Health-related NLP activities at IDSIA

Fabio Rinaldi[1,*], Andrea Franchini[1], Joseph Cornelius[1], Oscar Lithgow-Serrano[1], Sandra Mitrovic[1], Lorenzo Ruinelli[2], Roberta Noseda[2], Amos Colombo[2] and Alessandro Ceschi[2]

[1]*Dalle Molle Institute for Artificial Intelligence (IDSIA USI-SUPSI), CH-6962 Lugano, Switzerland*
[2]*Ente Ospedaliero Cantonale (EOC), CH-6500 Bellinzona, Switzerland*

Abstract

We briefly describe three projects at the intersection of AI and Health, conducted in partnership between the Dalle Molle Institute for Artificial Intelligence (IDSIA) and the Ticino Cantonal Hospital (EOC). The projects focus on improving the workflow of clinical documentation, from the generation of discharge summaries to the semi-automated detection of adverse drug reactions, and the identification and removal of personal health information for the purpose of secondary usage of clinical records.

Keywords

Generative AI, Clinical Documentation, LLMs

1. Introduction

The introduction of Artificial Intelligence in the field of medicine holds the promise of revolutionary changes in the treatment of diseases and better management of patients. The introduction of AI for diagnostic and treatment, although offering a potentially high return, faces resistance due to concerns related to errors and hallucinations, which could have serious consequences.

One area where the introduction of AI technologies appears to be less controversial is in documentation management. During a patient's hospital stay extensive clinical reports are produced, which are later consolidated into a discharge summary, which has the primary purpose to communicate essential patient information to healthcare providers responsible for the ongoing treatment of the patient.

In partnership with the Ticino Cantonal Hospital (EOC) we are conducting several projects aiming at relieving the burden of medical documentation and at enabling its usage for downstream medical and scientific purposes.

The production of clinical documentation is often seen a slow and cumbersome process which distracts the doctors from their primary duty of patient care. Our aim is to relieve clinicians and other medical practitioners of this serious burden, releasing precious time which can be dedicated to the patients.

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
*Corresponding author.
fabio.rinaldi@supsi.ch (F. Rinaldi); andrea.franchini@supsi.ch (A. Franchini); joseph.cornelius@supsi.ch (J. Cornelius); oscarwilliam.lithgow@supsi.ch (O. Lithgow-Serrano); sandra.mitrovic@supsi.ch (S. Mitrovic); Lorenzo.Ruinelli@eoc.ch (L. Ruinelli); Roberta.Noseda@eoc.ch (R. Noseda); Amos.Colombo@eoc.ch (A. Colombo); Alessandro.Ceschi@eoc.ch (A. Ceschi)
0000-0001-5718-5462 (F. Rinaldi); 0009-0009-6963-9412 (A. Franchini); 0000-0002-5427-5005 (J. Cornelius); 0000-0003-1995-1669 (O. Lithgow-Serrano); 0000-0002-5697-5865 (S. Mitrovic); 0000-0002-9144-2570 (L. Ruinelli); 0000-0002-4946-7926 (R. Noseda); 0009-0000-3638-8973 (A. Colombo); 0000-0002-4308-0405 (A. Ceschi)

2. Project AutoDischarge

Automatic discharge letter generation is a topic of high interest and is being investigated by various actors worldwide. Several publications and initiatives show its importance and relevance, such as the BioNLP ACL'24 Shared Task on Streamlining Discharge Documentation [1].

Discharge letters are a fundamental component of all medical institutions. They provide essential information about the patient, such as treatments, past medical history, and medical conditions upon admission and discharge. Insurance companies, general practitioners, hospital staff, post-treatment care, future re-admission, etc., employ discharge letters.

Despite their crucial role, redacting discharge letters is still a highly resource-intensive process. In the US, studies report that for every hour the medic spends with the patient, on average, two are invested in writing discharge letters, billing, coding, etc. [2]; therefore, discharge letters and similar administrative requirements are considered a significant culprit in medical professional burnout [3, 4].

Additionally, discharge letters often lack quality and timeliness, significantly contributing to adverse post-hospital care events and incident reports [5]. In 1999, bureaucratic costs comprised at least 31% of the total health spending in the United States, a similar estimate also for the Swiss health sector. Given these latent and unexploited savings in overall costs, medic's time, and safety, a solution to automatically draft a reasonable discharge summary to be reviewed and finalized by domain experts would positively impact the healthcare system.

Our primary objective in the InnoSuisse-funded AutoDischarge project is to develop an AI-based system capable of generating quality discharge letters from medical notes in collaboration with the Ente Ospedaliero Cantonale (EOC) and Ex Machina, targeting initially the Swiss-Italian region, and later expanding to the rest of Switzerland.

Our preliminary work [6] highlights that Large Language Models (LLMs) offer considerable potential for improving the summarization of clinical records in healthcare, particularly for the preparation of discharge letters. However, their adoption faces challenges. First and foremost, our experiment setup relies on a publicly available ChatGPT model, that raises significant privacy concerns, as the anonymization tool may fail to completely remove sensitive parts of the data. In our case, since the number of cases was limited, the risk was mitigated by manually inspecting the prompts. Secondly, it remains challenging to ensure that the generated summaries consistently maintain a high level of accuracy. On one hand, quantitative metrics excel at capturing the broad meaning of a text, yet they frequently struggle to pinpoint specific details crucial in clinical practice, such as nuanced terms like hallucination. On the other hand, qualitative evaluations still rely on expert evaluation, which is often subjective and also extremely expensive, as it requires the judgment of senior medical doctors.

In this project we intend to address both concerns by researching into a locally available solution and to verify the model using a novel external evaluation system that provides independent feedback on the quality of AI-generated outputs.

As a secondary goal we are considering the generation of patient-friendly discharge summaries. One way to increase patients' involvement in their own medical treatment is by making their medical documentation more accessible. In the US, the 21st Century Cures Act enforces Hospitals to release discharge letters to the patient upon request; and a similar regulation exists in Switzerland under the Right of Access to Case File. However, patients are likely to lack the medical literacy necessary to understand the content of a discharge letter; thus, they must either rely on the physician for explanations. Patient misunderstanding has been linked to increased costs in the health system. Several studies support the view that more conscious and aware patients lead to health-related improvements [7, 8]. Recent studies testing the automated generation of simplified discharge letters via LLMs show that the technology is mature enough for practical usage [9, 10, 11].

3. Project M2P2

Analysis of large quantities of clinical notes offers a great potential for medical knowledge discovery. However several factors limit their availability for scientific usage, primarily the private and sensitive nature of the information they contain. Coupled with that, the actual amount of annotated text available at any hospital in a specific medical area might be insufficient to enable the identification of relevant clinical signals. Reliable de-identification methods would enable sharing and pooling more significant quantities of medical records. Additionally, the different languages in Switzerland necessitate the development of multilingual text analysis capabilities.
In this project, we are developing multilingual de-identification techniques that will help to overcome these problems. We are currently exploring strategies based on the generation of synthetic text in the three target languages starting from publicly available datasets. In particular we are considering a subset of MIMIC [12] where PHIs have been manually removed. We plan to automatically replace the missing entities with surrogates and translate the documents in our three target languages, yielding a multilingual synthetic dataset, suitable for fine-tuning one of the recent open-weights models on the specific task of multilingual de-identification.
Additionally the project includes two areas of application: (1) extraction of clinically-relevant information from notes in oncology, and (2) summarization of discharge communication. The first case study focuses on the detection of adverse drug reactions in oncology. The second case study is about methods for generating and summarizing discharge letters, in order to simplify the administrative burden for doctors, and to adapt the language to the patient's needs. Thanks to the participation of three major hospitals (EOC, USZ, CHUV) we will implement solutions valid across Switzerland, using three national languages (Italian, French, and German).

4. Project QUADRATIC

The QUADRATIC project is a collaboration between IDSIA's NLP group and the Institute of Pharmacological Sciences of Southern Switzerland (ISFSI), funded by the Swiss National Science Foundation. It aims to investigate the use of NLP methods for automated detection of adverse drug reactions (ADRs) within clinical discharge letters in the context of pharmacovigilance. The project tackles two aspects of the pharmacovigilance process. First, it tries to detect automatically if a given report contains a mention of a reportable ADR. These reports are then presented to the human expert for further assessment. Second, for the reports which are judged positive, it tries to find the items of information that need to be reported (such as type of event, drug, dose, date, etc). The first aspect is treated as a Classification problem, the second as a Named Entity Recognition problem (NER).
We framed ADR detection as a binary classification task on 400 anonymized Italian discharge summaries (200 positive, 200 negative). Classical machine learning algorithms such as Logistic Regression and Random Forest were paired with three vectorization techniques, with hyperparameter tuning via grid search and stratified k-fold cross-validation. Logistic Regression with bag-of-words scored as the top model, based on our set of metrics. We then explored deep learning models by fine-tuning a BERT-based model pre-trained on Italian medical texts, and developing a hierarchical BERT architecture to mitigate the context length limitations. Our naive deep-learning implementation showed strong promise, suggesting that further optimization could surpass classical models. Finally, we deployed a daily pipeline in production that extracts new discharge letters, ranks the top 20 ADR candidates for expert review, with the objective of improving the pharmacovigilance efficiency.
The NER task is structured as a token classification problem, where each token (or sub-token, in the case of pre-trained language models (PLMs)) is assigned a label corresponding to one of three categories. We divide the terms into three labels: medical events, drugs, and other. We combined a rule-based approach that leverages specialized dictionaries with a PLM-based approach. Specifically, we use MedPsyNIT [13],

a domain-specific PLM built upon BERT. This model is fine-tuned on an Italian biomedical corpus to adapt it to the specific language and domain requirements. Although the results are significantly better than our dictionary-based baseline NER approach, the current NER solution is not yet fully deployed within the ISFSI pharmacovigilance baseline. The results of the project have recently been preliminarily presented in [14].

Acknowledgments

The activities described in this paper have been supported by grants 220564 and 10003518 (Swiss National Science Foundation), and grant 122.281 IP-ICT (Swiss Innovation Agency), to Dr. Fabio Rinaldi.

Declaration on Generative AI

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

References

[1] Stanford AIMI, Discharge Me!, GitHub Pages, n.d. URL: https://stanford-aimi.github.io/discharge-me/, retrieved September 3, 2024.
[2] C. S. Sinsky, Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties, Annals of Internal Medicine (2016).
[3] D. A. J. Salvagioni, Physical, psychological and occupational consequences of job burnout: A systematic review of prospective studies, PLoS One (2017).
[4] M. Panagioti, Association between physician burnout and patient safety, professionalism, and patient satisfaction: a systematic review and meta-analysis, JAMA Internal Medicine (2018).
[5] K. Weetman, What makes a "successful" or "unsuccessful" discharge letter? Hospital clinician and general practitioner assessments of the quality of discharge letters, BMC Health Services Research (2021).
[6] L. Ruinelli, A. Colombo, M. Rochat, S. G. Popeskou, A. Franchini, S. Mitrović, O. W. Lithgow, J. Cornelius, F. Rinaldi, Experiments in automated generation of Discharge Summaries in Italian, in: Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024, ELRA and ICCL, Torino, Italia, 2024, pp. 137–144.
[7] J. Greene, Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes, Journal of General Internal Medicine (2012).
[8] J. Hibbard, Do increases in patient activation result in improved self-management behaviors?, Health Services Research (2007).
[9] J. Zaretsky, Generative artificial intelligence to transform inpatient discharge summaries to patient-friendly language and format, JAMA Network Open (2023).
[10] M. Sudarshan, Agentic LLM Workflows for Generating Patient-Friendly Medical Reports, ArXiv (2024).
[11] H. Kim, Patient-Friendly Discharge Summaries in Korea Based on ChatGPT: Software Development and Validation, Journal of Korean Medical Science (2024).
[12] A. Johnson, L. Bulgarelli, T. Pollard, B. Gow, B. Moody, S. Horng, L. A. Celi, R. Mark, MIMIC-IV (version 3.1), PhysioNet, 2024. RRID:SCR_007345.

[13] C. Crema, et al., Advancing Italian biomedical information extraction with transformers-based models: Methodological insights and multicenter practical application, Journal of Biomedical Informatics 148 (2023) 104557.
[14] A. Ceschi, A. Colombo, J. Cornelius, A. Franchini, O. Lithgow Serrano, S. Mitrovic, F. Rinaldi, L. Ruinelli, NLP in support of Pharmacovigilance: QUality Adverse Drug Reaction AcTIve Control (QUADRATIC), in: Precision Health Day 2025, SUPSI, Lugano, 2025.