Structuring Clinical Notes of Italian ST-elevation Myocardial Infarction Patients

Vittorio Torri[1], Sara Mazzucato[2], Stefano Dalmiani[3], Umberto Paradossi[3], Claudio Passino[3,4], Sara Moccia[2], Silvestro Micera[2,4,5], Francesca Ieva[1,6]
[1]MOX - Modelling and Scientific Computing Lab, Department of Mathematics,
Politecnico di Milano, Italy
{name.surname}@polimi.it
[2]Biorobotics Institute, Department of Excellence in Robotics and AI,
Scuola Superiore Sant'Anna, Italy
{name.surname}@santannapisa.it
[3]Fondazione Toscana Gabriele Monasterio, Italy
{dalmiani, uparadossi, passino}@ftgm.it
[4]Health Science Interdisciplinary Center, Scuola Superiore Sant'Anna, Italy
[5]Bertarelli Foundation Chair in Translational Neural Engineering,
Center for Neuroprosthetics and Institute of Bioengineering,
Ecole Polytechnique Federale de Lausanne, Switzerland
[6]HDS - Health Data Science Centre, Human Technopole, Italy

Abstract

In recent years, it has become common for patients to get full access to their Electronic Health Records (EHRs), thanks to the advancements in the EHRs systems of many healthcare providers. While this access empowers patients and doctors with comprehensive and real-time health information, it also introduces new challenges, in particular due to the unstructured nature of much of the information within EHRs. To address this, we propose a pipeline to structure anamneses, providing patients with a clear and concise overview of their health data and its longitudinal evolution, also allowing clinicians to focus more on patient care during consultations. In this paper, we present preliminary results on extracting structured information from EHRs of patients diagnosed with ST-Elevation Myocardial Infarction from an Italian hospital. Our pipeline exploits text classification models to extract relevant clinical variables, comparing rule-based, recurrent neural network and BERT-based models. While various approaches utilized ontologies or knowledge graphs for Italian data, our work represents the first attempt to develop this type of pipeline. The results for the extraction of most variables are satisfactory (f1-score > 0.80), with the exception of the most rare values of certain variables, for which we propose future research directions to investigate.

Keywords: Natural Language Processing, Clinical Notes, EHR Summarization, ST-elevation myocardial infarction

1. Introduction

In the past decades, the adoption of Electronic Health Records (EHRs) has become widespread among healthcare providers. In recent years EHRs have also granted direct access to patients, bypassing the need for a physician's mediation (Klein et al., 2016). This advancement offers patients numerous benefits, including immediate access to their latest exam results and allows them to review their medical history at their convenience, ultimately enhancing the relational aspect of care (Blease et al., 2020). However, a significant portion of EHRs is still in the format of unstructured documents, with only a fraction of their data available in structured formats (Rosenbloom et al., 2011; Tayefi et al., 2021). This lack of structure complicates the work of physicians since, despite their familiarity with such documents, they still require substantial time to extract pertinent information, consequently impacting their interactions with patients (Friedberg et al., 2014).

A potential solution lies in leveraging Natural Language Processing (NLP), a field that has witnessed remarkable advancements in recent years. However, applying NLP models in the medical domain presents challenges due to the unique formats and terminology inherent in medical documents. Despite the attempts of the most recent models to achieve a certain level of multilingualism, the highest performances in NLP models predominantly continue to occur with English documents, which also serve as the primary focus for most benchmarks (Hedderich et al., 2021; Lai et al., 2023). The intersection of these two areas of complexity poses significant hurdles (Névéol et al., 2018).

In this study, we present preliminary findings on developing a pipeline to extract structured data from EHRs of patients diagnosed with ST-Elevation Myocardial Infarction (STEMI) at Fondazione Toscana Gabriele Monasterio (FTGM), an Italian hospital specialized in cardiology. Specifically, we analyze a dataset comprising 9275 clinical notes pertaining to 1730 patients, manually annotated by clinicians to identify the most relevant risk factors, comor-

Proceedings of CL4Health 2024 @LREC-COLING-2024, pages 37–43
20 May, 2024.

bidities, and clinical characteristics associated with STEMI.

To the best of our knowledge, this represents the first attempt to extract such granular clinical details, as the site of STEMI or the presence and location of stenosis, from Italian EHRs. Since all these features are categorical, we develop and compare different approaches for their extraction, ranging from rule-based to recurrent neural networks and transformer-based models, showing that different types of features necessitate distinct models based on their complexity levels. Current results are promising and we believe that this pipeline will enhance patients' experiences, both directly as they access their EHRs and indirectly during consultations with physicians.

2. Related Works

The sole study that has undertaken a similar approach with Italian documents is (Viani et al., 2019) but notable distinctions exist between their work and ours. They analyzed 75 cardiology reports, focusing on event extraction and specific attributes associated with these events. Consequently, their task is modelled as entity extraction, followed by event classification into four categories (problem, test, treatment, and occurrence) and with respect to some attributes (DocTimeRel, Polarity, Modality, and Experiencer). While their work is significant, we note that our task delves into extracting more detailed structured information, thus presenting a higher level of complexity. Additionally, the substantially larger dataset in our possession provides greater confidence in the generalizability of our results. In (Viani et al., 2018) a similar task was attempted with an ontology-based approach, requiring an elevated effort in the development of the ontology. Similarly, out of the cardiology domain, (Agnello et al., 2021) and (D'Auria et al., 2023) used ontologies and knowledge graphs to extract and link entities from Italian clinical notes.

Out of the Italian domain, as far as our knowledge extends, there exist no specific applications of NLP dedicated to extracting structured information from documents of STEMI patients. Expanding our scope to the wider cardiology domain, we find a few studies on data extraction from EHRs. Some rely solely on rule-based systems (e.g., Patterson et al., 2017, Berman et al., 2021), while others leverage BERT-based models (e.g., Silva et al., 2020, Richter-Pechanski et al., 2021, Singh et al., 2022) or the MedCat tool (Shek et al., 2021). However, these studies only focus on comorbidities or numerical measurements. Alternatively, there are works related to utilizing NLP for cardiology patient classification (e.g., Afzal et al., 2017, Ambrosy et al., 2021, Zaman et al., 2021, Berman et al., 2023). Yet, none of them align with our objective of identifying more granular clinical details, such as the site of STEMI or the presence and location of stenosis.

3. Material and Methods

3.1. Dataset

The dataset consists of 9275 EHRs obtained from STEMI patients at FTGM, a specialized cardiology hospital situated in Pisa, Italy. It covers a large period of time from May 2006 to April 2023, and encompasses records from 1730 patients. The experimental protocol was approved by the FTGM ethical committee. Informed consent was obtained from all participants according to the declaration of Helsinki. Each EHR corresponds to a note written by a physician after a specialized patient examination. On average, each patient has 4 records (with a first quartile of 2 and a third quartile of 7). The length of the notes ranges from 1040 to 2047 characters, with a median length of 1677 characters. Clinicians manually annotated these notes with a set of structured or semi-structured variables using Excel software, so the dataset serves as both training and test data for our pipeline.

Table 1 provides a summary of the variables extracted, delineating their types and values. It is worth mentioning that Coronary Artery Bypass Graft surgery history (CABG) and Percutaneous Transluminal Coronary Angioplasty history (PTCA) variables, along with the boolean indicator of non-culprit stenosis, exhibit a large number of missing values. This is primarily because it is not always feasible to assess the positive or negative values of these variables from the textual content of the note since in clinical documentation the primary focus is often on acute medical concerns rather than historical procedures like CABG or PTCA.

The location of the non-culprit stenosis is considered a semi-structured variable because it is reported as free text in the annotations, yielding 272 unique values. We manually consolidated these values into 12 distinct locations, acknowledging that each value might encompass multiple locations. These locations were further condensed into the same five groups used for the culprit vessel.

3.2. Methods

Our objective is to develop a pipeline able to extract these variables from clinical notes as they are frequently solely reported within free-form text within clinical practice. Due to the categorical nature of the clinical variables of interest and the semi-structured variable can be mapped to categories, we approach the problem of their extraction from the text of the clinical notes as a text classification problem. The workflow is reported in Figure 1. Consequently, our

Variable	Description	Type	Struct. Values (distribution)	% NA
SK	Smoker	S	Yes (41.8%) / No (58.2%)	1.6%
DB	Diabetes	S	Yes (19.5%) / No (80.5%)	1.8%
HC	Hypercholesterolemia	S	Yes (46.0%) / No (54.0%)	1.9%
HT	Hypertension	S	Yes (57.8%) / No (42.2%)	1.6%
CAD	Coronary Artery Disease Family history	S	Yes (33.6%) / No (66.4%)	1.6%
MI	Myocardial Infarction history	S	Yes (9.2%) / No (90.8%)	1.6%
CABG	Coronary Artery Bypass Graft surgery history	S	Yes (6.1%) / No (93.9%)	82.9%
PTCA	Percutaneous Transluminal Coronary Angioplasty history	S	Yes (22.3%) / No (77.7%)	79.3%
ECG	STEMI location from Diagnostic ECG	S	Front (42.0%) / DX (39.1%) / Lat (17.1%) / Post (0.8%) / Negative (1.1%)	1.8%
CV	Culprit vessel	S	DX (36.0%) / IVA-DA (42.4%) / CX (11.9%) / TC (1.0%) / Other (8.7%)	2.1%
SNC	Presence of a non-culprit stenosis	S	Yes (55.0%) / No (45.0%)	46.4%
LS	Non-culprit stenosis location*	SS	DX (26.7%) / IVA (39.9%) / CX (30.1%) / TC (4.7%) / Other (19.4%)	8.8%

Table 1: Summary information for the variables to be extracted from the notes. S = Structured, SS = Semi-structured.*LS numbers are only applicable to records with a positive value for SNC; their percentages may exceed 100% due to possible multiple locations. For coronary arteries the corresponding English terms are: DX = RCA, IVA = LDA, CX = LCx, TC = LMCA

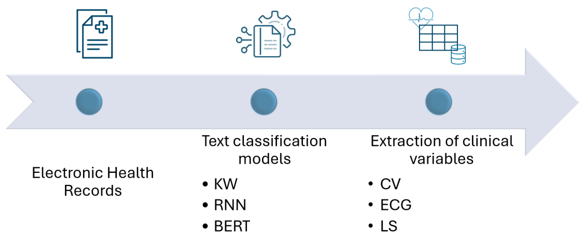

Figure 1: Schema of the proposed pipeline

pipeline consists of a series of text classification models, some of which also require an additional pre-processing of the text: (i) Keyword-based (KW), (ii) Recurrent Neural Network (RNN), (iii) Bidirectional Encoder Representations from Transformers-based (BERT).

The KW model examines the presence of predefined keywords and their synonyms within the text associated with each variable. To facilitate this analysis, the text undergoes lowercase conversion and lemmatization. We employ the Spacy *it_core_news_sm* model for lemmatization [1].

The RNN model architecture is composed of an embedding layer, followed by an Long Short-Term Memory layer and a fully connected layer, each integrated with dropout. Training the network involves utilizing the Adam optimizer with early stopping and cross-entropy as loss function. The hyper-parameters, including layer sizes, dropout rates, and learning rates, are selected via grid search. Preprocessing for this model involves lowercasing, lemmatization, and the removal of words with very low frequencies ($< 0.5\%$), in addition to eliminating Italian stopwords defined in the *nltk* Python library (Bird et al., 2009) and punctuation.

The BERT-based model is a fine-tuned adaptation of the latest Italian version of BERT, Umberto (Tamburini et al., 2020). We further pre-trained the Umberto model on a publicly available corpus of clinical documents (Bernardo Magnini et al, 2020). Since this corpus includes documents in five languages, we automatically translated into Italian, using Google Translate, all documents that were in other languages. This additional pre-training allows the model to acquire knowledge of Italian medical terms. In our classification tasks, fine-tuning occurs solely in the last two layers of the transformer model, while the previous ones are kept frozen during training. The final layer consist of a fully connected layer with sigmoid activation function for binary variables and softmax for multiclass variables. For the location of the non-culprit stenosis, the final layer has multiple binary outputs, one for each potential location. The model is trained with the AdamW optimizer with early stopping and using cross-entropy as loss function. No preprocessing steps are needed since the BERT base model is pre-trained on data without any specific preprocessing.

[1] https://spacy.io/models/it#it_core_news_sm

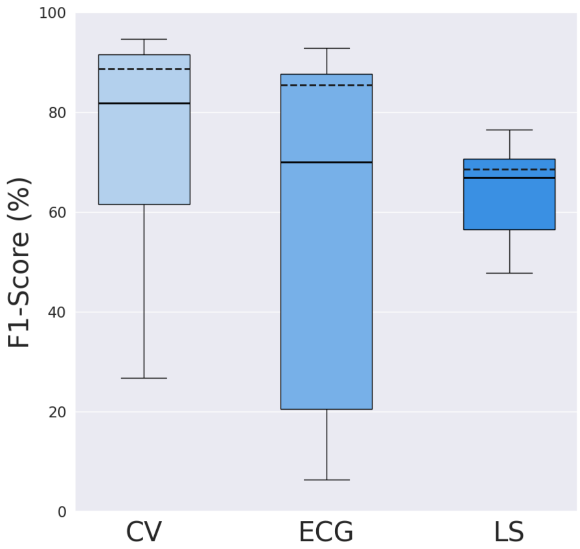

Figure 2: F1-score for the non-binary clinical variables over their possible values. Dashed lines correspond to weighted average.

The models are evaluated with 10-fold stratified cross-validation. Evaluation metrics include precision, recall, f1-score, and area under the receiver operating characteristic curve (AUC).

4. Results

In Table 2, we report the results for each binary variable, comparing the metrics among the three models. In Figure 2, we outline the results for non-binary variables. We report boxplots of the f1-score distribution over the different values that can be assumed by these variables, and we highlight the weighted average. This is relevant as these variables exhibit highly unbalanced value distributions, resulting in outstanding performance for common values but potentially poor outcomes for rare ones. The least favourable results are associated with *Posterior* and *Negative* values of ECG and the *TC* value of both CV and LS. For these variables the best model is BERT.

5. Discussion and Conclusions

The study presents a novel method for extracting structured information from Italian unstructured EHRs, with a focus on STEMI patients. By developing a pipeline and evaluating various classification models, our preliminary results show encouraging outcomes for many variables, with notable achievements such as an f1-score of 89.6% for HT. However, they are less satisfactory for the less frequent values of some variables, like CV, ECG and LS. It is not possible to have a precise comparison with previous works, due to the different variables and type of data, but for the comorbidities variables our results are aligned with similar works reported in Section 2. Furthermore, our study contributes to the field by evaluating various NLP approaches for Italian data, representing the first attempt to develop such a pipeline in this language. Comparing the different approaches, BERT outperforms in variables requiring deeper contextual understanding and semantics as the case of SNC, whereas for simpler variables, RNN suffices. Specifically, in the case of CABG where the best model results KW, it is expressed very clearly with a few keywords, rendering the use of more complex methods unnecessary. To address the less frequent values, in future work, we propose testing generative models to generate additional training examples for these values. This might be helpful also to cover certain types of expressions that are less frequent and less clear to interpret, such as *"pregresso intervento rivascolarizzazione miocardica mediante triplice BPAC" ("Previous myocardial revascularization surgery by triple CABG")* which requires knowing that a triple CABG implies a stenosis on TC (LMCA), CX (LCx) and IVA (LAD). Another potential expansion to be explored is a joint model to extract multiple variables simultaneously, enhancing the efficiency of the pipeline. Additionally, the integration of explainability methods could provide insights into how the model makes decisions, improving its interpretability and trustworthiness. We also defer the comparison with open-source large language models to future work. A limitation of our study is its restriction to data sourced from a single center and from patients affected by a single specific disease (STEMI). To address this problem, validation on a different dataset would be beneficial to ensure the generalizability of our findings across diverse patient populations and healthcare settings. Despite the identified limitations, our pipeline holds significant utility for patients. By organizing the data embedded within their EHRs, we provide clinicians and patients with a more transparent comprehension of their health status and treatment possibilities. This not only encourages patient involvement in their healthcare decisions but also nurtures deeper interactions between patients and physicians during clinical encounters. Therefore, our research contributes to the continuous advancement of patient care and healthcare delivery through the innovative utilization of NLP technologies.

6. Acknowledgements

The present research is part of the activities of “Dipartimento di Eccellenza 2023-2027”. Work partially supported by Proximity Care Project and #NEXTGENERATIONEU (NGEU) and funded by

Metric	Model	SK	DB	HC	HT	CAD	MI	CABG	PTCA	SNC
P	KW	70.0 %	52.6 %	82.4 %	83.7 %	60.4 %	23.2 %	90.1 %	37.8 %	63.1 %
P	RNN	**90.1 % (4.4 %)**	71.4 % (2.1%)	**88.4 % (2.7 %)**	87.8 % (3.4 %)	**91.6 % (3.1 %)**	67.5 % (1.5 %)	**93.4 % (2.2 %)**	**86.5 % (4.3 %)**	77.8 % (2.1 %)
P	BERT	86.0 % (1.4 %)	**80.4 % (3.2%)**	81.9 % (1.6 %)	**89.6 % (1.0 %)**	88.1 % (0.9 %)	**70.1 % (3.1 %)**	89.0 % (4.1 %)	84.9 % (5.2 %)	**84.8 % (1.9 %)**
R	KW	79.9 %	**79.9 %**	23.7 %	63.0 %	35.6 %	53.7 %	**84.5 %**	21.3 %	37.4 %
R	RNN	83.0 % (4.5 %)	78.9 % (2.6 %)	77.1 % (2.9 %)	**91.7 % (6.6 %)**	85.2 % (3.3 %)	**76.1 % (2.6 %)**	77.6 % (1.1 %)	**73.0 % (5.2 %)**	72.0 % (3.8 %)
R	BERT	**85.9 % (1.4 %)**	71.1 % (2.9 %)	**81.8 % (1.5 %)**	89.5 % (1.0 %)	**88.0 % (0.8 %)**	74.8 % (3.2%)	82.0% (2.0 %)	68.7% (8.8 %)	**83.8 % (5.2 %)**
F1	KW	74.6 %	63.4 %	36.8 %	71.9 %	44.8 %	32.4 %	**87.2 %**	27.3 %	46.7 %
F1	RNN	**86.9 % (4.5 %)**	**73.4 % (2.5 %)**	**83.5 % (3.0%)**	**89.6 % (9.6 %)**	**88.3 % (3.6 %)**	**71.5 % (3.0 %)**	82.9 % (2.8 %)	**78.8 % (4.3 %)**	74.6 % (3.1 %)
F1	BERT	85.9 % (1.4 %)	71.2 % (2.8 %)	81.8 % (1.6 %)	89.5 % (1.0 %)	88.0 % (0.9 %)	71.3 % (2.8 %)	85.4 % (2.9 %)	75.6 % (6.0 %)	**84.0 % (4.9 %)**
AUC	RNN	85.0 % (4.5 %)	**78.9 % (2.6 %)**	**83.7 % (2.9 %)**	**87.2 % (2.6 %)**	**88.7 % (3.4 %)**	69.4 % (2.6 %)	**79.5 % (5.8 %)**	**75.2 % (5.2 %)**	75.3 % (3.9 %)
AUC	BERT	**92.7 % (1.1 %)**	77.0 % (1.1 %)	82.8 % (1.3 %)	86.8 % (1.5 %)	86.5 % (1.2 %)	**72.5 % (0.7 %)**	79.4 % (5.1%)	73.3 % (2.8 %)	**82.0 % (4.5 %)**

Table 2: Results for the binary variables on 10-fold stratified cross-validation, reported as mean (std dev) for each model. P = Precision, R = Recall, F1 = F1-Score, AUC = Area Under the receiver operating characteristic Curve. Standard deviation is not reported for KW since there is no training set. AUC is not reported for KW since it does not output a probability. Best results for each metric and variable are highlighted in bold.

the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project BRIEF (IR0000036) - Biorobotics Research and Innovation Engineering Facilities (DN. 103 17.06.2022).

7. Bibliographical References

Naveed Afzal, Sunghwan Sohn, Sara Abram, Christopher G Scott, Rajeev Chaudhry, Hongfang Liu, Iftikhar J Kullo, and Adelaide M Arruda-Olson. 2017. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. Journal of vascular surgery, 65(6):1753–1761.

Patrizia Agnello, Silvia Maria Ansaldi, Fabio Azzalini, Giovanni Gangemi, Davide Piantella, Emanuele Rabosio, Letizia Tanca, et al. 2021. Extraction of medical concepts from Italian natural language descriptions. In 29th Italian Symposium on Advanced Database Systems, SEBD, volume 2994, pages 275–282.

Andrew P Ambrosy, Rishi V Parikh, Sue Hee Sung, Anand Narayanan, Rajeev Masson, Phuong-Quang Lam, Kevin Kheder, Alan Iwahashi, Alexander B Hardwick, Jesse K Fitzpatrick, et al. 2021. A natural language processing–based approach for identifying hospitalizations for worsening heart failure within an integrated health care delivery system. JAMA Network Open, 4(11):e2135152–e2135152.

Adam N Berman, David W Biery, Curtis Ginder, Olivia L Hulme, Daniel Marcusa, Orly Leiva, Wanda Y Wu, Nicholas Cardin, Jon Hainer, Deepak L Bhatt, et al. 2021. Natural language processing for the assessment of cardiovascular disease comorbidities: The cardio-Canary comorbidity project. Clinical Cardiology, 44(9):1296–1304.

Adam N Berman, Curtis Ginder, Zachary A Sporn, Varsha Tanguturi, Michael K Hidrue, Linnea B Shirkey, Yunong Zhao, Ron Blankstein, Alexander Turchin, and Jason H Wasfy. 2023. Natural Language Processing for the Ascertainment and Phenotyping of Left Ventricular Hypertrophy and Hypertrophic Cardiomyopathy on Echocardiogram Reports. The American Journal of Cardiology, 206:247–253.

Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

Charlotte Blease, Leonor Fernandez, Sigall K Bell, Tom Delbanco, and Catherine DesRoches. 2020. Empowering patients and reducing inequities: is there potential in sharing clinical notes? BMJ Quality & Safety, 29(10):1–2.

Daniela D'Auria, Vincenzo Moscato, Marco Postiglione, Giuseppe Romito, and Giancarlo Sperlí. 2023. Improving graph embeddings via entity linking: a case study on Italian clinical notes. Intelligent Systems with Applications, 17:200161.

Mark W Friedberg, Peggy G Chen, Kristin R Van Busum, Frances Aunon, Chau Pham, John Caloyeras, Soeren Mattke, Emma Pitchforth, Denise D Quigley, Robert H Brook, et al. 2014. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. Rand health quarterly, 3(4).

Paul Hager, Friederike Jungmann, Kunal Bhagat, Inga Hubrecht, Manuel Knauer, Jakob Vielhauer, Robbie Holland, Rickmer Braren, Marcus Makowski, Georgios Kaisis, and Daniel Rueckert. 2024. Evaluating and Mitigating Limitations of Large Language Models in Clinical Decision Making. medRxiv.

Michael A Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. 2021. A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 2545–2568.

Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.

Jared W Klein, Sara L Jackson, Sigall K Bell, Melissa K Anselmo, Jan Walker, Tom Delbanco, and Joann G Elmore. 2016. Your patient is now reading your note: opportunities, problems, and prospects. The American journal of medicine, 129(10):1018–1021.

Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. 2023. Chatgpt beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. arXiv preprint arXiv:2304.05613.

Bernardo Magnini, Begona Altuna, Alberto Lavelli, Manuela Speranza, and Roberto Zanoli. 2021. The E3C project: European clinical case corpus. Language, 1(L2):L3.

Aurélie Névéol, Hercules Dalianis, Sumithra Velupillai, Guergana Savova, and Pierre Zweigenbaum. 2018. Clinical natural language processing in languages other than English: opportunities and challenges. Journal of biomedical semantics, 9(1):1–13.

Olga V Patterson, Matthew S Freiberg, Melissa Skanderson, Samah J. Fodeh, Cynthia A Brandt, and Scott L DuVall. 2017. Unlocking echocardiogram measurements for heart disease research through natural language processing. BMC cardiovascular disorders, 17:1–11.

Phillip Richter-Pechanski, Nicolas A Geis, Christina Kiriakou, Dominic M Schwab, and Christoph Dieterich. 2021. Automatic extraction of 12 cardiovascular concepts from German discharge letters using pre-trained language models. Digital Health, 7:20552076211057662.

Joseph Root, Natalia V Oster, Sara L. Jackson, Roanne Mejilla, Jan Walker, and Joann G Elmore. 2016. Characteristics of patients who report confusion after reading their primary care clinic notes online. Health communication, 31(6):778–781.

S Trent Rosenbloom, Joshua C Denny, Hua Xu, Nancy Lorenzi, William W Stead, and Kevin B Johnson. 2011. Data from clinical notes: a perspective on the tension between structure and flexible documentation. Journal of the American Medical Informatics Association, 18(2):181–186.

Anoop D Shah, Emily Bailey, Tim Williams, Spiros Denaxas, Richard Dobson, and Harry Hemingway. 2019. Natural language processing for disease phenotyping in UK primary care records for research: a pilot study in myocardial infarction and death. Journal of biomedical semantics, 10(Suppl 1):20. PMID: 31711543.

Seyedmostafa Sheikhalishahi, Riccardo Miotto, Joel T Dudley, Alberto Lavelli, Fabio Rinaldi, Venet Osmani, et al. 2019. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR medical informatics, 7(2):e12239.

Anthony Shek, Zhilin Jiang, James Teo, Joshua Au Yeung, Ajay Bhalla, Mark P Richardson, and Yee Mah. 2021. Machine learning-enabled multitrust audit of stroke comorbidities using natural language processing. European Journal of Neurology, 28(12):4090–4097.

João Figueira Silva, João Rafael Almeida, and Sérgio Matos. 2020. Extraction of family history information from clinical notes: deep learning and heuristics approach. JMIR Medical Informatics, 8(12):e22898.

Pulkit Singh, Julian Haimovich, Christopher Reeder, Shaan Khurshid, Emily S Lau, Jonathan W Cunningham, Anthony Philippakis, Christopher D Anderson, Jennifer E Ho, Steven A Lubitz, et al. 2022. One clinician is all you need–cardiac magnetic resonance imaging measurement extraction: Deep learning algorithm development. JMIR Medical Informatics, 10(9):e38178.

Fabio Tamburini et al. 2020. How “BERTology” changed the state-of-the-art also for Italian NLP. In CEUR WORKSHOP PROCEEDINGS, volume 2769, pages 1–7. CEUR-WS.

Maryam Tayefi, Phuong Ngo, Taridzo Chomutare, Hercules Dalianis, Elisa Salvi, Andrius Budrionis, and Fred Godtliebsen. 2021. Challenges and opportunities beyond structured data in analysis of electronic health records. Wiley Interdisciplinary Reviews: Computational Statistics, 13(6):e1549.

Natalia Viani, Cristiana Larizza, Valentina Tibollo, Carlo Napolitano, Silvia G Priori, Riccardo Bellazzi, and Lucia Sacchi. 2018. Information extraction from Italian medical reports: An ontology-driven approach. International journal of medical informatics, 111:140–148.

Natalia Viani, Timothy A Miller, Carlo Napolitano, Silvia G Priori, Guergana K Savova, Riccardo Bellazzi, and Lucia Sacchi. 2019. Supervised methods to extract clinical events from cardiology reports in Italian. Journal of biomedical informatics, 95:103219.

Sameer Zaman, Camille Petri, Kavitha Vimalesvaran, James Howard, Anil Bharath, Darrel Francis, Nicholas S Peters, Graham D Cole, and Nick Linton. 2021. Automatic diagnosis labeling of cardiovascular MRI by using semisupervised natural language processing of text reports. Radiology: Artificial Intelligence, 4(1):e210085.

8. Language Resource References

Bernardo Magnini et al. 2020. E3C - European Clinical Case Corpus.