MODELLISTICA E CALCOLO SCIENTIFICO

MODELING AND SCIENTIFIC COMPUTING

MOX-Report No. 36/2024

A NLP-based semi-automatic identification system for delays in follow-up examinations: an Italian case study on clinical referrals

Torri, V.; Ercolanoni, M.; Bortolan, F.; Leoni, O.; Ieva, F.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it https://mox.polimi.it

A NLP-based semi-automatic identification system for delays in follow-up examinations: an Italian case study on clinical referrals

Vittorio Torri MSc[1*], Michele Ercolanoni MSc[2], Francesco Bortolan MSc[3], Olivia Leoni PhD[3], Francesca Ieva PhD[1,4]

[1*]MOX - Modelling and Scientific Computing Lab, Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, 20133, Italy.
[2]ARIA s.p.a - Azienda Regionale per l'Innovazione e gli Acquisti, Via Taramelli 26, Milan, 20124, Italy.
[3]U.O. Osservatorio Epidemiologico, DG Welfare, Regione Lombardia, Piazza Città di Lombardia 1, Milan, 20124, Italy.
[4]HDS - Health Data Science Centre, Human Technopole, Viale Rita Levi-Montalcini 1, Milan, 20157, Italy.

*Corresponding author(s). E-mail(s): vittorio.torri@polimi.it;
Contributing authors: michele.ercolanoni@ariaspa.it; francesco@bortolan.org; olivia_leoni@regione.lombardia.it; francesca.ieva@polimi.it;

Abstract

Background: This study aims to propose a semi-automatic method for monitoring the waiting times of follow-up examinations within the National Health System (NHS) in Italy, which is currently not possible to due the absence of the necessary structured information in the official databases.
Methods: A Natural Language Processing (NLP) based pipeline has been developed to extract the waiting time information from the text of referrals for follow-up examinations in the Lombardy Region. A manually annotated dataset of 10 000 referrals has been used to develop the pipeline and another manually annotated dataset of 10 000 referrals has been used to test its performance. Subsequently, the pipeline has been used to analyze all 12 million referrals prescribed in 2021 and performed by May 2022 in the Lombardy Region.

Results: The NLP-based pipeline exhibited high precision (0.999) and recall (0.973) in identifying waiting time information from referrals' texts, with high accuracy in normalization (0.948-0.998). The overall reporting of timing indications in referrals' texts for follow-up examinations was low (2%), showing notable variations across medical disciplines and types of prescribing physicians. Among the referrals reporting waiting times, 16% experienced delays (average delay = 19 days, standard deviation = 34 days), with significant differences observed across medical disciplines and geographical areas.
Conclusions: The use of NLP proved to be a valuable tool for assessing waiting times in follow-up examinations, which are particularly critical for the NHS due to the significant impact of chronic diseases, where follow-up exams are pivotal. Health authorities can exploit this tool to monitor the quality of NHS services and optimize resource allocation.

Keywords: Natural language processing, Referrals, Follow-up examinations, Public healthcare system, Quality of healthcare

1 Background

In Italy, the National Health System (NHS) plays a crucial role in healthcare, covering a significant portion of healthcare expenditure (75.6% in 2021 [1]) and ensuring universal coverage for citizens. Given the profound impact of healthcare services on citizens' quality of life and on the government budget, it is essential for public authorities to monitor the services provided by the NHS.
Chronic diseases exert a substantial burden on the health system, often requiring frequent follow-up examinations. The management of these examinations is often complex for both patients and General Practitioners (GPs) [2–4], but their inadequate provision can lead to increased rehospitalizations and worsened patient outcomes [5–14]. Moreover, several studies have examined the costs associated with chronic diseases, evidencing the cost-effectiveness of a timely provision of follow-up examinations [15–17].
The Italian NHS mandates a referral for follow-up examinations, that can be filled in by both GPs and specialized physicians. To guarantee appropriate waiting times, physicians are required to specify the priority of an examination in the referral. The standard referral form has four possible priority values (*U, B, D, P*), that correspond to 3, 10, 30-60 (depending on whether it is a specialist or instrumental examination) and 120 days of maximum waiting time, respectively [18]. While this priority mechanism has had an overall positive effect on the health care system [19], the standard priority classes inadequately capture disease-specific timing requirements for follow-up examinations. For this type of exams, physicians should select the default priority class (P) and manually specify the exact timing in the referral's free-text field. Consequently, assessing whether a specific follow-up examination has been provided in a timely manner becomes challenging, resulting in the exclusion of follow-up examinations from waiting time monitoring as per the Italian National Plan for the Management of Waiting Lists [18].

To address this limitation and enable waiting time monitoring, it is necessary to analyze the unstructured textual field that contain temporal information. Automating this process requires the utilization of Natural Language Processing (NLP) techniques. NLP is a subfield of artificial intelligence at the intersection of computer science, statistics and linguistics [20, 21], dealing with the representation and analysis of textual data. In the healthcare domain, a vast amount of unstructured textual data exists, particularly in Electronic Health Records (EHRs) clinical notes. These texts contain valuable information that is only partially represented in the structured EHR fields [22]. In the last decade, NLP has begun to find applications in this domain, despite facing challenges such as the specific lexicon, high ambiguity, and numerous context-specific abbreviations [23]. However, a key limitation of NLP is that most tools and models are developed for English texts, even though recent years have seen an increase in NLP applications to other languages [24].

In this paper, we present a tailored NLP-based pipeline for extracting temporal information from the textual fields of Italian referrals. This tool aims to facilitate the monitoring of waiting times by healthcare authorities.

Our pipeline was applied to a dataset of referrals from the Lombardy Region, Italy's most populous region, with 10 million inhabitants. Italy's NHS administration is largely decentralized, with the Ministry of Health issuing laws and regulations while the 19 regions, plus the 2 autonomous provinces of Trento and Bolzano, bear responsibility for organizing, administering, and evaluating NHS services in their respective territories [25]. Therefore, our work is expected to pave the way for the definition of regulations regarding the indication and monitoring of waiting times for follow-up examinations in the Lombardy Region.

The data sources used in this work are sourced from administrative databases, which contain routinely collected data for operational purposes in the Lombardy Region. Although administrative databases offer comprehensive coverage of the population and do not require additional data collection costs [26], they do pose challenges related to data quality, source integration, and the lack of specific research design [27, 28]. Despite these challenges, administrative data have been extensively used for statistical analysis in the healthcare domain [29–32].

To the best of our knowledge, this is the first work applying NLP techniques on Italian referrals, using open-source software (Venturelli et al [33] applied NLP techniques to verify prescription appropriateness on similar data, but with proprietary software). Additionally, our work is the first to address the issue of waiting times for follow-up examinations in Italy, providing an essential tool for monitoring this critical indicator of healthcare service quality.

The rest of this paper is organized as follows: in Section 2 we present the data and the methods adopted for our analysis, Section 3 shows the results, discussed, together with limitations and possible developments, in Section 4, while Section 5 contains conclusive remarks.

All analyses were conducted in Python [34] and the code is available by the authors upon request.

Table 1 Description of the dataset

Variable	Source	Description
Anonymized Doctor ID	(1)	Unique ID of the prescribing physician
Anonymized Referral ID	(1)	Unique ID of the referral
Date of referral	(1)	Date when the physician filled out the referral
Date of booking	(2)	Date when the patient booked the examination
First proposed date	(2)	First date proposed to the patient by the booking system
Accepted date	(2)	Date accepted by the patient to perform the examination (may be later than the first proposed date)
Type of examination	(1)	Type of specialist or instrumental examination requested
O/Z Flag	(2)	Flag indicating if the examination is subject to waiting time monitoring (O) or not (Z)
Healthcare facility	(2)	Hospital or outpatient clinic where the examination is performed
ATS (LHA - Local Health Authority)	(1)	LHA of the prescribing physician
Clinical question	(1)	Free text containing the reason for the referral and possibly its timing

List of variables included in the analyzed dataset, with source and description. Sources: (1) = Anonymized Electronic Referrals, (2) = Anonymized Performed Examinations

2 Methods

In this section, we present the data utilized in our analysis (Section 2.1) and describe the pipeline we developed (Section 2.2).

2.1 Data structure and contents

The dataset employed for this analysis consists of all referrals for specialists and instrumental examinations in Lombardy Region, prescribed in 2021 and performed up to May 2022.
The data have been obtained from two distinct sources:

1. *Anonymized Electronic Referrals*: this dataset includes information entered by the prescribing physician, such as the type of examination and the clinical question.
2. *Anonymyzed Performed Examinations*: this dataset contains data entered by healthcare facilities where the examinations are conducted. It includes information like the date of booking, the first date proposed to the patient, the actual examination date, and the healthcare facility details.

Both datasets contain records for individual prescribed examinations. However, since a referral can encompass multiple examinations (up to eight), some fields pertain to each individual examination (e.g., examination type, date performed), while others relate to the entire referral (e.g., referral date, physician's ID), and these shared values apply to all records within the same referral.
Table 1 summarizes the variables used in our analysis.

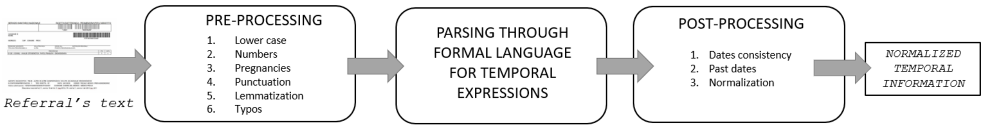

Fig. 1 Schema of the pipeline for the extraction of temporal information from textual fields of referrals

Since both physician and referral IDs have been replaced with anonymous identifiers, no sensitive information is present in the structured data. The free text field does not include sensitive information, since it is used only to report information about the reasons behind the referral and the time for follow-up examinations. In any case, to make sure that we do not fall under the scope of the GDPR (General Data Protection Regulation), a preliminary check of the presence of sensitive information (e.g., names, phone numbers, SSN identifiers) has been performed by Lombardy Region, removing them from texts, where present. The procedure is detailed in Additional File 1, Section D.

Exclusion criteria were applied to remove referrals not relevant for waiting time monitoring: screenings, emergency room (ER) visits, urgent referrals (labelled with priority class *U*, typically necessitating ER visits), and lab tests (which do not require booking in Lombardy).

Of the variables shown in Table 1, the most delicate is the O/Z flag. It has the value Z for all the types of referrals that are not subject to the monitoring of waiting times, which currently include follow-up examinations, but also a few other types of referrals, in particular psychiatry, sports medicine and dialysis, for which all the referrals, independently from being follow-up or first exams, should have the Z flag. While for specialist examinations there are two codes for each type of specialist examination, one corresponding to first access and one to follow-up, instrumental examinations have a unique code for each type of exam, so this flag is the best proxy to assess if an instrumental examination is a follow-up or not. Considering that our analysis of instrumental examinations will be focused on radiological exams, we do not expect to have a relevant number of them with the Z flag without being a follow-up examination. Both in our analysis of the full dataset and in selecting the data for the "training" and test set for the pipeline, we use follow-up codes for specialist examinations and the Z flag for radiology examinations.

Regarding the type of examination, the dataset encompasses 1304 different types of examinations. Among them, the 59 specialist examination types and the 130 radiology exam types (MRIs, CTs, X-rays, ultrasounds, and mammography scans) are the most relevant ones due to their inclusion in the priorities for waiting times set by the Lombardy Region [35].

The focus of our analysis is on the free text field (*Clinical Question*), which has a value in 99.9% of referrals. Table 2 provides examples of its values along with their English translations.

Table 2 Examples of clinical questions

Clinical question (ITA)	Clinical question (ENG)
- follow up in Ca endometroide	- follow-up in endometroid carcinoma
- CONTROLLO IN PREGRESSA POLIPECTOMIA	- MONITORING IN PREVIOUS POLYPECTOMY
- regolarizzazione 4 dito mano dx	- regularisation 4th finger right hand
- ca mammario in follow up; scan osseo 07 20 neg; dolore dorso-lombare dndd che si irradia a sinistra	- breast carcinoma in follow up; bone scan 07 20 neg; dorso-lumbar pain of nature to be determined radiating to the left
- Dermatite	- Dermatitis
- diabete mellito ottobre 2021	- diabetes mellitus October 2021
- sarcoma di ewing emibacino sx	- ewing's sarcoma left haemibacinus
- F.A. CRONICA	- CHRONIC ATRIAL FIBRILLATION
- 01282 - Altre forme di tubercolosi respiratoria, risultato dell'esame batteriologico o istologico non conosciuto (allo stato attuale	- 01282 - Other specified respiratory tuberculosis, bacteriological or histological examination unknown (at present
- per ter apia valutazione qt	- for ter hapy qt evaluation

Examples of clinical questions from the dataset. Italian abbreviations are expanded in the English translation.

2.2 Methods

The extraction of temporal information from text has been extensively studied in the literature [36], with rule-based and hybrid approaches being commonly used.
These approaches make extensive use of regular expressions, a powerful NLP technique that allows the definition of a formal language using specific symbols and patterns [37]. This language defines the set of strings that conform to the specified expressions. The main advantages of using regular expressions for information extraction is that they do not require a large annotated training set and they are a white-box system, overcoming the explainability challenges associated with many machine learning and deep learning models [38]. However, existing tools for extracting temporal information have limitations in terms of language and domain coverage. Most tools are initially developed for English and subsequently extended to other languages. Among these tools, HeidelTime [39] is the only one that currently covers Italian [40]. Another limitation is the domain-specific nature of temporal information, with expressions varying depending on the specific domain. In the medical domain, there are often abbreviations and particular temporal expressions that may not be adequately covered by general-purpose tools. Notably, the i2b2 challenge in 2012 was centered around identifying temporal relations in English clinical texts, with the winning team utilizing regular expressions [41]. While some tools exist for extracting temporal information from clinical texts [42], none of them currently cover the Italian language.
To address these limitations we developed a pipeline specifically tailored for extracting temporal information from Italian referral texts.
The pipeline proposed in this paper is depicted in Figure 1. It encompasses three key steps:

1. Pre-processing: the referral text undergoes a pre-processing step aimed at simplifying the text, enabling subsequent steps to be more effective.
2. Parsing: the pre-processed text is parsed using a formal language to extract temporal indications related to waiting times.
3. Post-processing: the extracted temporal indications are normalized, resulting in the production of Normalized Temporal Information (NTI). This allows for the computation of delays and facilitates further analysis.

In our pipeline, the core component is the parsing step, which employs a formal language defined through nested regular expressions to capture various types of temporal indications. We implement this step using the *reparse* Python library [43], facilitating the definition of a formal language using regular expressions.

This formal language for temporal expressions is developed based on a manually annotated dataset of 10 000 clinical questions randomly selected from follow-up examinations. The relatively large size of the annotated dataset is due to the low frequency of temporal indications observed during the annotation process.

Using these annotations, the formal language is developed and validated on another dataset of 10 000 manually annotated texts. The language incorporates rules for different types of temporal indications identified during the annotation process, as well as common elements shared among multiple types for temporal indications (e.g.: a *time unit* can be used in both an *interval of time* and a *precise time*). More details on the annotation procedure, the test set and the formal language can be found in Additional File 1, Section A, B and C, respectively.

In addition to parsing, our pipeline includes pre and post-processing steps to address specific cases and mitigate challenges such as misspelt words. Pre-processing includes lower-casing, transforming literal numbers into digits, removing confounding elements related to pregnancies, partial deletion of punctuation (excluding punctuation symbols used in dates), lemmatization using the Spacy Python library, and partial correction of typos. Stop words, often removed in NLP pre-processing [44], are retained in our pipeline as they provide essential context for understanding temporal relations. Lemmatization [45] helps normalize temporal-related terms, reducing the need for explicit handling of these cases in the formal language. More details on pre-processing can be found in Additional File 1, Section D.

Post-processing steps allow the completion of some missing information, the deletion of inconsistent values and the normalization of the extracted information. They are detailed in Additional File 1, Section E.

Once these steps are performed, we are able to quantify the delay, since it is possible to compare the extracted *NTI* with the first date proposed to the patient by the booking system.

When a patient needs to book an examination, the Lombardy Regional Booking System proposes the available slots in a selected province and he/she can select among them. For the computation of delays, we consider as starting time the time of booking, that is when the patient books the exam As the date on which the exam has been performed, we consider the first proposed date, i.e., the date of the first slot that was proposed to the patient when he/she booked the exam. The NHS cannot be

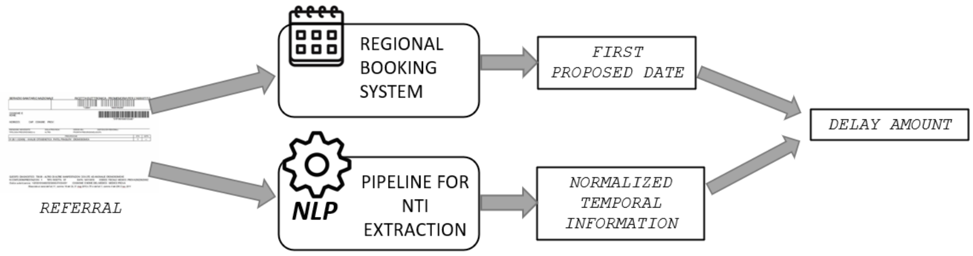

Fig. 2 Schema of the procedure for the computation of delays for follow-up examinations

Table 3 Results of the pipeline on the test set for the temporal information extraction

Metric	Value
Presence/Absence of the temporal indication	
Accuracy	0.999
Precision	0.990
Recall	0.973
Correctness of the *NTI*	
Accuracy on all referrals	0.998
Accuracy on referrals with any temporal indication	0.953
Accuracy on referrals with a follow-up temporal indication	0.948

Results of the pipeline on the test set, measuring different metrics with respect to both the binary classification problem of presence/absence of the temporal indication and to the accuracy of the extracted NTI

held responsible for delays resulting from late booking or choice of a later date by the citizen. Figure 2 summarizes the procedure for the computation of delays.

3 Results

In this section, we present the results of our analysis. In particular, in Section 3.1 we present the results for the pipeline on the manually annotated test set and in Section 3.2 we present the outcome of the analysis on the full 2021 dataset.

3.1 Pipeline for the NTI extraction

The pipeline's evaluation is conducted on two levels: the identification of temporal information presence/absence (a binary classification problem) and the correctness of the extracted Normalized Temporal Information (NTI). The accuracy of the NTI extraction is assessed on both the entire test set and the subset containing temporal information to avoid bias caused by the low frequency of temporal information. The extraction results are presented in Table 3.

Analyzing the errors in the test set of 10 000 examinations, we observe 8 (0.08%) false negatives (FN), 3 (0.03%) false positives (FP), and 7 (0.07%) indications with incorrect values. False positives are mainly caused by expression ambiguities and abbreviations, while false negatives result from uncorrected misspellings and unaccounted expressions such as “today”. The extraction of incorrect values is primarily due to misspellings and multiple temporal indications within a single clinical question. Additional File 1, Section F, provides a detailed analysis of false positives, false negatives, and wrong values.

3.2 Temporal information and delays on the entire dataset

At the end of this procedure, the dataset consists, for the year 2021, of 14 502 861 records, corresponding to 12 035 177 unique referrals. Within this dataset, there are 24 736 physicians, with 71% being specialized physicians and the remaining being GPs. Applying the pipeline to the entire dataset enables the insertion of new variables related to follow-up examinations: the value and unit of measure of the extracted NTI. Among all Z referrals, 1.18% contain temporal indications. This percentage increases to 2.02% for follow-up specialist examinations and 2.16% for radiological exams with the Z flag.

Analyzing the presence of temporal information based on the type of prescribing physician reveals that 0.43% of Z referrals from GPs have a temporal indication, while this percentage increases to 2.38% for referrals from specialized physicians.

Figure 3 shows the presence of temporal information, stratified by different types of specialist examinations, radiological exams, and the different ATS (LHA) of prescribing physicians. Notable findings include a high percentage number (9.5%) of orthopaedic examinations reporting temporal indications. Being orthopaedic examinations very common, these imply they have by large the highest absolute number of temporal indications (21 298). Other exams with high percentage numbers correspond to lower absolute numbers: vulnological (10.5% - 262), angiological (9.1% - 376), obstretical (8.6% - 6 196), plastic surgery (7.6% - 1 325) and oncohaematological (6.6% - 1 671).

Among the 102 000 temporal indications that are extracted, the average delay is -39 days, indicating that, on average, the examinations are performed in a timely manner. The median delay is -5 days, with an interquartile range of 43 days, suggesting the presence of outliers. Among the subset of examinations with identified temporal indications, 16% exhibit delays, with an average delay of 19 days (SD=34). The median delay is 6 days, with an interquartile range of 17 days.

Figure 4 provides a visual representation of the delayed follow-up referrals, stratified by type of specialist examination, type of radiological exam, and ATS (LHA). Additionally, boxplots illustrate the distribution of delays for the delayed exams.

In medical examinations, there are evident differences between various types of specialist examinations in terms of the number of delayed examinations and the extent of delays. The highest percentages of delayed examinations can be observed in diabetology (41%), oncohaemaotology (32%) and oncology (28%). Considering the entity of the delays, the highest delays are observed in nephrology (median 19 days), followed by diabetology (median 17 days) and pulmonology (median 16 days).

Regarding radiological follow-up exams, higher percentages of delays are observed in mammographies (17%), MRIs (16%), and X-rays (16%), with lower delays in MRIs compared to other radiological exams (median 2 vs 14-17 days).
Variations in delays are evident across different ATS (LHA). ATS 7 has the highest percentage of delayed follow-up examinations (36%), while ATS 11 exhibits the best performance.

4 Discussion

This paper aims to analyze texts of Italian referrals to investigate waiting times for follow-up examinations. To the best of our knowledge, this is the first attempt in this direction.
The proposed approach has evident strengths and potential. The developed pipeline has proved to be able to extract and normalize temporal information from these texts with high accuracy, on both the recognition of the presence of temporal information and the correctness of the extracted normalized temporal information. One of the significant advantages of our pipeline is its inherent explainability. Unlike many machine-learning and deep-learning-based models, the rationale behind the system's output is transparent and understandable. This attribute is crucial when a model is intended for use by health authorities. It is worth noting that this achievement was accomplished despite having used as a "training set" only a small percentage of the total available records. As we look forward, the availability of a larger annotated dataset will enable further fine-tuning of the system. While some extensions may be required to cover previously unseen types of temporal expressions, the pipeline's architecture, based on a formal language, facilitates the introduction of new rules or extensions to existing ones.
Executing the pipeline when the majority of referrals for follow-up examinations contain a temporal indication will allow precise monitoring of the quality of follow-up examinations, in terms of waiting times. Manual analysis of this magnitude would entail an unsustainable cost for the NHS, estimated at over 30 000 hours of work, assuming 10 seconds per referral.
Limitations should also be acknowledged. In particular, it is essential to remark that the results regarding delays, including variations among geographical areas and medical disciplines, presented here serve as examples of potentially valuable insights for decision-makers. However, these findings should be validated in subsequent analyses when a more significant proportion of referrals will include timing indications.
Another limitation is related to the fact that we considered as follow-up instrumental examinations all instrumental examinations with the Z flag since it is the best proxy to identify follow-up instrumental exams, but Z referrals may also include other types of referrals, and follow-up referrals might not always be marked as Z. This should not particularly affect the radiological exams we analyzed, but it might affect other types of instrumental examinations.

[1] Delays are calculated as $[NTI] -$ [time from date of booking to date proposed by the booking system to patient]

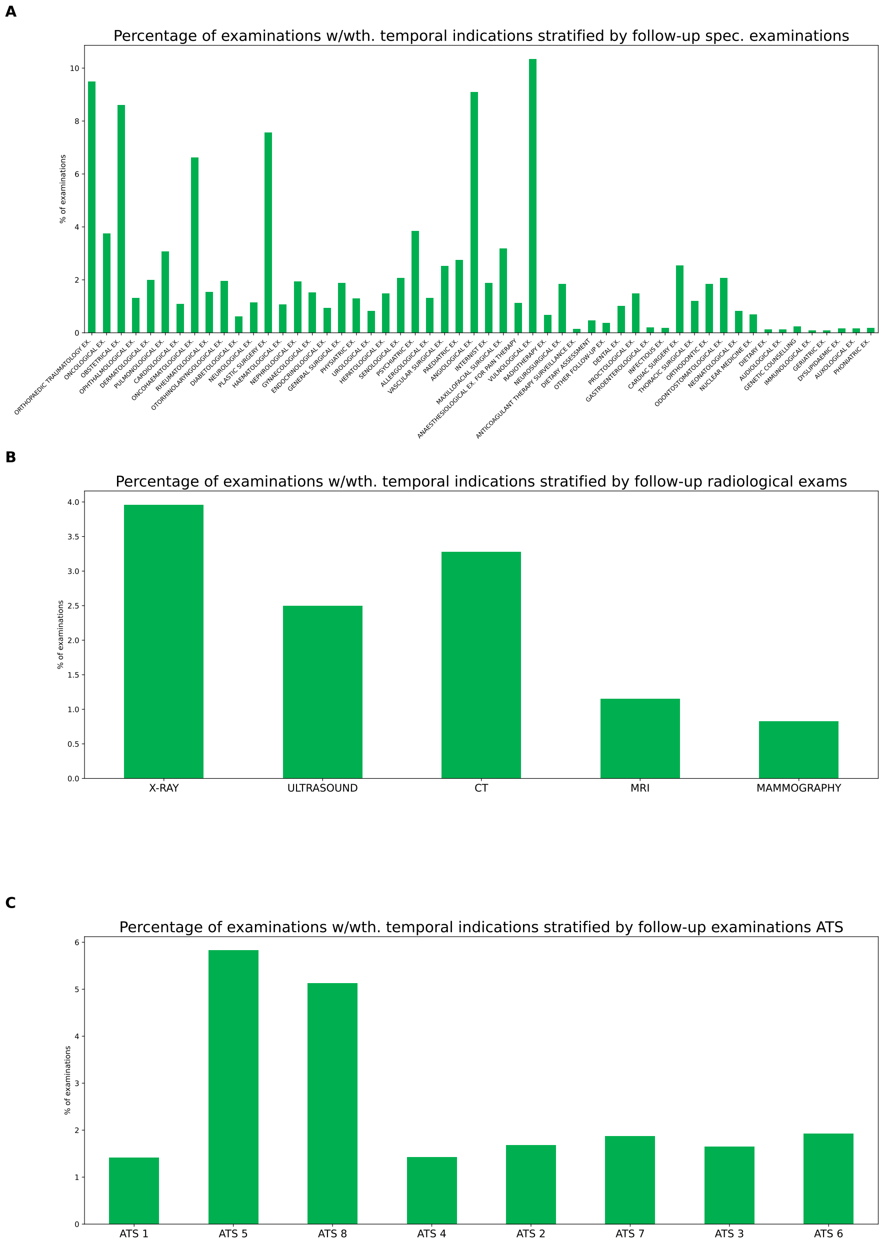

Fig. 3 Presence of the temporal indication for follow-up examinations by type of specialist examination, type of radiological exam and ATS (LHA). Only examinations with > 1000 records are included.

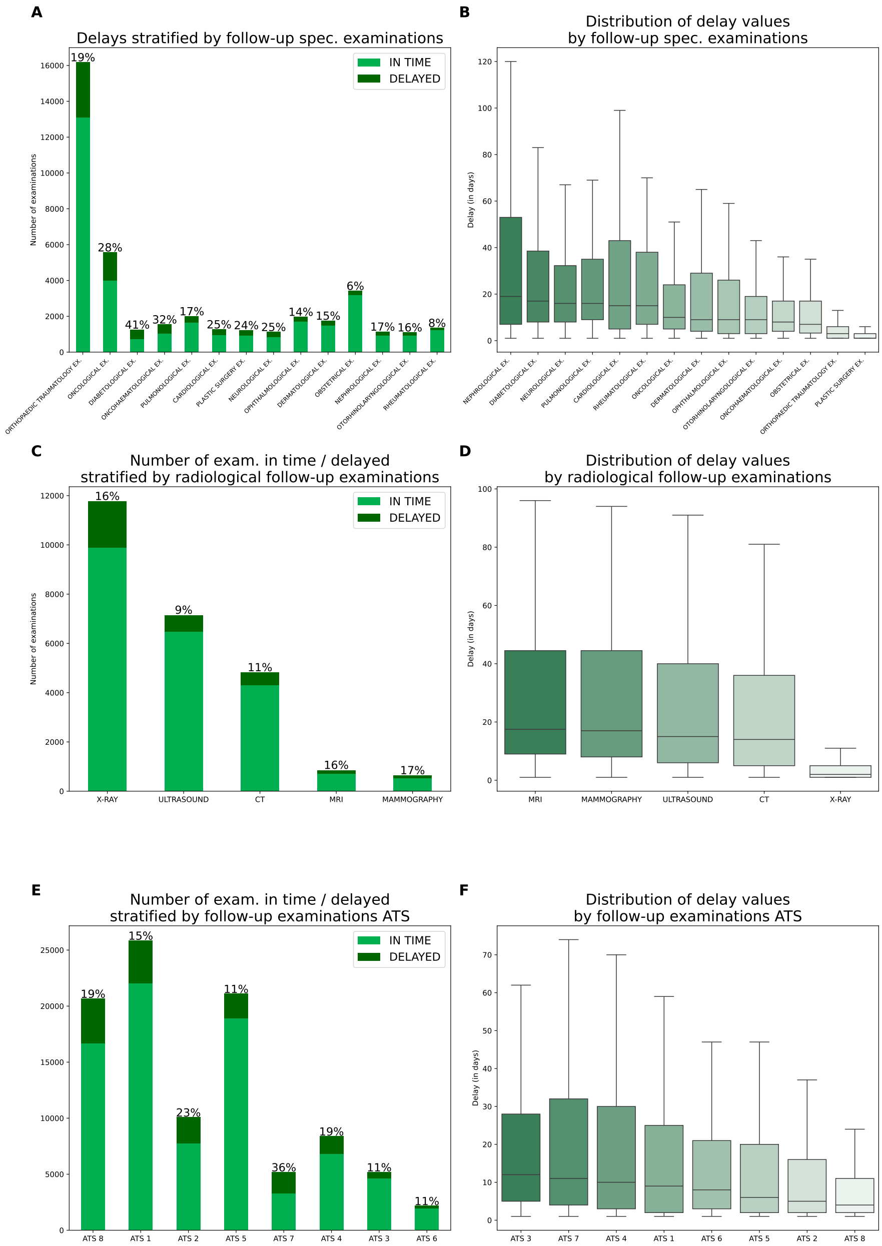

Fig. 4 Delays[1] for delayed exams and amount of delayed exams by type of examination, type of radiological exam and ATS (LHA). Boxplots are ordered by median, while barplots are ordered by the absolute number of delayed exams. Only examinations with > 1000 records are included.

Moreover, in the computation of delays in the provision of follow-up examinations, we considered the time between the first date proposed to the patient by the booking system and the date of booking. This implies excluding the impact of the following elements:

1. the time that might elapse between the request of a follow-up examination communicated by the specialist physician to the patient and the filled-out of the corresponding referral by the patient's GP, when the specialist physician does not directly fill it out
2. the time between the date of the referral and the date of booking
3. the geographical position of the healthcare facility where the first slot proposed to the patient is located

We do not have access to data which allow quantifying the impact of the first and the third point, and we did not consider the second point since the focus of this analysis is the monitoring of the delays for which the NHS can be considered responsible. We report in Additional File 1, Section G, plots describing the number and the amount of postponed follow-up examinations. Different types of examinations show differences in both the number of postponed exams (from 29% of diabetological exams to 4% of oncoematological exams) and the amount of postponement. Some types of exams, such as orthopaedics, have shorter postponements. This can be expected due to the higher urgency of certain diseases, which makes patients less likely to postpone their follow-up exams with respect to other types of diseases, such as diabetes and cardiological ones, that might remain more stable and less urgent for long periods of time. Relevant differences can also be observed among different ATSs: ATS 1 has 29% of postponed examinations, while they are $< 1\%$ in ATS 4. This is probably due to the fact that certain ATSs cover more extensive territories, with more hospitals where the first proposed date could be offered, potentially far from patients' residency. The numbers reported for the postponements increase the effective delay experienced by citizens, despite this is not taken into account by the quality metrics currently in use in the Italian NHS. Future analyses could better assess the reasons behind these postponements, which might lead health authorities to intervene both on the geographical factor and on patient awareness of the importance of in-time follow-up examinations for chronic diseases.

Among the possible extensions of our pipeline, we could enhance the recognition of typing mistakes, as they are prevalent in these texts. Additionally, we could consider incorporating the capability to associate temporal indications with specific examinations in cases where referrals contain multiple examinations with multiple temporal indications, all reported within the clinical question of the referral.

5 Conclusions

NLP and textual analysis represent promising tools for enabling semi-automatic performance monitoring in public administration. In particular, in this paper, we have placed our focus on the analysis of health referrals, in order to monitor the waiting times for follow-up examinations.

Follow-up examinations constitute a substantial portion of the services provided by the NHS, and their impact is particularly significant, especially when considering chronic diseases. Consequently, they can no longer be disregarded in the assessment of the quality of NHS services. Mandating physicians to include time indications in referrals for follow-up examinations would empower healthcare authorities to proactively allocate resources and intervene when necessary. This proactive approach would ensure the timely and effective delivery of care, supported by data that would be impractical to produce through manual analysis due to the extensive time it would require.
In summary, the application of NLP and textual analysis in monitoring follow-up examinations can contribute to reducing healthcare burdens, improving patient well-being, and optimizing the allocation of resources within the NHS.

Supplementary information.

- Name: Additional file 1
  Format: pdf
  Title: Supplementary Material
  Description: contains more details about the pipeline, the dataset and the error analysis

List of abbreviations.

1. NHS: National Health System
2. NLP: Natural Language Processing
3. GP: General Practitioner
4. EHR: Electronic Health Record
5. ER: Emergency Room
6. ATS/LHA: Agenzia di Tutela della Salute / Local Health Authority
7. GDPR: General Data Protection Regulation
8. SSN: Social Security Number
9. SD: Standard Deviation
10. IQR: Interquartile Range
11. MRI: Magnetic Risonance Imaging
12. CT: Computerized Tomography
13. NTI: Normalized Temporal Information
14. FP: False Positive
15. FN: False Negative

Declarations

Ethics approval and Consent to participate

The need for ethics approval is deemed unnecessary according to national regulations. In particular, according to the rules from the Italian Medicines Agency (available at: http://www.agenziafarmaco.gov.it/sites/default/files/det_20marzo2008.pdf), retrospective studies using administrative databases do not require Ethics Committee protocol approval.

The need for consent to participate is deemed unnecessary according to national regulations. In particular, see the rules of the Italian Privacy Authority available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9124510 and the European Union regulation n. 2016/679 (GDPR) Art. 1 c. 1 (https://gdpr.eu/article-1-subject-matter-and-objectives-overview/), Art. 4 c.1 (https://gdpr.eu/article-4-definitions/) and Recital 26 (https://gdpr-info.eu/recitals/no-26/).

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from Regione Lombardia but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Regione Lombardia.

Code availability

The code is available by the authors upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

No fundings supported this research.

Authors' contributions

VT: acquisition, analysis and interpretation of data, drafting of the manuscript ME: concept and design, acquisition, analysis and interpretation of data FB: concept and design, acquisition, analysis and interpretation of data OL: concept and design, acquisition, analysis and interpretation of data FI: concept and design, acquisition, analysis and interpretation of data, critical revision of the manuscript, supervision. All authors read and approved the final manuscript.

Acknowledgments

We thank the Welfare Directorate of Regione Lombardia, in the person of Dr. Giovanni Pavesi (Agreement between Politecnico di Milano and Regione Lombardia DGR n. XI/5301-27/09/2021). The present research is part of the activities of "Dipartimento di Eccellenza 2023-2027".

References

[1] OECD.: Health spending. Available from: https://www.oecd-ilibrary.org/content/data/8643de7e-en.

[2] Kersting C, Weltermann B. Electronic reminders to facilitate longitudinal care: a mixed-methods study in general practices. BMC Medical Informatics and Decision Making. 2016;16(1):1–9.

[3] Haq R, Heus L, Baker NA, Dastur D, Leung FH, Leung E, et al. Designing a multifaceted survivorship care plan to meet the information and communication needs of breast cancer patients and their family physicians: results of a qualitative pilot study. BMC medical informatics and decision making. 2013;13(1):1–13.

[4] Singh H, Wilson L, Petersen LA, Sawhney MK, Reis B, Espadas D, et al. Improving follow-up of abnormal cancer screens using electronic health records: trust but verify test result communication. BMC medical informatics and decision making. 2009;9:1–7.

[5] Anderson G, Horvath J. The growing burden of chronic disease in America. Public health reports. 2004;119(3):263–270.

[6] Nanthamongkolkul K, Hanprasertpong J. Longer waiting times for early stage cervical cancer patients undergoing radical hysterectomy are associated with diminished long-term overall survival. Journal of Gynecologic Oncology. 2015;26(4):262–269.

[7] van Harten MC, de Ridder M, Hamming-Vrieze O, Smeele LE, Balm AJ, van den Brekel MW. The association of treatment delay and prognosis in head and neck squamous cell carcinoma (HNSCC) patients in a Dutch comprehensive cancer center. Oral oncology. 2014;50(4):282–290.

[8] Kulkarni GS, Urbach DR, Austin PC, Fleshner NE, Laupacis A. Longer wait times increase overall mortality in patients with bladder cancer. The Journal of urology. 2009;182(4):1318–1324.

[9] Russell KL, Holloway TM, Brum M, Caruso V, Chessex C, Grace SL. Cardiac rehabilitation wait times: effect on enrollment. Journal of Cardiopulmonary Rehabilitation and Prevention. 2011;31(6):373–377.

[10] Ross H, Howlett J, Arnold JMO, Liu P, O'Neill B, Brophy J, et al. Treating the right patient at the right time: access to heart failure care. Canadian journal of Cardiology. 2006;22(9):749–754.

[11] Deslauriers S, Raymond MH, Laliberté M, Lavoie A, Desmeules F, Feldman DE, et al. Access to publicly funded outpatient physiotherapy services in Quebec: waiting lists and management strategies. Disability and rehabilitation.

2017;39(26):2648–2656.

[12] Desai AS, Stevenson LW. Rehospitalization for heart failure: predict or prevent? Circulation. 2012;126(4):501–506.

[13] Adib-Hajbaghery M, Maghaminejad F, Abbasi A. The role of continuous care in reducing readmission for patients with heart failure. Journal of caring sciences. 2013;2(4):255.

[14] Calman L, Beaver K, Hind D, Lorigan P, Roberts C, Lloyd-Jones M. Survival benefits from follow-up of patients with lung cancer: a systematic review and meta-analysis. Journal of Thoracic Oncology. 2011;6(12):1993–2004.

[15] Detournay B, Pribil C, Fournier Mt, Housset B, Huchon G, Huas D, et al. The SCOPE study: health-care consumption related to patients with chronic obstructive pulmonary disease in France. Value in health. 2004;7(2):168–174.

[16] Garattini L, Tediosi F, Chiaffarino F, Roggeri D, Parazzini F, Coscelli C, et al. The outpatient cost of diabetes care in Italian diabetes centers. Value in Health. 2001;4(3):251–257.

[17] Rea F, Ronco R, Martini N, Maggioni AP, Corrao G. Cost-effectiveness of posthospital management of acute coronary syndrome: A real-world investigation from Italy. Value in Health. 2022;25(2):185–193.

[18] Italian Ministry of Health.: National Plan for the Management of Waiting Lists 2019-2021. Available from: https://www.salute.gov.it/imgs/C_17_pubblicazioni_2824_allegato.pdf.

[19] Mariotti G, Siciliani L, Rebba V, Fellini R, Gentilini M, Benea G, et al. Waiting time prioritisation for specialist services in Italy: the homogeneous waiting time groups approach. Health Policy. 2014;117(1):54–63.

[20] Sharma AR, Kaushik P. Literature survey of statistical, deep and reinforcement learning in natural language processing. In: 2017 International Conference on Computing, Communication and Automation (ICCCA). IEEE; 2017. p. 350–354.

[21] Locatelli M, Seghezzi E, Pellegrini L, Tagliabue LC, Di Giuda GM. Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis. Buildings. 2021;11(12):583.

[22] Friedman C, Elhadad N. Natural language processing in health care and biomedicine. In: Biomedical informatics. Springer; 2014. p. 255–284.

[23] Iroju OG, Olaleke JO. A systematic review of natural language processing in healthcare. International Journal of Information Technology and Computer Science. 2015;8:44–50.

[24] Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical natural language processing in languages other than english: opportunities and challenges. Journal of biomedical semantics. 2018;9(1):1–13.

[25] Ferré F, de Belvis AG, Valerio L, Longhi S, Lazzari A, Fattore G, et al. Italy: health system review. Health systems in transition. 2014;.

[26] Timotfe D, Stoian AP, Hainarosie R, Diaconu C, Iliescu B, Balan G, et al. A review on the advantages and disadvantages of using administrative data in surgery outcome studies. Journal of Surgery. 2018;14(3):97–99.

[27] Hand DJ. Statistical challenges of administrative and transaction data. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2018;181(3):555–605.

[28] Mazzali C, Paganoni AM, Ieva F, Masella C, Maistrello M, Agostoni O, et al. Methodological issues on the use of administrative data in healthcare research: the case of heart failure hospitalizations in Lombardy region, 2000 to 2012. BMC health services research. 2016;16(1):1–10.

[29] Ieva F, Jackson CH, Sharples LD. Multi-State modelling of repeated hospitalisation and death in patients with Heart Failure: the use of large administrative databases in clinical epidemiology. Statistical methods in medical research. 2017;26(3):1350–1372.

[30] Ieva F, Paganoni AM, Secchi P. Mining administrative health databases for epidemiological purposes: a case study on acute myocardial infarctions diagnoses. In: Advances in Theoretical and Applied Statistics. Springer; 2013. p. 417–426.

[31] Corrao G, Rea F, Di Martino M, De Palma R, Scondotto S, Fusco D, et al. Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy. BMJ open. 2017;7(12):e019503.

[32] Corrao G, Mancia G. Generating evidence from computerized healthcare utilization databases. Hypertension. 2015;65(3):490–498.

[33] Venturelli F, Ottone M, Pignatti F, Bellocchio E, Pinotti M, Besutti G, et al. Using text analysis software to identify determinants of inappropriate clinical question reporting and diagnostic procedure referrals in Reggio Emilia, Italy. BMC health services research. 2021;21(1):1–13.

[34] Foundation PS.: Python Language Reference, version 3.7. Available from: http://www.python.org.

[35] Lombardy Region.: DGR XI / 6002 - Determinations regarding the 2022 waiting list plan. Available from: https://areadocumentale.servizirl.it/atti/download/AAAAWOWkQM9/PrCeYPCLXUJpI0ksRK0WkbCi0vQJyuZ6OQ3AwaydPHT

ZjqwEJQCqduNbQppW5ZxjQD5oppBnrsdA8AMxlsNxjB0lRAVa5vTSHu0XpmwLIWit/ioAAACAQpuf6d4a4TuH9BvJxFayYj9DAEvGHN2iz5HeuzeGH+HuH0qKVix6ktJPMulHpfT4pvOa2DZmanGFFEp6xdWROjFywLK/u3kbajD5Stm+9lZ5PXmfvvhrJCuk4+ePSS1ojpTjAidsmIQlhvYg4AlL3+8jQh2Q4GbPk4zoIGDCobYAAAAIhHxEx2uaIvk=.

[36] Lim CG, Jeong YS, Choi HJ. Survey of temporal information extraction. Journal of Information Processing Systems. 2019;15(4):931–956.

[37] Yu S. Regular languages. In: Handbook of formal languages. Springer; 1997. p. 41–110.

[38] Ho TK, Luo YF, Guido RC. Explainability of Methods for Critical Information Extraction From Clinical Documents: A survey of representative works. IEEE Signal Processing Magazine. 2022;39(4):96–106.

[39] Strötgen J, Gertz M. Heideltime: High quality rule-based extraction and normalization of temporal expressions. In: Proceedings of the 5th international workshop on semantic evaluation; 2010. p. 321–324.

[40] Manfredi G, Strötgen J, Zell J, Gertz M. HeidelTime at EVENTI: Tuning Italian resources and addressing TimeML's empty tags. HeidelTime at EVENTI: Tuning Italian Resources and Addressing TimeML's Empty Tags. 2014;p. 39–43.

[41] Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. Journal of the American Medical Informatics Association. 2013;20(5):806–813.

[42] Hamon T, Grabar N. Tuning HeidelTime for identifying time expressions in clinical texts in English and French. In: Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi); 2014. p. 101–105.

[43] Chase A.: Reparse Python library. Available from: https://github.com/andychase/reparse.

[44] Kannan S, Gurusamy V, Vijayarani S, Ilamathi J, Nithya M, Kannan S, et al. Preprocessing techniques for text mining. International Journal of Computer Science & Communication Networks. 2014;5(1):7–16.

[45] Balakrishnan V, Lloyd-Yemoh E. Stemming and lemmatization: A comparison of retrieval performances. Lecture Notes on Software Engineering. 2014;.

Appendix A Annotation procedure

The manual annotation process for both the “training” and the test set has been performed by initially assigning to each text a binary indication of the presence/absence of the temporal indication. Subsequently, all the texts in which a temporal indication has been recognized have been labelled with the following elements:

1. a binary indication for the temporal information to be referred to the date of the next follow-up examination;
2. the portion of the string containing the temporal indication;
3. a normalized value of the temporal indication (dates in the format *dd-mm-yyyy*, other indications as numeric values);
4. the unit of measure of the normalized value (e.g., date, month, week);
5. the type of temporal indication (e.g., *Da* (Since), *Tra* (In), *Data completa* (Full date));

This process has been repeated three times and discrepancies have been manually analyzed to reduce the probability of errors in the annotations.

Appendix B Test dataset

The manually annotated test set consists of 10 000 clinical questions selected randomly among follow-up examinations. Details on the annotations are reported in Table B1.

Table B1 Types of temporal information in the manually annotated test set

Type of indication	N. of records
Follow-up timings	**291**
Tra (In)	134
A (in) + amount of time or month	32
Per (for) + month	5
Alla x settimana (at the x-th week)	2
Dopo (After) + amount of time	1
Entro (within) + amount of time	6
Time indication (e.g., *2 months*, *October*)	30
Precise date	81
Other timings	**27**
Past exams	16
Week of pregnancy	5
Pregnancy termination date	1
Date of last period (for pregnancy)	1
Date of origin of symptoms/disease	4

Analysis of the different types of temporal indications manually annotated in the test set, including temporal indications not referred to waiting times

Appendix C Formal language for temporal information

The formal language we developed to extract temporal information is reported here, using the notation of formal grammars.

```
TEMPORAL_INDICATION →CONTROL_INDICATION | DA | FA

CONTROL_INDICATION →FUTURE_DISTANCE | IL | FULL_DATE |
    ↪ IMPRECISE_FUTURE

SPACES →\s SPACES | ϵ

FUTURE_DISTANCE →INTERVAL_FUTURE_DISTANCE |
    ↪ PRECISE_FUTURE_DISTANCE

INTERVAL_FUTURE_DISTANCE →NUMBER SPACES SEPARATOR SPACES NUMBER
    ↪ SPACES TIME_UNIT

NUMBER →[0-9]+

SEPARATOR →- | /

TIME_UNIT →DAY_UNIT | WEEK_UNIT | MONTH_UNIT | YEAR_UNIT

DAY_UNIT →gg | giorno

WEEK_UNIT →sett | settimana

MONTH_UNIT →mese

YEAR_UNIT →anno

PRECISE_FUTURE_DISTANCE →NUMBER SPACES TIME_UNIT

IL →il SPACES (FULL_DATE | DAY_MONTH_DATE | MONTH_YEAR_DATE | DAY)

FULL_DATE →DAY DATE_SEPARATOR MONTH DATE_SEPARATOR YEAR

DAY →([0-3][0-9]) | [0-9]

DATE_SEPARATOR →- | / | . | +

MONTH →NUMERICAL_MONTH | LITERAL_MONTH

NUMERICAL_MONTH →(1[0-2]|0[1-9]|[1-9])
```

```
LITERAL_MONTH →SHORT_LITERAL_MONTH | FULL_LITERAL_MONTH

SHORT_LITERAL_MONTH →gen|feb|mar|apr|mag|giu|lug|ago|set(?!timana)
    ↪ |sett(?!imana)|ott|nov|dic

FULL_LITERAL_MONTH →gennaio|febbraio|marzo|aprile|maggio|giugno|
    ↪ luglio|agosto|settembre|ottobre|novembre|dicembre

YEAR →FULL_YEAR | SHORT_YEAR

FULL_YEAR →20[0-9][0-9]

SHORT_YEAR →[0-9][0-9]

DAY_MONTH_DATE →DAY DATE_SEPARATOR MONTH

MONTH_YEAR_DATE →MONTH DATE_SEPARATOR YEAR

IMPRECISE_FUTURE →DAY_MONTH_DATE | MONTH_YEAR_DATE | LITERAL_MONTH

DA →(da|dal) SPACES FUTURE_DISTANCE

FA →FUTURE_DISTANCE SPACES fa
```

Appendix D Pre-processing steps

To ensure the absence of sensitive information of both doctor and patients in the free text fields, texts have been pre-filtered with the following regular expressions:

```
# A name (from predefined list of names) and surname (assuming the
    ↪ word after the name is a surname)
'\b(' + r'|'.join(names) + r')\b\s*[a-z]+\b\s*'

# A title like doctor, professor, followed by a surname (assuming the
    ↪ word after the title is a surname)
'\b(dott|dottor|dr|dot|dottoressa|dottssa|dssa|dotssa|drssa|sig|signor
    ↪ |signora|sig\.?ra|sig\.?na|signorina|prof|prof\.?ssa|professor|
    ↪ professoressa)\.?\b\s*[a-z]+\b\s*'

# An Italian SSN (Codice Fiscale)
'\b[a-z]{6}\d{2}[a-z]\d{2}[a-z]\d{3}[a-z]\b'

#Italian phone numbers
```

```
'\b((3\d{2}[\s|\-|\.|\/]?\d{7})|(0\d[\s|\-|\.|\/]?\d{8}) | (0\d{2}[\s
    ↪ |\-|\.|\/]?\d{7})|(0\d{3}[\s|\-|\.|\/]?\d{6}))\b'
```

Before running our parser we then performed the following operation:

- replace all numbers expressed in words with their corresponding digits
- remove all expressions of the type `\d+ \s* settimana` (*n-th week*) in sentences where a pregnancy is mentioned
- remove all dates that immediately follow the word `gravidanza` (*pregnancy*) or the expression `termine il` (*termination on*)
- for punctuation removal, remove all punctuation symbols present in `string.punctuation` from the Python library `string`, except for the following characters: `./()-`, since they are commonly used in dates.
- to correct some typos, replace all the words with less than 5 occurrences in the dataset that have a similar word with more than 5 occurrences. The similarity is computed with the `SequenceMatcher.ratio(s1,s2)` function, which corresponds to $2M/T$, where M is the number of matching characters between $s1$ and $s2$ and T is the total number of characters in $s1$ and $s2$. The threshold is set to 0.8.
- to simplify the grammar, lemmatize words using the *Spacy* Python library, with the model *it_core_news_sm*

Appendix E Post-processing steps

The *reparse* library allows defining functions to be executed during the grammar parsing, i.e. during the process in which a string is analyzed to understand if it belongs or not to the language defined by the grammar. These functions act as post-processing functions (corresponding to the third block of Figure 1 of the main manuscript), and we defined them so as to normalize the representation of the extracted temporal information, making it possible to use this information to compute delays. We consider two possible types of normalized formats: dates and amounts of time. Dates are all transformed in the standard format *dd-mm-yyy*, while amounts of time have a numeric value and a unit of measure (*days, weeks, months* or *years*).

The post-processing also completes missing information, considering the referral date. The rules used to complete missing information are reported in Algorithm 1. It should be noted that when a type of temporal indication is part of another (e.g., `LITERAL_MONTH` in `FULL_DATE`) the completed information can be overwritten by the container information (e.g., the year will be the one reported in the `FULL_DATE`, not the one completed in the `LITERAL_MONTH`). At this point, we also discard all the dates that are not consistent (e.g., *31.06*, since the day *31* is not present in the month *06*). Doing this only at post-processing allows to avoid an explicit specification of all the particular date cases in the grammar.

Eventually, all dates that are prior to the referral date are discarded.

Algorithm 1 Rules to complete missing information in extracted time indication

```
ref_month ← referral.referral_date.month
ref_day ← referral.referral_date.day
ref_year ← referral.referral_date.year
time_ind ← extract_time_indication(referral.clinical_question)

if type(time_ind)==MONTH_YEAR_DATE or type(time_ind)==LITERAL_MONTH then
    time_ind.day ← get_last_day_of_month(time_ind.month)
end if

if type(time_ind)==IL and isna(time_ind.month) then
    if time_ind.day ≤ ref_day then
        time_ind.month ← (ref_month + 1)%12
    else
        time_ind.month ← ref_month
    end if
end if

if type(time_ind)==DAY_MONTH_DATE or type(time_ind)==LITERAL_MONTH or (type(time_ind) ==
IL and isna(time_ind.year)) then
    if time_ind.month ≤ ref_month or (time_ind.month == ref_month and
time.day ≤ ref_day) then
        time_ind.year ← ref_year + 1
    else
        time_ind.year ← ref_year
    end if
end if
```

Appendix F Errors analysis

In this section we perform a detail reporting of the errors registered on the test set by the NTI extraction pipeline.

The false positives are the following values:

1. *recidiva di panniculite (dopo un mese di augmentin) - recurrence of panniculitis (after one month of augmentin)*: probably referred to a previously prescribed medication and not to the fact the next exam should be done after one month of augmentin
2. *frattura biossea 1/4 distale avambraccio destro - right forearm distal 1/4 bony fracture*: *1/4* is not a date here
3. *allergia a sedano, dic in rivalutazione - allergy to celery, irritative contact dermatitis under reevaluation*: here the problem is in the acronym *dic*, which is often used as an abbreviation of the month December (*Dicembre*) in Italian, but in this case, it stands for *dermatite irritativa da contatto (irritative contact dermatitis)*

The first one is the most difficult to correct since the sentence has a degree of ambiguity and could be even interpreted in the sense of waiting one month before doing the next examination. The second one is referred to orthopaedics and may suggest inserting specific rules for these cases. The last one is determined by the usual problem of ambiguity in abbreviations. It is difficult to avoid since even in the field of allergology *dic* could be used to indicate *december*.

The false negatives are the following ones:

1. *controllo in tx fegato < 1 annoMGEINF-N6prelievo in ambulatorio infermieristico - control in liver tx < 1 yearMGEINF-N6drawal in nursing clinic*: the problem is that the word *anno (year)* is not separated from the following one and consequently it is not recognized
2. *fr 3 mtc sx tr auna settimana ca - fr 3 mtc sx i none week ca*: the problem is that the word *tra (in)* has one letter attached to the next word *una (one)*, so the latter is not recognized
3. *ipotiroidismo cod 191 - tra 2 mesi- - hypothyroidism cod 191 - in 2 months-*: the problem is related to the dash attached to *mesi (months)*, which causes an incorrect lemmatization
4. *retinopatia diabetica, AMB.CHIRURGIA COMPLESSA PER OGGI - diabetic retinopathy, COMPLETE SURGERY AMB.SURGERY FOR TODAY*: we did not consider *today* among the time indications, since it is anomalous that a follow-up examination is requested for today
5. *k mammella operato, pregresso LH, controllo tra circa 6-7mesi - k breast operated, previous LH, check in about 6-7months*: the absence of a space in *6-7mesi (6-7months)* causes a wrong lemmatization
6. *sospetta DAC - controllo per lettura patch dopo circa 48 ore - suspected DAC - check for patch reading after about 48 hours*: we did not include temporal indication expressed as hours, since it seemed to be anomalous for follow-up examinations, but this suggests its inclusion
7. *controllo il gg11 Lombalgia - check on day11 Lumbago*: the problem is due to the absence of a space between gg and 11
8. *ciclo di 8 ctr in D Umore c/o CTDD (95B002) dal 2022 - cycle of 8 ctr in D Mood c/o CTDD (95B002) from 2022*: here the indication is related to the starting time of the follow-up examinations, and not to the time within which they should be performed, so it is not really a false negative

While it is possible to fix the problems related to dashes attached to words, which cause incorrect lemmatization, it is harder to detect missing spaces.

The texts from which the pipeline extracted a wrong values are the following ones:

1. *reflusso acido - cerume periodico / per Luglio-Agost - acid reflux - periodic cerumen / for July-August*: the pipeline returns 31/07/2021 instead of 31/08/2021, due to a missing letter in *Agosto (August)*
2. *Controllo ad 1 mese e gispotomia in frattura biossea avambraccio dx (chirurgia il 15/3/21) - 1-month check-up and gispotomy in right forearm bony fracture (surgery on 15/3/21)*: the pipeline returns 1 month, which is the most correct value. The labelling included also 15/3/21, since it is still a date of a future examination

3. *ulcera gamba dx-visita fine agosto e fine settembre 2020 - ulcer right leg-visit end of August and end of September 2020*: the pipelines returns 31/08/2021, which is wrong, since both September and August probably refers to 2020, even if it is not completely unambiguous
4. *frattutra base V MTT piede sinistro 7 giorni fa. In sala gessi 3 set - fracture base V MTT left foot 7 days ago. In plaster room 3 set*: the pipeline returns 3/09/2021, but the label is 3 weeks. The text is in effect ambiguos, since *set* is an abbreviation for both September and week, in Italian.
5. *controllo dopo rimozione di collare ecrvicale in trauma distorsivo rachide cervicale: visita da programmare per il 24-25 di giugno - check after removal of ecrvical collar in cervical spine distortion injury: visit to be scheduled for 24-25 June*: the pipeline returns 30/06/2021 instead 26/06/2021, due to the double number 24-25 which is not recognized
6. *k sene V8901 per dicembre 21- genaioo 2022 - k breast V8901 for December 21-January 2022*: the pipeline returns 31/12/2021 instead of 31/01/2022, due to the mispelling of *gennaio (January)*
7. *fratt tibia distale destra - medicazione giovedi 25/11 - RX e visita a 30 gg - right distal tibia fracture - dressing Thursday 25/11 - X-ray and examination at 30 days*: the pipeline returns 25/11/2021, which is correct, but there is also the examination at 30 days

Some of these wrong values can be fixed, like the errors due to multiple days separated by dashes. The most complex situation is when there is more than one indication since it is not enough to recognize them all, but it should be understood which is the correct one for the examination that is being considered.

Appendix G Postponments

Figure G1 described the number of postponed examinations and the amount of postponments by type of examination and ATS.

Appendix H Number of examinations

Figure H2 described the number of examinations by type of specialist examinations, radiological examinations and ATS.

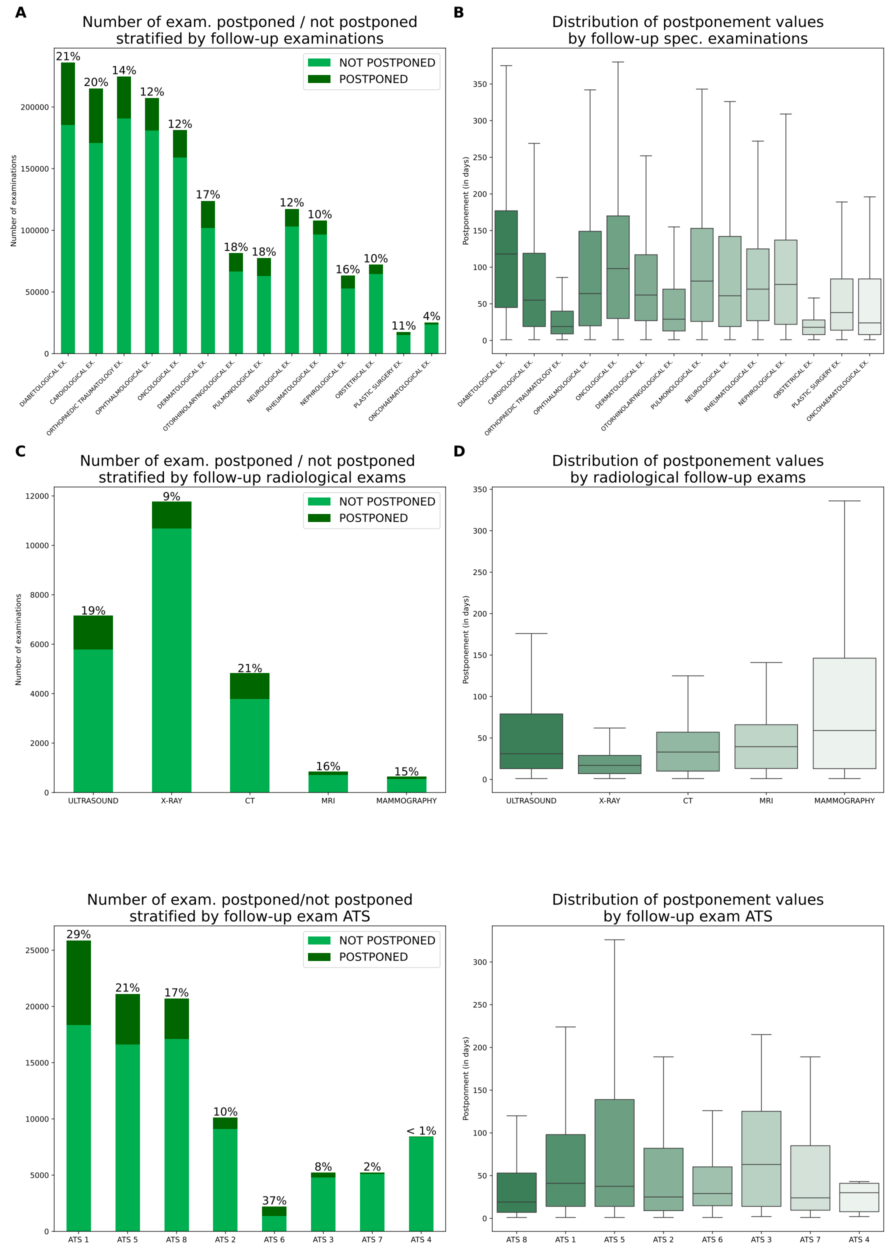

Fig. G1 Postponments of follow-up examinations by type of specialist examination, type of radiological exam and ATS

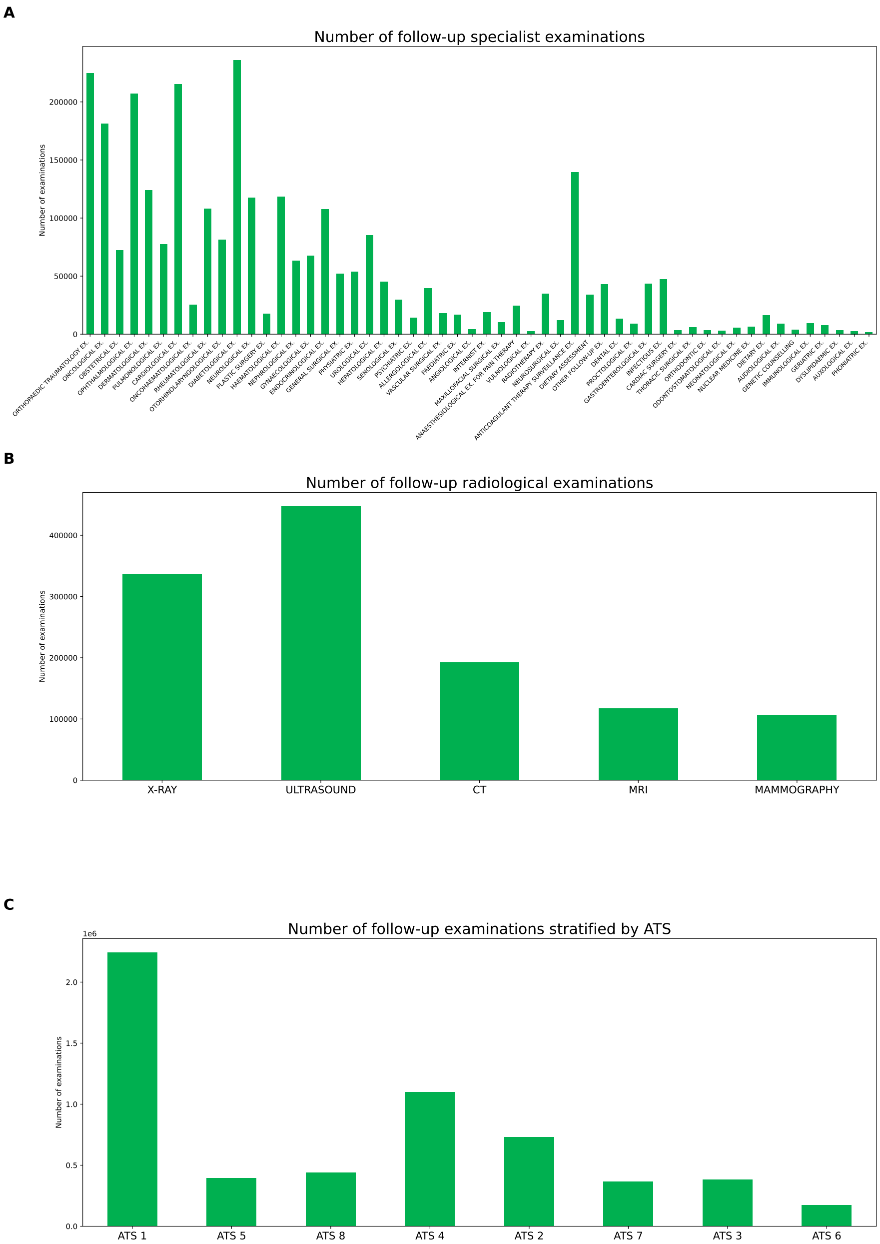

Fig. H2 Number of follow-up examinations by type of specialist examination, type of radiological exam and ATS

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

Corti, M.
Exploring tau protein and amyloid-beta propagation: a sensitivity analysis of mathematical models based on biological data

34/2024 Corti, M.
Exploring tau protein and amyloid-beta propagation: a sensitivity analysis of mathematical models based on biological data

35/2024 Botti, L.; Botti, M.; Di Pietro, D.A.; Massa; F.C.
Stability, convergence, and pressure-robustness of numerical schemes for incompressible flows with hybrid velocity and pressure

32/2024 Ziarelli, G.; Parolini, N.; Verani, M.
Learning epidemic trajectories through Kernel Operator Learning: from modelling to optimal control

31/2024 Gambarini, M.; Agate, G.; Ciaramella, G.; Miglio, E.; Maran, S.
Modeling and optimization for arrays of water turbine OWC devices

30/2024 Gregorio, C.; Rea, F.; Ieva, F.; Scagnetto, A.; Indennidade, C.; Cappelletto, C.; Di Lenarda, A.; Barbati, G.
Flexible approaches based on multi-state models and microsimulation to perform real-world cost-effectiveness analyses: an application to PCSK9-inhibitors

29/2024 Palummo, A.;, Arnone, E.; Formaggia, L.; Sangalli, L.M.
Functional principal component analysis for incomplete space-time data

28/2024 Magri, M.; Riccobelli, D.
Modelling of initially stressed solids: structure of the energy density in the incompressible limit

27/2024 Antonietti, P.F.; Beirao da Veiga, L.; Botti, M.; Vacca, G.; Verani, M.
A Virtual Element method for non-Newtonian fluid flows

26/2024 Bucelli, M.; Regazzoni, F.; Dede', L.; Quarteroni, A.
Robust radial basis function interpolation based on geodesic distance for the numerical coupling of multiphysics problems