Healthcare information system engineering: AI technologies and open source approaches

Shen, Z.

Chapter 1

Introduction

This dissertation focuses on the development of Healthcare Information Systems (HIS). The research designs and develops a number of HIS with various Machine Learning (ML) and Natural Language Processing (NLP) techniques that address a number of issues in healthcare. Further investigations are conducted on how to improve HIS engineering with Open Source methodology. All chapters together present an overview of HIS engineering in today's healthcare.

To provide a background of this dissertation, the introduction first presents the concepts of HIS engineering and various types of HIS related to this research in section 1.1. Then, Open Source Software and its application in HIS engineering are discussed, respectively in section 1.2 and section 1.3. Next, the main research question and related research questions are introduced, along with the main research methods used to investigate them in section 1.4 and section 1.5. Finally, the remaining chapters of this dissertation are outlined in section 1.6, and the outcomes listed in section 1.7.

1.1 Healthcare Information System Engineering

Healthcare engineering is defined as engineering involving all aspects of healthcare, from prevention to treatment. Its purpose is to improve human health and well-being through engineering approaches (Chyu et al., 2015). Chyu et al. (2015) lists Healthcare Information Systems (HIS) engineering as an important subject of healthcare engineering. A healthcare information system is defined as a system designed to manage healthcare data, including data collection, processing, reporting, and analytics of such data (Wager et al., 2021). These systems intend to provide better and timely decision-making with respect to patient treatments and the provision of healthcare services (Fichman et al., 2011; Ljubicic et al., 2020). The scope of HIS engineering includes the following domains of interest: Electronic Health Record, E-Health, Data Mining and Big Data, M-Health, Telemedicine, Wireless Technology, and Information Security (Chyu et al., 2015). This dissertation addresses the first three domains in depth.

1.1.1 Electronic Health Records

As a prominent example of HIS, electronic health records (EHR) have moved beyond the simple digitization of healthcare data that are collected during routine delivery of health care. More and more resources are dedicated to utilize data in EHR to support clinical decision-making (Evans, 2016). For instance, the large amount of patient data stored in the EHR consists of a large variety of free-text documents, such as discharge summaries and medical notes. This spawned the need for the use of NLP to read, process and transform free-text data with standardized codes such as ATC and ICD10.

1.1.2 E-Health

An E-Health service refers to a service that uses information and communication technology to improve healthcare (Pagliari et al., 2005). EHR is often included as one of the common applications of E-Health. However, we consider it as a separate field because of its size and importance in the HIS domain. Blaya et al. (2010) presents a list of E-Health categories, including laboratory information management, pharmacy information systems, patient registration or scheduling, monitoring, evaluation, and patient tracking systems, clinical decision support systems, patient reminder systems and research data collection systems. In particular, clinical decision support systems (CDSSs) have received a lot of attention in recent years (Sutton et al., 2020). A CDSS supports clinicians in their complex decision-making processes with a combination of clinical knowledge, patient information, and other health information (Osheroff et al., 2012). CDSSs are often provided to clinicians as web or mobile applications.

1.1.3 Data Mining and Big Data

Big data and data mining are among the most important topics in today's IT world which are discussed and applied in almost all industries and research domains (Che et al., 2013). Its importance in healthcare has grown significantly as a vast amount of the healthcare data is collected and stored (Yang et al., 2021). Data mining enables the discovery of hidden knowledge from big healthcare data so that healthcare professionals can use such knowledge to solve real-world problems. Applications that prove the benefits of data mining in healthcare can be found in a variety of healthcare research (Li et al., 2013; Menger et al., 2018a; Kavakiotis et al., 2017; Carchiolo et al., 2019; Albahri et al., 2020).

Besides the common structured data, healthcare data has a wide range of other data types, including semi-structured and unstructured data, such as images, texts, web pages and videos. Both researchers and practitioners in healthcare have developed or applied specific data mining techniques to cope with such data variability. Among them, NLP is an increasingly crucial component in facilitating healthcare services since textual data exists in almost all processes of healthcare, ranging from patient intake to discharge. Furthermore, a huge amount of healthcare literature has been accumulated over the decades, and the number is still growing on a daily basis. For example, in chapter 4 we investigate how to use NLP to encode patient health records, and in chapter 5 we extract knowledge from full-text literature.

1.1.4 Trends and Challenges

Ngafeeson (2015) discussed a number of important trends in the field of HIS. The scope of this dissertation focuses primarily on the following three trends:

From Local to Global HIS: With the development of Internet and information security, HIS has moved beyond a single hospital. There are many HIS that are built for national or international settings (e.g. chapter 3).

From using patient data to all healthcare data: Aside from patient data, a lot of other healthcare related data has been accumulated, such as healthcare literature. How to uncover knowledge from a huge amount of openly accessible full-text literature has gained more interest (e.g. chapter 6).

From Numeric Data to More Complex Types: As mentioned previously, there are a variety of data types of healthcare data, including text, images and videos. To fully unlock the potential of healthcare data, researchers are moving toward analyzing more complex types of healthcare data (e.g. chapter 5).

1.2 Open Source Software (OSS)

Open source software refers to software distributed with a license that guarantees free access to its source code, which allows users to freely redistribute the software, to create derived works, and to unrestrictedly use the software (Bretthauer, 2002). Its origin could be traced back to the 1950s. Decades of development have transformed OSS into a sophisticated engineering domain that has produced well-known open source projects such as Linux and Apache, among many others (Midha and Palvia, 2012). Therefore, OSS is an indispensable component of today's software industry. To facilitate the development of OSS, a number of IT tools have been created. Among them is GitHub, which hosts source codes of millions of OSS, and supports collaboration of software developers around the world (Shen and Spruit, 2019). Therefore, researchers have learnt more about OSS via studying open source projects hosted on GitHub (Kalliamvakou et al., 2015; Shen and Spruit, 2019).

Openly accessible OSS can benefit the IT development of both academia and industry in many ways. To begin with, the free accessibility of OSS allows us to develop our IT systems more efficiently by easily reusing existing tools or codes instead of creating everything from scratch. It, therefore, significantly reduces the IT development cost and time (McDonald et al., 2003). Secondly, the open source philosophy believes two heads are better than one, and a million heads can move mountains (Brabham, 2008). Therefore, OSS welcomes developers all around the world to contribute to the development and maintenance of software. This ensures its reliability and robustness due to the world-wide developer community. Lastly, the effective utilization of open-source software could significantly boost IT development in developing countries, especially in resource-poor areas where there are few IT resources (Shen et al., 2019a). Without OSS, developing countries might have difficulty building state of the art software, and their gap with developed nations would become wider and wider.

1.3 OSS in HIS Engineering

With the progress of healthcare IT in recent years, open source software has become an essential part of HIS engineering. More and more people in healthcare advocate OSS (de Abajo and Ballestero, 2012; Nolden et al., 2013; Shen and Spruit, 2019). Moreover, numerous open source software projects are created to solve healthcare problems (Athey et al., 2013; Hansen and Sørensen, 2013; Bankhead et al., 2017; Cavelaars et al., 2015; Gibson et al., 2018). Well-known and successful examples include openEHR (open Electronic Health Records), cTAKES (clinical Text Analysis and Knowledge Extraction System), and others. These projects have been used in multiple studies and some are even adopted in the real cinical settings. Table 1.1 shows more details about this selection of well-known open source healthcare software.

OSS	Description	License	Year	Contributors	Dev Status
openEHR	As one of the most popular open source electronic health records and medical practice management solutions, it supports a variety of clinical practices.	GNU GPL	2002	188	Active
cTAKES	cTAKES is a NLP system for extraction of information from electronic medical record clinical free-text.	Apache License 2.0	2006	2	Active
OpenClinica	OpenClinica is an open source software for Electronic Data Capture (EDC) and Clinical Data Management (CDM) used to optimize clinical trial workflow in a smart and secure fashion.	GNU LGPL	2005	22	Inactive
tranSMART	tranSMART is a knowledge management platform that enables scientists to develop and refine research hypotheses by investigating correlations between genetic and phenotypic data, and assessing their analytical results in the context of published literature and other work.	GNU GPL	2013	45	Inactive
Gadgetron	Gadgetron is an open source software for medical image reconstruction.	MIT license	2013	48	Active
NiftyNet	NiftyNet is an open-source convolutional neural networks platform for research in medical image analysis and image-guided therapy.	Apache License 2.0	2017	41	Inactive

Table 1.1: A selection of well-known OSS in healthcare

However, given the large amount of OSS in healthcare, how to choose the most appropriate OSS to solve a particular healthcare problem is challenging. For instance, there are numerous OSS for EHR: OpenEMR, OSEHRA VistA, OpenMRS, GNU Health, OSCAR, Hospital OS, and ClearHealth; each of them has its own strengths and weaknesses (Purkayastha et al., 2019). Which OSS to use in building an EHR system is a difficult question. As a challenging topic, reusing OSS still faces many difficulties, particularly in the healthcare domain (Shen. and Spruit., 2019). The lack of resources to support the selection of OSS from tens of thousands of open source projects is one of them. Although systematic literature studies on clinical OSS could help us obtain a deeper understanding of the existing tools and their performances (Rehim et al., 2017; Guaitoli et al., 2014; Marien et al., 2017), they are not particularly useful in selecting OSS since they cover only a small proportion of the vast amount of clinical software. Furthermore, the rapid updates of OSS in healthcare are not reflected appropriately due to the slow scientific publication process.

1.4 Research Questions

As previously mentioned, HIS engineering plays a crucial role in addressing various clinical challenges, and facilitating the digitalization of healthcare. However, developing and implementing HIS in today's healthcare organizations still faces many challenges. Therefore, this dissertation poses the following main research question:

> **MRQ** – How can we employ artificial intelligence technologies – such as machine learning algorithms, knowledge systems and natural language processing techniques – based on open source principles to accelerate healthcare information system engineering in solving real-world clinical problems?

To answer the main research question, we formulate six research questions. The first three questions focus on a specific use case. First, we design an intelligent clinical decision support system for addressing polypharmacy reviews. Then, based on the use case, the next three research questions investigate the use of NLP for different biomedical/clinical challenges. The opportunities of open source clinical software for clinical research and practices are also explored in these questions.

> **RQ1** – How can we design a GDPR-compliant clinical decision support system that supports a multi-center multi-lingual clinical trial?

Polypharmacy, defined as the chronic use of multiple medications by a patient, often leads to severe clinical problems or accidents if it is inappropriate, including adverse drug effects, underprescribing, overtreatment, low patient compliance and decreased drug adherence (Björkman et al., 2002; Claxton et al., 2001; Munger, 2010; Steinman et al., 2006; Wright et al., 2009). The Systematic Tool to Reduce Inappropriate Prescribing (STRIP) is a clinical intervention method crafted to deal with polypharmacy problems which are incurred by the concurrent use of multiple drugs. STRIP has been proven to be effective and is included in the Dutch national guideline for polypharmacy. To boost the usage of STRIP in clinical practices, we developed a web application called the STRIP Assistant (STRIPA) and a number of Dutch physicians further evaluated it in terms of user-friendliness, efficiency and effectiveness (Meulendijk et al., 2015a,b, 2016). However, to use the Dutch-based STRIPA tool in a large multinational randomized clinical trial (RCT), we need to address several issues, including multilingual support, clinical data security, data accessibility, consistency and GDPR-compatibility. Therefore, chapter 2 presents an overhauled STRIPA prototype with a lightweight data integration component that supports multinational implementations, ensures data consistency across countries, and maintains data accessibility and security.

RQ2 – How can we improve rapid and cost-effective development of clinical NLP systems with external NLP APIs?

Natural language processing (NLP) has become essential for secondary use of clinical data. Over the last two decades, many clinical NLP systems were developed in both academia and industry. However, nearly all existing systems are restricted to specific clinical settings mainly because they were developed for and tested with specific datasets, and they often fail to scale up. Therefore, using existing NLP systems for one's own clinical purposes requires substantial resources and long-term time commitments for customization and testing. Moreover, the maintenance is troublesome and time-consuming. Therefore, chapter 2 presents a lightweight approach for building clinical NLP systems with limited resources. Following the design science research approach, we propose a lightweight architecture which is designed to be composable, extensible, and configurable. It takes NLP as an external component which can be accessed independently and orchestrated in a pipeline via web APIs. To validate its feasibility, we developed a web-based prototype for clinical concept extraction with six well-known NLP APIs and evaluated it on three clinical datasets.

RQ3 – How can we utilize NLP techniques to automatically extract adverse drug reactions from the summary of product characteristics in European drug product labels?

Drug product labels are regulatory documents required as part of the marketing authorization of each medicine. They provide up-to-date and comprehensive information about the risks, benefits, and pharmacological properties of marketed medicines. As such, extracting the clinical knowledge stored in product labels and making it available in the form of computationally accessible knowledge bases would benefit several applications in the area of drug safety surveillance and assessment (Banda et al., 2016; Roberts et al., 2017). Therefore, in chapter 3 we investigate how to utilize NLP techniques to automatically extract adverse drug reactions (ADR) from standardized European product labels, namely SmPC. To answer the question, we develop an NLP pipeline to extract adverse drug reactions from SmPC.

RQ4 – How can we design an open source big data framework to facilitate cost-effective, large-scale, biomedical literature mining?

A huge amount of biomedical literature has been produced over the decades. For example, as the leading biomedical literature database, PubMed Central (PMC) has archived over 5.3 million research papers, of which around 2.4 million full-text articles are easily accessible at the PMC Open Access Subset (PMC OA Subset) (Pubmed Central, 2019; Pubmed Open Access, 2019). The massive size of available biomedical literature requires researchers to utilize novel big data technologies in data storage and analysis. Among them is cloud computing which has become the most popular solution for big data applications in industry. Therefore, in chapter 4 we propose a cloud-based big data framework that enables researchers to perform reproducible and scalable large-scale biomedical literature mining in an efficient and cost-effective way. Additionally, a cloud agnostic platform was constructed and then evaluated on two open access corpora with millions of full-text biomedical articles. The results demonstrate that our framework indeed supports scalable and efficient large-scale biomedical literature mining.

> **RQ5** – How can we support clinical developers' decisions in the software development process with a reproducible and scalable method for systematic studies on open source clinical software?

The plethora of open source projects offers great reuse opportunities for developers to build tools at lower cost and at a faster pace. Therefore, software reuse, also called code reuse, has become an essential topic in software development (Frakes and Fox, 1996; Zaimi et al., 2015). Zhang and Ho (2017) have recognized this importance and have called for more reuse of open source projects in clinical settings (Zhang and Ho, 2017). However, effective software reuse still faces many difficulties, particularly in clinical software development (de Oliveira, 2015). To address the problem, chapter 5 contributes to clinical software reuse research by conducting a large-scale analysis of open source repositories in the clinical domain. In particular, its purpose is to shed light on the following questions, so that clinical developers can make more informed decisions: 1). what is the current status of open source clinical software? 2). what are the impacts of various factors, such as the number of contributors and the number of commits, on the popularity of an open source clinical software? 3). what are the main focus areas of all the collected open source clinical software?

> **RQ6** – How can we design an online platform where clinical developers can easily locate and confidently select appropriate open-source clinical software based on associated scientific literature?

Numerous open-source tools have been created across a variety of domains after decades of open source software advocacy (Anthes, 2016). With the accelerating popularity of open science, more and more tools will be added to open repositories. Open-source clinical software covers various research topics and clinical practices, including medical image analysis (McCormick et al., 2014), medical text processing (Cunningham et al., 2013), clinical trials management systems (Haak et al., 2016), and electronic health record systems (de Abajo and Ballestero, 2012). The plethora of tools offers great opportunities for both clinical researchers and practitioners to accelerate their work with available open-source tools and algorithms. However, it also leaves many people unable to easily locate the tools most suitable for their clinical research or practices. Therefore, chapter 6 addresses this challenge by designing a search tool that links open-source clinical software to their literature. With this LOCATE tool, clinical researchers and practitioners are able to find literature related to open-source clinical software of their interest so that they can make better informed decisions.

1.5 Research Methods

The dissertation adopts the Design Science Research (DSR) framework. Different DSR approaches are used to answer the above research questions. The DSR framework is a well-known and widely used research method in information systems (Gregor and Hevner, 2013). Known for its strength and popularity in solving real-world problems by designing and building innovative IT artifacts (Hevner et al., 2008), this research follows the design science research methodology (DSRM) proposed by Peffers et al. (2007), which consists of six steps: problem identification and motivation, a definition of the objectives for a solution, design and development, demonstration, evaluation, and communication. Figure 1.1 explains how these steps are applied in the studies. Furthermore, the following methods have been applied in this dissertation: prototyping, computational experiment, and quantitative analysis.

1.5.1 Prototyping

In information systems development, prototyping is a quick and inexpensive way to improve requirements, commitment from end-users, and quality of code (Beynon-Davies et al., 1999). Prototypes can be built in the early stages, eliciting or validating requirements, in the middle stages, confirming the behavior of a system, or in the late stages, investigating the operational system. In addition to a development tool, prototyping can also be used as a research method, to increase knowledge and understanding of a system (O'Leary, 1988). Building such a prototype can additionally be used to quantify the performance of a system. In this research, we use prototyping in four chapters (chapters 2 to 4 and 7). In each study, a working prototype with all key functionalities of the system is constructed to validate its workings.

1.5.2 Computational experiments

An experiment is one of the common quantitative research methods in research, which investigates the effect of modifying an independent variable on the dependent variable. The focus of a computational experiment lies in the theoretical analysis and empirical testing of a computational method, such as approximation or optimization algorithms. Zelkowitz and Wallace (1998) for example categorize experimental research in software engineering into observational, historical, and controlled methods. In this work, three chapters (chapters 3 to 5) use computational experiments to conduct their research.

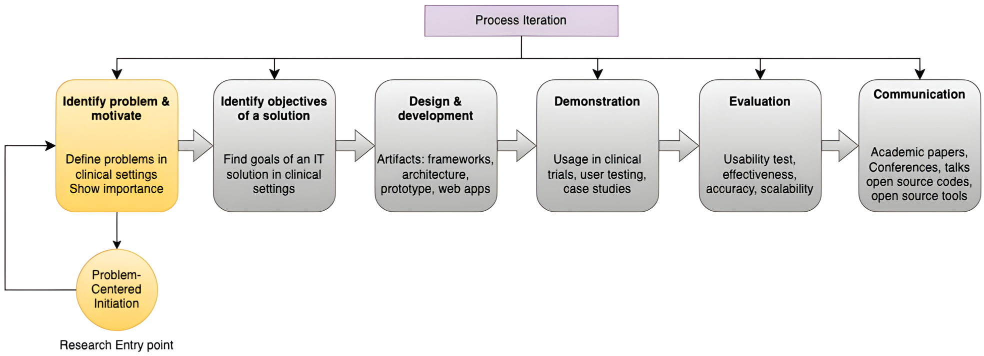

Figure 1.1: Process of design science research methodology

1.5.3 Quantitative analysis

Creswell et al. (2004) define quantitative analysis as a research method that involves the collection of quantified data and statistical analysis for supporting or refuting claims. Different statistical methods are applied when it comes to different collected data. Numerical and textual data are common data types collected in research. Researchers use statistical models to analyze numerical data to discover trends, correlations and other insights. For textual data, natural language processing techniques are applied before using any statistical models. For example, chapter 6 collects both numerical and textual data and performs quantitative analysis on them.

1.6 Dissertation Outline

Chapters 2-7 of the dissertation investigate the research questions RQ1-RQ6, with each chapter corresponding to one research question. All chapters are written as papers, published in proceedings of scientific conferences or in scientific journals. Table 1.2 offers an overview of Chapters 2-7 and their corresponding research question ID, research methods, and a summary of the datasets used in each chapter.

Ch.	RQ	Research Methods	Data	Outcomes	Source codes
2	RQ1	Prototyping	Table 2.1 in Chapter 2	Architecture and Web Application running in production to support multinational clinical trials	Per request
3	RQ2	Prototyping, computational experiment, Quantitative analysis	100 EHRs in free-text and manual annotation of these documents	A prototype (web application) that processes free-text EHRs	Yes (GitHub)
4	RQ3	Prototyping, computational experiment, Quantitative analysis	Adverse drug reactions sections for 647 common medicines scraped from the Electronic Medicines Compendium	A reusable data pipeline in Python and structured ADRs knowledge base	Yes (GitHub)
5	RQ4	Computational experiment, Quantitative analysis	Over one million full-text biomedical articles from PubMed Open Access Subset; 29437 labeled full-text biomedical articles from PubMed	A framework supports big clinical data analysis	Yes (GitHub)
6	RQ5	Computational experiment, Quantitative analysis	Metadata of 14971 clinical-related open-source GitHub repositories	A resuable data pipeline for conducting systematic studies on GitHub	Yes (GitHub)
7	RQ6	Prototyping	Metadata of 5119 clinical-related open-source GitHub repositories; 8820 scientific papers that related to the selected repositories	A public web application as prototype	Yes (GitHub)

Table 1.2: Overview per research question of the research methods and datasets used in Chapters 2-7

1.7 Dissertation outcomes

This section lists used and created datasets, available source codes, implemented prototypes and scientific publications of this thesis. Table 1.2 outlines the research questions, methods and outcomes from Chapter 2-7.

1.7.1 Data

The i2b2 2008 Obesity Challenge - consists of 611 discharge letters. The data provider (National Center for Biomedical Computing) annotated all discharge letters with 16 different medical condition terms in the context of obesity, including asthma, gastroesophageal disorder, and depression. Terms could be either annotated as being in the document, not being in the document, or undecided/unknown which was treated as not being in the document. The dataset is available at `https://www.i2b2.org/NLP/Obesity`.

The i2b2 2009 Medication Challenge - contains 1243 de-identified discharge letters with the gold standard annotations. Medication names in the annotations are used for the evaluation. The dataset is available at `https://www.i2b2.org/NLP/Medication`.

Adverse drug reactions corpora - contains section 4.8 Undesirable effects of the Summary of Product Characteristics (SmPC) of 647 medications scraped from the Electronic Medicines Compendium (EMC). This data is either free-text or tabular text. The datasets are available at `https://github.com/ianshan0915/ade-extraction/tree/master/data`.

PubMed Open Access Subset - has millions of journal articles and preprints that are made available under license terms that allow reuse. The dataset was obtained via the PMC FTP service as zip files and it contains 1010787 full-text biomedical articles (Shen et al., 2019b). This data could be found at `https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist`.

Labeled full-text biomedical texts from PubMed - contains 29437 full-text biomedical articles from PubMed which are divided into three groups: breast cancer, lung cancer, and prostate cancer (Ye et al., 2016). This dataset was used for evaluating document classification models (Ye et al., 2016; Shen et al., 2019b) and is available at `https://figshare.com/articles/dataset/New_draft_item/3796302`.

1.7.2 Code and Prototypes

The STRIP Assistant (STRIPA) is a web application that was built to help physicians to perform medication reviews in the case of polypharmacy. It was used in a large-scale randomized clinical trial (RCT) across four European countries: the

Netherlands, Switzerland, Ireland and Belgium. A demo of the tool is available at `https://youtu.be/GMhge8GxaVk`.

MAB-NLP is a prototype that extracts clinical concepts from clinical free-texts, such as discharge summaries, clinical reports, etc. It was developed based on a scalable API-based architecture using Laravel. The code is available at: `https://github.com/ianshan0915/MABNLP`. A demo of the tool is available at `https://youtu.be/dGk9NQGWYfI`.

Automatic Extraction of Adverse Drug Reactions is a natural language processing pipeline that automatically scrapes the online summary of product characteristics of medications and then extracts structured adverse drug reactions from them. The code is available at: `https://github.com/ianshan0915/ade-extraction`.

LDABiotext is a spark application using a Latent Dirichlet Allocation (LDA) model to infer topics from a large collection of biomedical literature from the PMC OA Subset. It is written in Scala and can be deployed to a cloud infrastructure of your choice. The code is available at: `https://github.com/ianshan0915/Spark-LDA-biomedical-text`.

GitHub Data Pipeline is used to collect information about GitHub repositories via the GitHub API. Data analysis is conducted to uncover knowledge of the collected repositories. The code is available at: `https://github.com/ianshan0915/clinical-opensource-projects`.

LOCATE is a web application that is useful for both IT practitioners and researchers in the clinical community. It enables practitioners to explore related literature for a given open-source clinical software so that they could make an informed decision on which software to choose. Clinical researchers are provided with a list of potential useful open-source tools based on their research topics. Details of the prototype and its source codes are available at: `https://github.com/ianshan0915/locate`.

1.7.3 Publications

Publications that are related to the thesis:

1. Shen, Z., Meulendijk, M., and Spruit, M. (2016b). A federated information architecture for multinational clinical trials: Stripa revisited. *24th European Conference on Information Systems (ECIS)*. `https://aisel.aisnet.org/ecis2016_prototypes/2`

2. Shen, Z., van Krimpen, H., and Spruit, M. (2019a). A lightweight api-based approach for building flexible clinical nlp systems. *Journal of Healthcare Engineering*, 2019(1):3435609. `https://onlinelibrary.wiley.com/doi/abs/10.1155/2019/3435609`

3. Shen, Z. and Spruit, M. (2021). Automatic extraction of adverse drug reactions from summary of product characteristics. *Applied Sciences*, 11(6). `https://www.mdpi.com/2076-3417/11/6/2663`

4. Shen, Z., Wang, X., and Spruit, M. (2019b). Big data framework for scalable and efficient biomedical literature mining in the cloud. In *Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*, NLPIR '19, page 80–86, New York, NY, USA. Association for Computing Machinery. `https://doi.org/10.1145/3342827.3342843`

5. Shen, Z. and Spruit, M. (2019). A systematic review of open source clinical software on github for improving software reuse in smart healthcare. *Applied Sciences*, 9(1). `https://www.mdpi.com/2076-3417/9/1/150`

6. Shen., Z. and Spruit., M. (2019). Locate: A web application to link open-source clinical software with literature. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019) - HEALTHINF*, pages 294–301. INSTICC, SciTePress. `https://doi.org/10.5220/0007378702940301`

1.7.4 Other Publications by the Author

1. van Tuijl, G., Leenen, W., **Z. Shen**, van de Weerd, I., and **S. Brinkkemper** (2011). Prioritizing requirements: An experiment to test the perceived reliability, usability and time consumption of bubblesort and the analytical hierarchy process. In *Proceedings of the International Requirements Engineering Efficiency Workshop (REEW 2011)*, pages 37–48. `https://api.semanticscholar.org/CorpusID:53900285`

2. Crowley, E. K., Sallevelt, B. T., Huibers, C. J., Murphy, K. D., **Spruit, M.**, **Shen, Z.**, Boland, B., Spinewine, A., Dalleur, O., Moutzouri, E., et al. (2020). Intervention protocol: Optimising therapy to prevent avoidable hospital admission in the multi-morbid elderly (operam): a structured medication review with sup-

port of a computerised decision support system. *BMC health services research*, 20(1):1–12. https://doi.org/10.1186/s12913-020-5056-3

3. Omta, W. A., van Heesbeen, R. G., **Z.Shen**, de Nobel, J., Robers, D., van der Velden, L. M., Medema, R. H., Siebes, A. P., Feelders, A. J., **S. Brinkkemper**, Klumperman, J. S., **M. Spruit**, Brinkhuis, M. J., and Egan, D. A. (2020b). Combining supervised and unsupervised machine learning methods for phenotypic functional genomics screening. *SLAS Discovery*, 25(6):655–664. https://doi.org/10.1177/2472555220919345

4. Omta, W. A., van Heesbeen, R. G., **Shen, Z.**, Feelders, A. J., Brinkhuis, M., Egan, D. A., and **Spruit, M.** (2020a). Purifyr: An r package for highly automated, reproducible variable extraction and standardization. *Systems Medicine*, 3(1):1–7. https://api.semanticscholar.org/CorpusID:215898718

5. Blum, M. R., Sallevelt, B. T. G. M., Spinewine, A., O'Mahony, D., Moutzouri, E., Feller, M., Baumgartner, C., Roumet, M., Jungo, K. T., Schwab, N., Bretagne, L., Beglinger, S., Aubert, C. E., Wilting, I., Thevelin, S., Murphy, K., Huibers, C. J. A., Drenth-van Maanen, A. C., Boland, B., Crowley, E., Eichenberger, A., Meulendijk, M., Jennings, E., Adam, L., Roos, M. J., Gleeson, L., **Shen, Z.**, Marien, S., Meinders, A.-J., Baretella, O., Netzer, S., de Montmollin, M., Fournier, A., Mouzon, A., O'Mahony, C., Aujesky, D., Mavridis, D., Byrne, S., Jansen, P. A. F., Schwenkglenks, M., **Spruit, M.**, Dalleur, O., Knol, W., Trelle, S., and Rodondi, N. (2021b). Optimizing therapy to prevent avoidable hospital admissions in multimorbid older adults (operam): cluster randomised controlled trial. *BMJ*, 374

6. Sallevelt, B., Huibers, L., Op Heij, J., Egberts, T., Puijenbroek, E., **Shen, Z.**, **Spruit, M.**, Jungo, K., Rodondi, N., Dalleur, O., Spinewine, A., Jennings, E., O'Mahony, D., Wilting, I., and Knol, W. (2021). Frequency and acceptance of clinical decision support system-generated stopp/start signals for hospitalised older patients with polypharmacy and multimorbidity. *Drugs & Aging*, 39. https://doi.org/10.1007/s40266-021-00904-z

7. Shen, Z., Meulendijk, M., Knol, W., Huibers, L., Wilting, I., Jansen, P., and Spruit, M. (2016a). Stripa investigational medical device dossier (imdd). Technical report, UU/UMCU. https://marcospruit.nl/pub/2016%20-%20Shen%20et%20al%20-%20IMDD.pdf