Structured radiology report text analysis using natural language processing for automatic billing

Conference Paper · March 2024

DOI: 10.1117/12.3005518

10 authors, including:

Natalie Heracleous
Groupe 3R
569 PUBLICATIONS **39,800** CITATIONS
SEE PROFILE

Cyril Thouly
Réseau radiologique romand, suisse
17 PUBLICATIONS **18** CITATIONS
SEE PROFILE

Henning Müller
HES-SO Valais-Wallis
964 PUBLICATIONS **23,374** CITATIONS
SEE PROFILE

Federica Zanca
General Electric
12 PUBLICATIONS **76** CITATIONS
SEE PROFILE

Structured radiology report text analysis using Natural Language Processing for automatic billing

Nicolas Devanthery[a], Natalie Heracleous[b] ,Benoit Dufour[b], Jean-Daniel Fardel[b], Benoit Rizk[b], Hugues Brat[b], Cyril Thouly[b], Henning Müller[a], Federica Zanca[b]

[a]University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
[b]Groupe 3R, R&D Department, Rue de la Dixence 8,1950 Sion, Switzerland

ABSTRACT

Purpose: The aim of this study was to develop an algorithm for automated quality control of structured radiology reports and to automatically obtain the correct invoicing codes for the performed exam. Ultrasound (US) exams of the abdomen were selected as use case, including Doppler exams.
Method: To build a correct algorithm for automatic billing, the billing tree for the Ultrasound exams was studied. In Switzerland, TARMED, is the tariff structure used for billing outpatient medical services. The 4600 services listed in TARMED are divided into chapters that group together all services with a well-defined common characteristic. For example, Chapter 39 covers all medical imaging services. These chapters are further subdivided into subchapters for greater precision.
Using this information a modular Natural Language Processing algorithm based on the Natural Language Toolkit (NLTK) library was developed. A second NLP algorithm based on SPACY was also developed, with the objective of a double validation of the first developed NLP algorithm.
To train and test the algorithm a dataset of 170 exams corresponding to US abdominal examinations along with their radiology report were extracted from our RIS.
The results of the algorithm were validated by an experienced technologists which identified possible discrepancy between the algorithm results and the correct billing. This check was carried out on a batch of data containing 95 samples. A confusion matrix was used to analyze the results.
Results: In all 95 data samples, the NLTK algorithm was able to detect the billing codes correctly 100% of the time. In all our 95 data samples, the Spacy algorithm was able to detect the billing codes correctly in 86.3% of cases. This algorithm tends to overestimate the type of abdominal examination present in the report. Indeed, the 13 cases in which the algorithm made an error were cases where it detected a full abdominal ultrasound when the examination was a simple lower or upper abdomen.
Conclusion: The NLTK model provides reliable and efficient estimation of billing codes for abdominal ultrasound, facilitating the task of the technologies who saves time and avoids possible human errors.

1. DESCRIPTION OF THE PURPOSE

A very large number of radiology reports are produced for the radiological examinations carried out at the Groupe 3R each day. Quality control of these reports is therefore essential to ensure that billing is carried out correctly and accurately.
In Switzerland, TARMED, short for Tarif Médical, is the tariff structure used for billing outpatient medical services. It provides a totally uniform structure throughout Switzerland and serves as the basis for billing medical services (Office Fédéral De La Santé Publique - OFSP, 2017).
The 4600 services listed in TARMED are divided into chapters that group together all services with a well-defined common characteristic. For example, Chapter 01 covers all basic services, Chapter 39 covers all medical imaging services. These chapters are further subdivided into subchapters for greater precision. Taking chapter 39 as an example, it is subdivided as follows (Figure 1):

39 Medical Imaging
- 39.01 Medical imaging: generalities
- 39.02 Radiology
- 39.03 Ultrasound, doppler and interventions
- 39.04 CT
- 39.05 MRI
- 39.06 Angiography

The services listed in TARMED all have a unique number, starting with the chapter number to which they belong. There are two types of services which can be invoiced: basic services, which can be billed on their own and additional services that are linked to a basic service and must therefore be billed together with the corresponding basic service.
For the invoicing, TARMED works on the basis of tariff points. Each ambulatory service corresponds to a defined number of points. Tariff points are separated into two categories (Office Fédéral De La Santé Publique - OFSP, 2017): points for medical services that remunerate medical staff for their work with patients; points for technical services that compensate the infrastructure required for the service and also remunerate all non-medical staff.
The total number of points in these two categories defines the total number of points for a service. The total points are then multiplied by the point price to give the price of the service. The point value is negotiated at the cantonal level by all partners and approved by the authorities. For example, the point price in Valais in 2022 is set at 89 centimes, while in Geneva it is 97 centimes. (Office Fédéral De La Santé Publique - OFSP, 2017).
In order to select the correct type of services provided with the exam, the technologist carries out a manual quality control of each exam. Specifically, the TARMED codes are prefilled in our RIS (EDL) after the exam is performed and need to be validated by the technologist or modified, if necessary. The codes are grouped based on the service provided and the material or product used. To verify their correctness the radiographer first opens the radiology report to verify the type of service provided, distinguishing between base service and additional service, in function of the protocol used for the exam. Subsequently the images are opened to confirm that the protocol mentioned in the radiology report is exactly what was implemented and to verify the presence or not of contrast and therefore identify the right code for the material and products used. This process is time consuming and error prone, with as a consequence possible economic loss.
The purpose of this study was therefore to develop an algorithm for automated quality control of structured radiology reports and to automatically obtain the correct invoicing codes for the performed exam. Ultrasound (US) exams of the abdomen were selected as use case, including Doppler exams.

2. METHOD

2.1 Decision tree for billing

To build a correct algorithm for automatic billing, the current billing tree for the Ultrasound exams were studied.

Detection of examination type
All US examinations have a common billing basis, consisting of two TARMED positions:

- 39.0010, which is a basic consultation in a radiological institute. Present on all examinations.
- 39.3800, which is the basic technical service for an ultrasound examination.

For a specific abdominal US, several optional billings are available, depending on the exam performed:
- Position 39.3510 Vessel sonography: should be added only if the exam was performed with Doppler. This term should also be included in the radiology report. If this term is not included in the report, even if the examination was performed, it is not billable.
- Position 39.3265 Gastrointestinal tract: should be added if the section of the radiology report concerning the gastrointestinal tract contains details other than the basic sentence stating that the examination is not dedicated to it.
- Position 39.3280 Post-void residual volume: is added if the section of the radiological report concerning the bladder mentions this analysis.

To complete the billing of an abdominal ultrasound, we need to add one more detail: whether the ultrasound is specific to the upper abdomen, the lower abdomen or the whole abdomen is analyzed. To do this, here are the 3 possible positions:
- Position 39.3250 defines an ultrasound of the upper abdomen. If the additional findings in the report state that the lower abdominal check shows no abnormality, the examination is focused on the upper abdomen.
- Position 30.3260 defines an ultrasound of the lower abdomen. If the additional findings in the report state that the upper abdominal check shows no abnormality, the examination is focused on the lower abdomen.
- Position 39.3240 defines an ultrasound of the entire abdomen (upper and lower). If neither of the above two conditions is met, this means that the report concerns the entire abdomen, so this position is used.
An example of abdominal US structured report is:

"Kidneys: no cavity expansion. No detectable calculi, subject to technical limitations. Global vascularity preserved on color Doppler. Discrete irregularity of the contour of the left kidney without atrophy.
Peritoneum: no free fluid.
Bladder: walls regular and not thickened.
Other findings: a check of the upper abdomen showed no abnormalities. "

As we can see from this report, there is one mention of the term "Doppler", so position 39.3510 should be billed. On the other hand, there is no mention of the digestive tract or post-void residual, which rules out positions 39.3265 and 39.3280.
Regarding the type of examination, we note that it states that "a check of the upper abdomen ...", which indicates that it was not the priority, and that this examination is therefore an ultrasound of the lower abdomen.
Adding the basic positions for ultrasound, here are the codes concerned by this examination:

Position TARMED	**Description**	**PT**	**PM**
39.0010	Basic consultation/operating unit Outpatient radiology institute	78.43 pts	6.25 pts
39.3800	Basic technical service 0, large ultrasound examination, outpatient	10.68 pts	0 pts
39.3510	Vessel sonography as part of organ or soft tissue examination	32.04 pts	42.48 pts
39.3260	Trans-abdominal ultrasound examination	25.64 pts	37.76 pts
Total Points		**146.79 pts**	**86.49 pts**
Price	**With a factor of 0.89CHF for the region where the exam was performed**	**130.64 CHF**	**76.97 CHF**

2.2 Model development

Using this information a modular Natural Language Processing algorithm based on the Natural Language Toolkit (NLTK) library was developed.
Detection of exam type (abdominal US)
The first module is used to detect a specific TARMED service (abdominal US in this case).
Detection of anatomical region of the abdomen
In order to detect the type of exams (codes 39.3240, 39.3250 and 39.3260), two steps were developed.
The first step is to analyze the section of the structured report named: "other findings" section. In fact, if in this section it is written "a check of the lower abdomen shows no abnormalities", we can deduce that the examination performed is an examination of the upper abdomen. Conversely, if it's stated that the check of the upper part shows no abnormalities, it's an inferior examination. If the "other findings" section mentions neither an upper nor a lower examination, it's a full abdominal examination.
The problem arises when the section is missing from the structured report, for one reason or another.
This is where the second step comes in. The aim of this second section is to see which organs are described in the report, and to determine whether they are upper or lower. Some organs are exclusive to the lower exam, and others to the upper, but many fall equally into both. We had to find the list of organs exclusive to each examination, which is as follows:
- Upper examination: liver, spleen, gallbladder, pancreas
- Lower examination: genitals, bladder

Three situations can be derived from these two lists:

1. One or more organs from the upper group are present, none from the lower group: the examination is an upper abdominal ultrasound.
2. One or more organs from the lower group are present, none from the upper group: the examination is a lower abdominal ultrasound.
3. One or more organs from the lower group are present, one or more organs from the upper group are present: the examination is a full abdominal ultrasound.

Based on this analysis, we can bill the following positions:
- 39.3540 for a full abdominal ultrasound
- 39.3250 for upper abdominal ultrasound
- 39.3260 for lower abdominal ultrasound

Detection of Doppler exam
To detect if it was a vessel sonography - position 39.3510, we simply checked for the presence of certain keywords in the radiology report. Since the presence of vessel sonography can be found in almost every section of the report, we have be targeting the entire report. The keywords we have been looking for are: "doppler", "suprahepatic", "vascularized", "vein", "hepatopetal", "hepatofuge".
These words define the presence of vessel sonography, so the position can be billed without doubt.

Detection of post-void residual
As with vessel sonography, the aim here is to detect the presence of the word or part of the word "void", which is specific to a post-void residual examination (position 39.3280).
Detection of a digestive tract examination
In the case of an examination of the digestive tract (position 39.3265), we analyzed the section specific to it. There are three possibilities:

1- The section does not exist, and the position is not billed.
2- The section exists, and contains a "non-dedicated examination" mention, the examination is not billed.
3- Section exists and contains information. The examination is billed.

Detection of soft tissue examination
The presence of a soft-tissue examination (39.3420) is currently discussed directly between the technologist and the radiologist, as it is rarely performed. However, we have deduced a list of key words that represent the presence of this position, as follows: "hernia", "white line", "ventration", "Valsalva", "inguinal".
If one of these words is present in the radiology report, then position 39.3420 - soft tissue examination is billed.
Detection of an examination of the renal arteries
For this examination, the presence of certain words or terms must be checked. Here is the list of keywords corresponding to this position: "intraparenchymal", "systolic velocity", "systolic peaks", "resistance index". systolic velocity", "systolic peaks", "resistance index", "resistivity index", "flow analysis", "renal hilum", "hilar renal artery", "anastomosis", "intrarenal".
All these words indicate the presence of a renal artery examination and allow billing of position 39.3610.
A second NLP algorithm based on SPACY has also been developed, with the objective of a double validation of the first developed NLP algorithm. If both algorithms obtain the same billing, then it's most likely correct. If, on the other hand, the two algorithms produce different results, human intervention may be required to decide between them. Detailed on this algorithm are here omitted because of space limit.

To train and test the algorithm a dataset of 170 exams corresponding to US abdominal examinations along with their radiology report were extracted from our RIS.
The results of the algorithm were validated by an experienced technologists which identified possible discrepancy between the algorithm results and the correct billing. This check was carried out on a batch of data containing 95 samples, extracted from their RIS system. A confusion matrix was used to analyze the results.

3. RESULTS

In all 95 data samples, the NLTK algorithm was able to detect the billing codes correctly 100% of the time. In all our 95 data samples, the Spacy algorithm was able to detect the billing codes correctly in 86.3% of cases. This algorithm tends to overestimate the type of abdominal examination present in the report. Indeed, the 13 cases in which the algorithm made an error were cases where it detected a full abdominal ultrasound when the examination was a simple lower or upper abdomen.
However, this over-assessment is always due to an element that does indeed give rise to doubt. In such cases, it is worthwhile having an MRT re-check the report in question.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

Two models to predict the correct billing code for US abdomen examinations has been developed and tested on clinical data. Our best model (NLTK) was found to predict the correct billing code with an 100% accuracy, while the second model reach only 86%.

5. CONCLUSION

The NLTK model provides reliable and efficient estimation of billing codes for abdominal ultrasound, facilitating the task of the technologies who saves time and avoids possible human errors. However, the model has not yet been validated externally in clinical workflow, on a larger amount of data and on possibly less structured reports. To do so we are working to integrate the developed tool in our RIS. We are also working on extending these algorithms to other modalities and indications.

6. SUBMISSION ELSEWHERE

This work is not submitted elsewhere.