang, H. Lu, B. Yu, S. Peng, Y. Ma, D. Li, A hybrid model based on deep convolutional network for medical named entity recognition, Journal of Electrical and Computer Engineering 2023 (2023) 1–11. doi:10.1155/2023/8969144.

[5] I. Segura-Bedmar, P. Raez, Cohort selection for clinical trials using deep learning models, Journal of the American Medical Informatics Association 26 (2019) 1181–1188.

[6] ibs.GRANADA, El Gobierno aprueba la convocatoria para lanzar la nueva infraestructura de Medicina de Precisión, IMPaCT, https://www.ibsgranada.es/en/el-gobierno-aprueba-la-convocatoria-para-lanzar-la-nueva-infraestructura-de-medicina-de-precision-impact/, 2020.

[7] A. Arefeen, S. Fessler, S. M. Mostafavi, C. S. Johnston, H. Ghasemzadeh, MEALMETER: Using multimodal sensing and machine learning for automatically estimating nutrition intake, arXiv preprint arXiv:2503.11683 (2025).

[8] E. Sandri, G. Cerdá Olmedo, M. Piredda, L. U. Werner, V. Dentamaro, Explanatory ai predicts the diet adopted based on nutritional and lifestyle habits in the spanish population, European Journal of Investigation in Health, Psychology and Education 15 (2025) 11.

[9] P. Whig, B. Y. Kasula, N. Yathiraju, A. Jain, S. Sharma, Tracking Nutrient Intake with AI: How AI Can Help Monitor and Optimize Nutrient Consumption, in: AI-Based Nutritional Intervention in Polycystic Ovary Syndrome (PCOS), Springer, 2025, pp. 167–183.

[10] H. Ding, J. Tian, W. Yu, D. I. Wilson, B. R. Young, X. Cui, X. Xin, Z. Wang, W. Li, The application of artificial intelligence and big data in the food industry, Foods 12 (2023) 4511.