

MATIIGEN: A Model-Driven Framework for Automated DNA Variant Interpretation in Precision Medicine

Ana Pérez-Gómez^{1,*†}, Mireia Costa^{1,*†}, Alberto García^{1,*†}, Giovanni Giachetti^{1,2} and Óscar Pastor^{1,*†}

¹Universitat Politècnica de València, Camino de Vera s/n, Valencia, Spain

²Universidad Andres Bello, Facultad de Ingeniería, Chile

Abstract

Software engineering plays a key role in developing high-quality healthcare systems, particularly in Precision Medicine, where advanced systems are required to tailor clinical interventions based on individual genetic, environmental, and lifestyle factors. Therefore, it is necessary to rethink software engineering methods to address the significant challenges in healthcare systems, especially in system design, complex data analysis, traceability, and AI explainability. This paper presents the MATIIGEN project, which aims to address these challenges. MATIIGEN introduces an intelligent system for automated DNA variant interpretation. The project combines Model-Driven Software Engineering (MDSE) principles with Explainable Artificial Intelligence (XAI) to support the transition of genomic tools from laboratory validation environments (TRL4) to operational clinical settings (TRL6). MDSE provides the structural and methodological basis for building robust, traceable, and interoperable software systems, and XAI provides mechanisms to make the system's outputs more transparent and understandable to clinicians. The applicability of MATIIGEN platform is demonstrated through a historically overlooked chronic disease: Lipedema. By focusing on a condition that is under-researched in clinical settings, we can demonstrate that the MATIIGEN platform can generate insights in areas of medicine where data is scarce and complex. Our work demonstrates that adopting MDSE principles provides a systematic methodology for transforming multifaceted clinical data into actionable medical insights and establishes a scalable architecture for high-quality, software-intensive systems in the healthcare domain.

Keywords

Model-Driven Engineering, Precision Medicine, Lipedema, DNA Variant Interpretation

1. Introduction

Precision Medicine represents a fundamental shift in clinical practice, moving away from population-level treatments toward interventions tailored to each patient's genetic, environmental, and lifestyle variability [1]. Central to this shift is the analysis of DNA—the fundamental blueprint of life—and the study of how genetic differences, or variants, impact human health. This process, known as variant interpretation, involves systematically identifying and classifying these variants to determine their pathogenicity and relevance to a patient's specific condition. However, despite its importance, variant interpretation remains a primary bottleneck in the clinical pipeline.

This bottleneck is driven by the sheer volume and fragmentation of genomic data. A single human genome typically harbors between 4.1 and 5 million variants, the vast majority of which hold no clinical significance. Compounding this complexity, genomic data production is projected to range from 2 to 40 exabytes in the coming years, yet this critical information remains dispersed across more than 1,700 heterogeneous data sources. This places an unprecedented strain on healthcare infrastructures; to be effective, modern variant interpretation systems must integrate these massive, fragmented datasets, support complex analytical pipelines, and remain intuitively operable for clinical professionals who are not technical experts [2, 3].

CAiSE'26: 38th International Conference on Advanced Information Systems Engineering, June 08–12, 2026, Verona, IT

*Corresponding author.

† These authors contributed equally.

✉ anpego@vrain.upv.es (A. Pérez-Gómez); micossan@vrain.upv.es (M. Costa); algarsi@vrain.upv.es (A. García); ggiachetti@dsic.upv.es (G. Giachetti); opastor@vrain.upv.es (O. Pastor)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Software Engineering provides the systematic principles and tools necessary to build high-quality information systems required to manage this complexity. At the PROS Research Center (Universitat Politècnica de València), we have translated these foundations into the domain of precision medicine through the development of Delphi: a platform specifically engineered for the end-to-end automation of DNA variant interpretation. To build Delphi, we combined Model-Driven Software Engineering (MDSE) principles with Explainable Artificial Intelligence (XAI). Through MDSE, we enabled Delphi to allow the integration of thousands of heterogeneous data sources and the interpretation of the clinical relevance of millions of variants in minutes, while ensuring strict adherence to international clinical guidelines. Through XAI, we provide Delphi with the necessary mechanism to make the variant interpretation outputs transparent, traceable and understandable by clinicians. To date, our laboratory validation has achieved an accuracy exceeding 80% across a diverse range of pathologies. Consequently, we have positioned the platform at Technology Readiness Level 4 (TRL4).

Building on these promising results, the PROS group at UPV is currently leading the **MATIIGEN project**, an initiative funded by the Carlos III Health Institute under the Health Technology Development call. The primary objective of MATIIGEN is to advance Delphi from TRL4 to TRL6, transitioning the system from a controlled laboratory setting to a functional demonstration in a relevant clinical environment. This progression is essential to validate our MDSE-based architecture against the complexities of real-world clinical workflows, where data quality is inconsistent, and system reliability is paramount. By reaching TRL 6, we aim to demonstrate that our software engineering foundations can meet the rigorous demands of real clinical ecosystems—ensuring that Delphi is not only technically sound but also ready for seamless integration into the daily practice of precision medicine.

We have structured the transition of Delphi from TRL4 to TRL6 under this project along five strategic dimensions designed to bridge the gap between laboratory research and clinical practice. First, we will adapt our platform to work in a relevant clinical environment: Lipedema disease. Lipedema is a chronic, multifactorial adipose tissue disorder that predominantly affects women and severely impacts their quality of life [4, 5]. Despite strong evidence of a hereditary component, the precise genetic basis of Lipedema remains unknown, hindering early and accurate diagnosis [6, 7]. As a historically neglected and under-researched pathology that has long remained invisible to the medical establishment, Lipedema represents an ideal, high-complexity domain to demonstrate the transformative impact that a model-driven platform like Delphi can provide [8]

Second, platform usability will be prioritized by integrating specialized UX methodologies with a cloud-native infrastructure, ensuring the system is co-designed with clinical end-users to support intuitive operation. Third, we will establish a robust strategic framework for intellectual property and licensing that carefully balances our academic open-science values with the commercial sustainability required for long-term viability. Fourth, we will develop a comprehensive regulatory and technology transfer roadmap to facilitate the platform’s seamless adoption within the complex healthcare ecosystem. Finally, we will validate these improvements within the context of Lipedema in collaboration with hospitals and clinical experts of this disease.

By the end of the MATIIGEN project, we expect to deliver a TRL6 system for supporting variant interpretation in a clinically relevant environment. The project will serve as a proof of concept for the use of Model-Driven Software Engineering and Explainable Artificial Intelligence in genomic interpretation workflows. More broadly, the work will contribute an architectural and methodological basis for bringing research advances in genomic analysis closer to real precision medicine settings.

2. MATIIGEN Objectives

The main objective of MATIIGEN is to mature and validate the technology of a software prototype currently at TRL4, developed using MDD and Explainable Artificial Intelligence (XAI) techniques to automate the interpretation of genetic variants. Through MATIIGEN, the technology will be advanced to TRL6, demonstrating its applicability in real clinical environments, using Lipedema as a representative use case.

To achieve this goal, the project will pursue the following specific objectives:

Objective 1 – Adapt the conceptual and technical foundations of the platform to the Lipedema context. Currently, Delphi is based on an existing conceptual model of the genome [9, 10] that guides how knowledge is represented and managed in the platform. This model will be refined and instantiated to support clinical professionals involved in Lipedema diagnosis. Model-Driven Engineering (MDE) techniques will be applied to systematically capture expert knowledge, define and automate variant-interpretation pipelines, and ensure full traceability to clinical evidence. In parallel, the platform will incorporate Explainable AI (XAI) mechanisms to support transparent, traceable, and interpretable variant interpretation [11]. Additionally, a secure, scalable, and GDPR-compliant cloud architecture will be implemented, including role-based access control, data encryption, and performance optimization mechanisms.

This objective will result in:

- A Lipedema-specific conceptualization of genomic knowledge.
- Explainable AI components enabling transparent and traceable variant classification.
- A secure and scalable cloud infrastructure aligned with current regulatory requirements.

Objective 2 – Improve the platform usability for clinical professionals. The project will define design requirements for the interaction between clinical professionals and the platform, with particular attention to usability, interpretability, and integration into real diagnostic practice. Based on these requirements, the web-based interfaces will be redesigned to better support the different stages of genomic interpretation in the context of Lipedema. This will include the development of mock-ups and the integration of interactive visualization modules to improve the exploration and interpretation of genomic data.

This objective will result in:

- A set of user and interaction requirements for clinically oriented genomic interpretation.
- Mock-ups and interface prototypes supporting the main tasks of Lipedema diagnosis.
- Interactive visualization components for exploring and interpreting genomic evidence within the platform.

Objective 3 – Validate the MATIIGEN platform in real clinical settings using Lipedema patient cohorts. The platform will be validated in collaboration with reference hospitals (i.e., La Fe University and Polytechnic Hospital and the INCLIVA Institute for Health Research) and clinics by applying it to real Lipedema cases. This validation will assess the system's capability to prioritize genetic variants potentially associated with Lipedema and evaluate diagnostic precision and efficiency using quantitative metrics, such as concordance with expert interpretation and reductions in analysis time, demonstrating effective integration into real clinical workflows. In addition, the validation will assess the interpretability and clinical usefulness of XAI-generated explanations, analyzing their contribution to decision-making processes and user trust. This objective will result in:

- A clinically validated dataset of Lipedema cases.
- A structured clinical validation report based on retrospective and prospective Lipedema cohorts, including quantitative evidence of diagnostic performance (e.g., precision, concordance with expert assessments), efficiency gains (e.g., reduction in analysis time).
- The evaluation of XAI-based explanations in terms of interpretability, usefulness, and trustworthiness from clinical experts.
- Validated clinical protocols and integration guidelines supporting the reproducible deployment of the platform within real hospital workflows.

Objective 4 – Define and implement an intellectual property protection strategy for software exploitation. This strategy will support the protection and exploitation of the MATIIGEN

platform as a software-intensive clinical decision-support system. An appropriate intellectual property protection strategy will be defined and implemented, including patentability assessment, software copyright registration, and the establishment of differentiated licensing schemes for academic and commercial exploitation. This objective will result in:

- Registered software assets.
- Formal patentability assessment.
- A proposal of a licensing model to enable sustainable exploitation in both academic and commercial contexts.

Objective 5 – Develop a technology transfer and commercialization strategy for genomic decision-support solutions in women’s health, with an initial focus on Lipedema, including market analysis, identification of early adopters, pilot deployments in real environments, and the definition of a preliminary regulatory roadmap. This objective will result in:

- A detailed market analysis focused on genomic and data-driven solutions for women’s health.
- A roadmap for regulatory compliance and certification.
- A report identifying early adopters and pilot deployment scenarios.
- A strategic plan for technology transfer and commercialization.

3. MATIIGEN Structure and Timeline

MATIIGEN is organized into five work packages (WPs) aligned with the project’s core objectives. WP1 encompasses the conceptual model for genomic applications in Lipedema (WP1.1) and the enhancement of platform usability (WP1.2). WP2 is dedicated to clinical validation, while WP3 and WP4 address intellectual property protection and the definition of the commercialization strategy, respectively. Finally, WP5 focuses on dissemination and serves as a cross-cutting activity throughout the project’s duration.

Launched in January 2026 and scheduled to run until December 2028, the project is currently in its initial execution phase. Efforts are presently concentrated on the early stages of WP1.1 (conceptual model adaptation) and WP1.2 (usability improvement), alongside preparatory work for WP2 (clinical validation). These foundational activities are pivotal for driving the planned technological advancement from TRL4 to TRL6.

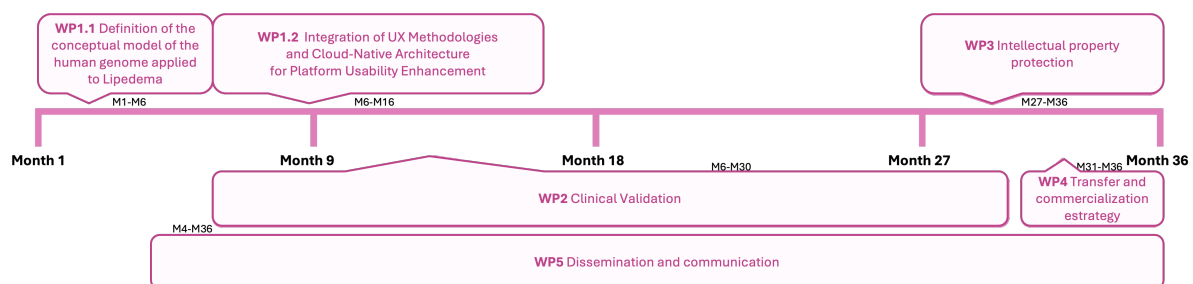


Figure 1: MATIIGEN Project timeline until Dec 2028

4. Relevance for the CAISE Community

MATIIGEN contributes to the CAiSE community by showing how fundamental principles of **conceptual modeling** and **Model-Driven Software Engineering (MDSE)** can be effectively applied in one of the most demanding and knowledge-intensive domains: precision medicine. Beyond a domain-specific

application, the project demonstrates how model-driven approaches can be scaled, adapted, and validated in real-world clinical settings.

This work makes a significant contribution to advancing MDE toward higher levels of technology readiness. While many model-driven solutions are still in the prototype stage, MATIIGEN focuses explicitly on the transition from TRL4 to TRL6, providing a concrete example of how conceptual models can guide the evolution of a system from research-grade artifacts to clinically validated tools. This progression provides insight into the practical challenges, trade-offs, and engineering decisions involved in implementing model-driven approaches in operational healthcare settings.

At the same time, the project emphasizes the importance of conceptual modeling as a basis for structuring genomic knowledge and facilitating traceability. Conceptual models in MATIIGEN support system design, integrate heterogeneous data sources, formalize interpretation pipelines, and explicitly link clinical evidence to decision outcomes. This viewpoint emphasizes the role of conceptual modeling in promoting transparency, reproducibility, and accountability in data-intensive healthcare systems, while enabling structured representations that can evolve toward ontology-driven and knowledge graph-based solutions.

Another relevant aspect is the incorporation of explainability as a primary concern within a model-driven context. By combining MDE and **Explainable AI** (XAI), the project moves beyond treating explanations as post-hoc artifacts and instead explores how they can be systematically embedded into system design. This opens up new research directions for the CAiSE community, particularly in terms of modeling, representing, and aligning explanations with clinicians' cognitive and decision-making processes, connecting with emerging CAiSE research on **AI-enabled decision support** and **human-centered intelligent systems**.

Furthermore, MATIIGEN demonstrates the applicability of model-driven approaches in domains with limited data and research. The use of Lipedema as a validation scenario demonstrates how structured conceptual representations can compensate for limited data availability, allowing the development of meaningful and usable systems even in situations where traditional data-driven approaches are severely limited, and illustrating the role of domain engineering in structuring emerging biomedical knowledge.

These contributions establish MATIIGEN as a step toward the creation of semantically grounded, model-driven precision medicine infrastructures. Such infrastructures have the potential to enable interoperability, continuous knowledge evolution, and reproducible clinical decision-making across a wide range of healthcare contexts, highlighting the importance of CAiSE research in addressing pressing challenges in modern healthcare systems.

5. Conclusions

MATIIGEN addresses one of precision medicine's most critical bottlenecks: the accurate, scalable, and clinically relevant interpretation of genetic variants. The project proposes a systematic approach that integrates Model-Driven Software Engineering (MDSE) with Explainable Artificial Intelligence (XAI) to overcome the limitations of current solutions, which frequently rely on opaque, rigid, and poorly maintainable pipelines that hinder traceability, adaptability, and clinical trust.

This work makes an important contribution by demonstrating how conceptual modeling can be used to structure genomic knowledge, orchestrate interpretation workflows, and ensure end-to-end traceability between raw data, analytical processes, and clinical decisions. Unlike existing tools, which often operate as "black boxes" with no systematic update mechanisms, MATIIGEN promotes a transparent, evolvable framework that explicitly represents domain knowledge, interpretation criteria, and clinical evidence and continuously integrates them.

Aside from its methodological contribution, MATIIGEN is positioned as a practical step toward closing the gap between research-grade genomic tools and their implementation in real-world clinical settings. The explicit emphasis on moving the system from TRL4 to TRL6, combined with planned validation using real patient cohorts and integration into hospital workflows, provides a practical example of how model-driven approaches can be scaled to high technology readiness levels in healthcare settings.

The use of Lipedema as a validation scenario strengthens this contribution. Lipedema represents an under-researched, data-scarce, and clinically underserved condition for which traditional data-driven approaches fail to provide value.

Nevertheless, a number of challenges remain unresolved. The clinical validation phase must show not only increases in diagnostic accuracy and efficiency, but also consistent adoption by clinicians in real-world workflows. Furthermore, the use of XAI raises fundamental questions about how explanations should be formalized, evaluated, and aligned with clinical reasoning practices. The system's real-world impact will be influenced by issues such as interoperability, regulatory compliance, and long-term genomic data governance.

Future work will focus on consolidating clinical validation, refining the underlying conceptual models to incorporate emerging biomedical knowledge, and expanding the platform to encompass broader precision medicine scenarios. In this regard, MATIIGEN is envisioned as the foundation for a semantically grounded, model-driven genomic interpretation infrastructure that encourages reproducibility, transparency, and scalability across a wide range of clinical contexts.

Acknowledgments

This work is part of the MATIIGEN project (DTS25/00159) funded by the Instituto de Salud Carlos III and co-funded by the European Union. Also, it is partially supported by the VARNETICA project (CNS2023-145422), the UNICA project (101218966), and the Universitat Politècnica de València through grant PAID-06-24.

Declaration on Generative AI

Generative AI tools (Gemini, Grammarly, and Quillbot) were used for grammar and spelling check and minor language refinement. All content was originally produced by the authors, who reviewed and edited the manuscript and take full responsibility for its content.

References

- [1] F. S. Collins, H. Varmus, A new initiative on precision medicine, *New England journal of medicine* 372 (2015) 793–795.
- [2] M. Martínez-García, E. Hernández-Lemus, Data integration challenges for machine learning in precision medicine, *Frontiers in medicine* 8 (2022) 784455.
- [3] R. Casanova, F. A. Villa-Garzon, J. W. Branch-Bedoya, Architectural patterns for health information systems: a systematic review, *Frontiers in Digital Health* 7 (2025) 1694839.
- [4] V. Cifarelli, Lipedema: progress, challenges, and the road ahead, *Obesity Reviews* 26 (2025) e13953.
- [5] U. B. Günay, G. R. Talak, A. Demiröz, Health-related quality of life among lipedema patients: A systematic review and meta-analysis, *Phlebology* (2025) 02683555251410009.
- [6] P. Kruppa, R. Crescenzi, G. Faerber, I. Forner-Cordero, M. Cornely, R. Shayan, T. Karnezis, J. L. Simarro, P. F. de Souza, K. L. Herbst, et al., Lipedema world alliance delphi consensus-based position paper on the definition and management of lipedema: Results from the 2023 lipedema world congress in potsdam, *Nature Communications* (2026).
- [7] S. Morgan, I. Reid, C. Bendon, M. Ishaq, R. Shayan, B. Pope, D. Park, T. Karnezis, A family-based study of inherited genetic risk in lipedema, *Lymphatic research and biology* 22 (2024) 106–111.
- [8] M. Solera-Martínez, et al., Diagnosis and treatment of lipedema in spain. results of an online survey, *Rehabilitacion* 56 (2021) 108–115.
- [9] A. Bernasconi, A. García S, S. Ceri, O. Pastor, A comprehensive approach for the conceptual modeling of genomic data, in: *International Conference on Conceptual Modeling*, Springer, 2022, pp. 194–208.

- [10] A. García S., et al., Towards the understanding of the human genome: A holistic conceptual modeling approach, *IEEE Access* 8 (2020) 197111–197123. doi:10.1109/ACCESS.2020.3034793.
- [11] R. Rosenbacke, Å. Melhus, M. McKee, D. Stuckler, How explainable artificial intelligence can increase or decrease clinicians' trust in ai applications in health care: systematic review, *Jmir Ai* 3 (2024) e53207.