



HeKDisco

Knowledge discovery for improved healthcare outcomes

To improve patient outcomes, the ITEA project HeKDisco (Healthcare Knowledge Discovery) developed a multimodal knowledge discovery platform, knowledge extraction modules and predictive artificial intelligence (AI) models that enhance real-time decision-making by physicians.

Medical errors and adverse healthcare-related events occur in ~8-12% of hospitalisations and 23% of EU citizens have been affected by medical error. While evidence shows that 50-70.2% of such harm can be prevented through comprehensive systematic approaches, a key challenge is standardisation: hospitals use varying coding systems and countries have their own standards. Additionally, most healthcare information systems were initially designed to support administrative and billing processes, limiting their clinical effectiveness unless optimised for patient care.

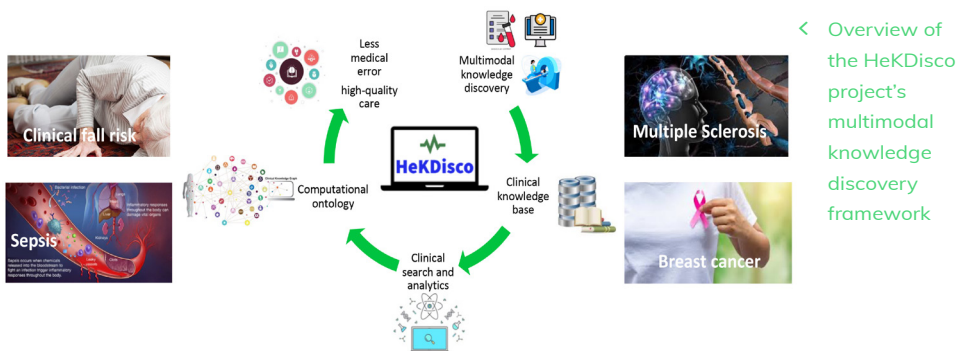
HeKDisco addressed these challenges by developing a multimodal knowledge discovery process to provide physicians with empirical evidence prior to decision-making. The project's approach utilised the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), an open community data standard designed to standardise the structure and content of observational data. While OMOP CDM has been widely adopted for retrospective research, its application in clinical settings remains limited due to real-time mapping challenges between hospital systems and OMOP-formatted data. HeKDisco focused on enabling real-time OMOP mapping in four use-cases – sepsis, multiple sclerosis (MS), breast cancer and clinical fall risk – thereby supporting knowledge extraction, data processing, and the development of a platform for monitoring, analytics and searching.

Technology applied

HeKDisco's focused on extracting and

transforming raw medical data – such as magnetic resonance imaging (MRI), pathology images, and reports – into structured clinical information using natural language processing (NLP) and image processing techniques, then standardising and mapping this to the OMOP CDM database. This enabled a prototype knowledge discovery platform that integrates knowledge from various sources into a unified system for analysis,

patient-reported cognitive assessments and enhanced the interpretability of brain MRI diagnostics by comparing MS patient data with age- and sex-matched healthy controls. For breast cancer, AI was applied to analyse pathology reports and generate structured outputs from histopathology images, including tumour morphology and grade. In the sepsis use-case, machine learning models were developed to predict sepsis onset using electronic medical records (EMRs) and the first 24 hours of intensive care unit data. For fall risk, the project integrated wearable sensors, AI-based pattern recognition, multimodal data fusion, and context-aware real-time monitoring to support timely interventions.



< Overview of the HeKDisco project's multimodal knowledge discovery framework

cohort identification, and clinical search. This can be exploited by healthcare providers as an entire system or as individual knowledge extraction modules that function independently and integrate with other systems. The project also developed AI prediction models per use-case, which support analytics to identify high- and low-risk patient characteristics. Comparing these populations can provide deeper insights into disease progression and risk factors. These models can also be marketed separately or as part of the system.

Diverse technical innovations have been applied across the use-cases. The MS use-case developed digital tools for capturing

Making the difference

HeKDisco adopted a novel approach to knowledge extraction by leveraging AI and big data analytics, integrating heterogeneous data sources such as structured EMRs, unstructured clinical notes, imaging, and sensor data. This addresses multiple gaps in the clinical informatics literature. For instance, clinical cohort-level queries – used to identify and analyse patient populations with specific characteristics – are often time-intensive as most systems rely on offline batch processing. HeKDisco partner Mantis enabled real-time online cohort querying and, in a pilot study with ~3,000 patients, reduced query time from over several seconds to under one second. This

transforms cohort analysis into a rapid, interactive process through a search platform, potentially reducing clinician wait times and streamlining data-driven decision support, especially at scale. The error rate in retrieving irrelevant or incorrect information was also reduced from ~10% to <1% in the pilot study, offering higher-quality care.

HeKDisco's use-case developments have opened opportunities for both commercial exploitation and internal innovation, with several initiatives underway. For example, Mantis and Bilbest have integrated the sepsis prediction model into the HeKDisco platform to better monitor and analyse high-risk patients. Virasoft automated pathology workflows using AI for report and image analysis, reducing pathologist workload and variability while improving diagnostic accuracy and patient outcomes. icometrix developed AI models for predicting MS progression based on its icobrain MRI analysis suite and incorporated the patient-reported cognitive assessment tools into its icompanion patient app,

thereby improving their service quality and competitiveness. Caretronic achieved 96% fall detection accuracy and 90% fall prediction accuracy in its pilot study, providing the foundation to launch a web dashboard and mobile app to make recommendations to family and caregivers.

The future

Having piloted the full system at Istanbul University, HeKDisco aims to introduce its innovations to hospitals globally, particularly to the several hundred hospital databases that are already mapped to OMOP CDM. This offers a springboard to extend the use-cases, such as in cardiology and other cancers. In the EU, comprehensive systematic approaches are estimated to be able to prevent 750,000 harm-inflicting medical errors per year, leading to over 3.2 million fewer days of hospitalisation, 260,000 fewer incidents of permanent disability and 95,000 fewer deaths per year. With its capacity to improve both diagnostics and prognostics, HeKDisco hopes to make a sizeable contribution to this.

Major project outcomes

Dissemination

- > 8+ publications and 15+ posters/oral presentations at international fairs/conferences.

Exploitation (so far)

- > **Clinical search engine** enabling users to conduct real-time cohort/population-level analyses on standardised clinical data mapped to OMOP CDM.
- > **Sepsis learning engine** allowing for real-time prediction, enhancing the platform's ability to generate insights to guide clinical decisions.
- > **MS prediction engine** integrable with icometrix's MS software, for predicting future clinical progression, disability accumulation, and conversion to more severe MS stages.
- > **Advanced bracelet for clinical fall assessment** that monitors gait stability and predicts high fall risk using a machine learning model.
- > **Pathology report and histopathology image analysis modules** extracting structured data from unstructured pathology reports and images using advanced NLP and image processing techniques, automating pathology workflows.
- > **icobrain MS population modeling** providing population-based graphs of brain lesions and volumes, compared to age- and sex-matched healthy controls and MS cohorts
- > **Brain MRI feature extraction module** that extends icometrix's existing software with new predictive MRI biomarkers, including MRI lesions and volumetry.
- > **Whole Slide Imaging (WSI) annotation dataset** for more than 10,000 images to be used in the development of deep learning models.
- > **HeKDisco knowledge discovery platform** integrating all knowledge extraction and predictive modules developed in the project with the clinical search engine.

Standardisation

- > **Contributing to OMOP CDM:** Incorporating new radiological and histopathological biomarkers as covariates for predicting the clinical course of multiple sclerosis and breast cancer within the OMOP CDM standardised vocabulary.

ITEA is the Eureka RD&I Cluster on software innovation, enabling a large international community of large industry, SMEs, start-ups, academia and customer organisations, to collaborate in funded projects that turn innovative ideas into new businesses, jobs, economic growth and benefits for society. ITEA is part of the Eureka Clusters Programme (ECP).

<https://itea4.org>

HeKDisco

20030

Partners

Belgium

- > icometrix

Slovenia

- > Caretronic d.o.o.

Türkiye

- > Bilbest Bilişim Sağlık Eğitim Dış Ticaret ve Sanayi Ltd.Şti.
- > Istanbul University Faculty of Medicine
- > Mantis
- > Virasoft Yazılım Ticaret A.Ş.

Project start - Project end

October 2021 - January 2025

Project leaders

- > Nurol Gençyılmaz, Bilbest Bilişim Sağlık Eğitim Dış Ticaret ve Sanayi Ltd.Şti.
- > Behzad Naderalvojud, Mantis

Project emails

- > ngencyilmaz@bilbest.com.tr
- > behzadn@mantis.com.tr

Project website

<https://hekdisco.com/>

<https://itea4.org/project/hekdisco.html>

