Published online: 2019-10-09

Validation of Semantic Analyses of Unstructured Medical Data for Research Purposes

Validierung von semantischen Analysen von unstrukturierten medizinischen Daten für Forschungszwecke

Authors

Roman Michael Pokora[1], Lucian Le Cornet[1, 2], Philipp Daumke[3], Peter Mildenberger[4], Hajo Zeeb[5], Maria Blettner[6]

Affiliations

1 Institute for Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center of the Johannes Gutenberg University Mainz, Mainz
2 Studienzentrale, Nationales Centrum für Tumorerkrankungen Heidelberg, Heidelberg
3 Averbis GmbH, Freiburg
4 Klinik und Poliklinik für Diagnostische und Interventionelle Radiologie, University Medical Center of the Johannes Gutenberg University Mainz, Mainz
5 Leibniz-Institut für Präventionsforschung und Epidemiologie (BIPS), Prevention and Evaluation, Bremen
6 Institut fur Medizinische Biometrie Epidemiologie und Informatik, Johannes-Gutenberg Universität Mainz, Mainz

Key words

secondary data, unstructured free text, text-mining, validation

Schlüsselwörter

Sekundärdaten, unstrukturierte Freitext, Text-Mining, Validierung

Bibliography

DOI https://doi.org/10.1055/a-1007-8540
Online-Publikation: 9.10.2019
Gesundheitswesen 2020; 82 (Suppl. 2): S158–S164

ISSN 0949-7013

Correspondence

Dr. Roman Michael Pokora
Institute for Medical Biostatistics, Epidemiology and Informatics (IMBEI)
University Medical Center of the Johannes Gutenberg University Mainz
Obere Zahlbacher Straße 69
55131 Mainz
Germany
pokora@uni-mainz.de

ABSTRACT

Background In secondary data there are often unstructured free texts. The aim of this study was to validate a text mining system to extract unstructured medical data for research purposes.

Methods From a radiological department, 1,000 out of 7,102 CT findings were randomly selected. These were manually divided into defined groups by 2 physicians. For automated tagging and reporting, the text analysis software Averbis Extraction Platform (AEP) was used. Special features of the system are a morphological analysis for the decomposition of compound words as well as the recognition of noun phrases, abbreviations and negated statements. Based on the extracted standardized keywords, findings reports were assigned to the given findings groups using machine learning methods. To assess the reliability and validity of the automated process, the automated and two independent manual mappings were compared for matches in multiple runs.

Results Manual classification was too time-consuming. In the case of automated keywording, the classification according to ICD-10 turned out to be unsuitable for our data. It also showed that the keyword search does not deliver reliable results. Computer-aided text mining and machine learning resulted in reliable results. The inter-rater reliability of the two manual classifications, as well as the machine and manual classification was very high. Both manual classifications were consistent in 93 % of all findings. The kappa coefficient is 0.89 [95 % confidence interval (CI) 0.87–0.92]. The automatic classification agreed with the independent, second manual classification in 86 % of all findings (Kappa coefficient 0.79 [95 % CI 0.75–0.81]).

Discussion The classification of the software AEP was very good. In our study, however, it followed a systematic pattern. Most misclassifications were found in findings that indicate an increased risk of cancer. The free-text structure of the findings raises concerns about the feasibility of a purely automated analysis. The combination of human intellect and intelligent, adaptive software appears most suitable for mining unstructured but important textual information for research.

ZUSAMMENFASSUNG

Hintergrund In Sekundärdaten existieren oftmals unstrukturierte Freitexte. In dieser Arbeit wird ein Text-Mining-System validiert, um unstrukturierte medizinische Daten für Forschungszwecke zu extrahieren.

Methoden Aus einer radiologischen Klinik wurden aus 7102 CT-Befunden 1000 zufällig ausgewählt. Diese wurden von 2 Medizinern manuell in definierte Befundgruppen eingeteilt. Zur automatisierten Verschlagwortung und Klassifizierung wurde die Textanalyse-Software Averbis Extraction Platform (AEP) eingesetzt. Besonderheiten des Systems sind u. a. eine morphologische Analyse zur Zerlegung zusammengesetzter Wörter sowie die Erkennung von Nominalphrasen, Abkürzungen und negierten Aussagen. Anhand der extrahierten standardisierten Schlüsselwörter werden Befundberichte mithilfe maschineller Lernverfahren den vorgegebenen Befundgruppen zugeordnet. Zur Bewertung von Reliabilität und Validität des automatisierten Verfahrens werden die automatisierten und 2 unabhängige manuelle Klassifizierungen in mehreren Durchläufen auf Übereinstimmungen hin verglichen.

Ergebnisse Die manuelle Klassifizierung war zu zeitaufwendig. Bei der automatisierten Verschlagwortung stellte sich in unseren Daten die Klassifizierung nach ICD-10 als ungeeignet heraus. Ebenfalls zeigte sich, dass die Stichwortsuche keine verlässlichen Ergebnisse liefert. Computerunterstütztes Textmining in Kombination mit maschinellem Lernen führte zu verlässlichen Klassifizierungen. Die Inter-Rater-Reliabilität der beiden manuellen Klassifizierungen, sowie der maschinellen und der manuellen Klassifizierung war sehr hoch. Beide manuelle Klassifizierungen stimmten in 93 % aller Befunde überein. Der Kappa-Koeffizient beträgt 0,89 [95 % Konfidenzintervall (KI) 0,87–0,92]. Die automatische Klassifizierung stimmte in 86 % aller Befunde mit der unabhängigen, zweiten manuellen Klassifizierung überein (Kappa-Koeffizient 0,79 [95 % KI 0,75–0,81]).

Diskussion Die Klassifizierung der Software AEP war sehr gut. In unserer Studie folgte sie allerdings einem systematischen Muster. Die meisten falschen Zuordnungen finden sich in Befunden, die auf ein erhöhtes Krebsrisiko hinweisen. Die Freitextstruktur der Befunde lässt Bedenken hinsichtlich der Machbarkeit einer rein automatisierten Analyse aufkommen. Die Kombination aus menschlichem Intellekt und einer intelligenten, lernfähigen Software erscheint als zukunftsweisend, um unstrukturierte aber wichtige Textinformationen der Forschung zugänglich machen zu können.

Introduction

Epidemiologic studies often rely on secondary data that have not been collected originally for research purposes [1, 2]. With the technical development and wider adoption of information systems, e. g. Radiology Information System (RIS) or Electronic Health/Medical Record Systems, large amounts of medical text data are produced in health institutions. These data include information about diagnosis and treatments, the patient's medical history, etc. In many cases the respective information is stored as unstructured data and therefore its usage is often hampered by different documentation styles and complex free texts as the extraction of relevant information is very time-consuming.

In a German and a combined European cohort study [3–5], we evaluated the risk of childhood cancer after exposure to ionizing radiation from computed tomography scans (CT). We collected data from the radiological information systems in 20 German hospitals to obtain the cohort and the exposure. These data were linked to the data from the German childhood cancer registry to assess the outcome. However, cancer risks associated with CT examinations must always be considered in the context of competing risks as well as potential confounders. It is possible that new cancers may occur due to underlying risk factors or that a cancer disease was already in progress but not yet diagnosed rather than being induced by the radiation exposure from the CTs. Information to control for these errors were available only in the radiological reports. As the reports are written as free text, it was difficult to retrieve and use information from these reports. To make the needed information accessible, text mining was considered as a solution to the problem. For this purpose, a criteria list of conditions and a categorization of overall diagnoses according to risk status was needed. Relevant risk groups were that patients were already diseased with cancer, had an elevated risk of cancer, had an elevated risk of mortality or had no elevated risk of cancer or mortality.

The aim of this work is to describe and evaluate a procedure to extract information from the data collected via the RIS in one of the participating hospitals.

Materials and methods

Data collection and management

Briefly, this study was performed by linking pseudonymized cohort data of children exposed to CT at the University Medical Centre Mainz. From the RIS, all available data for children receiving CTs before the age of 15 and between 1 January 1980 and 31 December 2010 were extracted as a comma-separated values file. Data regarding the date of the examination and the indication as well as the full radiologic report were extracted for all eligible examinations.

Approvals for the study were obtained from the ethics committee of the Medical Chamber of Rhineland Palatinate, and the data protection officer of the University Medical Centre Mainz [3].

From all CTs, a random sample (n = 1,000, "main sample") stratified by calendar year was drawn. From this main sample, separate stratified subsets were created. Radiological reports, which only referred to other reports or consisted of empty strings, were excluded.

Criteria list for elevated risk of cancer, risk of mortality, and the definition of the risk groups

The list of diseases associated with a higher risk for either cancer development (ICD-10: D70, D80–83, Q85, Q90–93, Q95, Q97–99) or mortality (A00–09, A15–19, A30–41, A75–99, B20–24, D68.

E84, E88, G00–09, G12–13, G80–83, G90–93, G95–96, G98, I00–09, I26, I28 I30, I34–38, I40–43, I46–51 I60–69, I71–74, J05, J09–18, K56, K65, K71–72, K74, M30, N17-N19, P10–11, P27, P36, P52, P77, P91, Q00–07, Q20–28, Q32-Q34, Q67, Q71–79, Q80–82, Q85–87, Q89, T00–09, T20–21, T27–32, T34–35, T36–50, T51–65, T74, T79) were adopted from a preceding study [6] and were reviewed by 2 physicians with epidemiological or radiological expertise. The lists were checked for completeness and for relevance in terms of the study population and the observation period. Diseases, which do not occur before the 15th year of life, were removed as well as mental diseases resulting from drug abuse.

Relevant risk groups were defined based on the preceding study on cancer risk after conventional x-ray examinations in children [6]:

- G1 definitely diseased with cancer
  Children who have been examined using CT in order to successfully verify a cancer or to treat a cancer or have had a cancer are excluded from the analysis.
- G2 a priori higher risk of cancer development
  Children who have been examined due to a suspected cancer which was not verified or children who suffer from a disease associated with a higher cancer risk.
- G3 higher risk of mortality
  Children who have a higher risk of dying due to prevalent or previous diseases.
- G4 no elevated risk of cancer or mortality
  All children who do not belong to any of the other groups.

Information extraction approaches

We evaluated different approaches which were implemented consecutively and are described below. Steps numbers represent the consecutive order of our approach.

Step 1 The radiological reports of test data 1 (n = 100) were classified manually using ICD-10 and the defined risk groups by a medical doctoral candidate (rater 1). A second subset (test data 2 (n = 104)) was drawn and classified by risk groups and diagnostic groups. The diagnostic groups were expanded on the basis of the data found. The groups represent a combination of distinct diagnoses and indications.

Step 2 The radiological reports of test data 1 and 2 (n = 204) were searched for keywords derived from the criteria lists. Ambiguous terms and abbreviations such as “trauma” were excluded. The terms were translated into regular expressions (RegEx) to include deviations and spelling errors. If a keyword was found, the respective result was assigned to the corresponding risk group.

Step 3 All consecutive steps were performed with the automated text mining tool Averbis Extraction Platform (AEP) [7]. It includes functions such as negation recognition, morpho-semantic analysis, hierarchical terminology and key word weighting. A radiological term mapping component that utilizes the German translation of RADLEX [8, 9] was first supplemented by the relevant terms from ICD-10. The included terms (RADLEX + ICD-10) were then assigned to the risk groups. Afterwards the test data was entered into the AEP. Any occurrence of one of the keywords within the radiological reports led to an assignment to the corresponding risk group.

The test data together with the manual classifications were inserted into the AEP and analyzed using the full functionality of the AEP. This automated procedure extracted more preferred terms than the manual classification. The relevance of the extracted terms was learned by analyzing their distributions within the risk groups which have been classified manually. The respective rules were generated automatically.

The validity (proportion of correct assignments) was tested by 10-fold cross-validations: the inserted dataset was divided into 10 separate sets. 9 sets served as training dataset to extract the rules with which the 10th set (classified dataset) would be classified. This procedure was repeated ten times. The average prediction accuracy of all ten iterations is the prediction accuracy of the classification model. For every processed radiological report, the assigned risk group and the probability for correct assignment were determined. The latter is called a confidence value and has a range from 0 to 1.

Step 4 The correct classification of the AEP was verified in 100 additional radiological reports with high confidence values. If needed, the rater 1 corrected the risk groups and they were reassigned; otherwise they were marked as approved. These manually verified reports were flagged and given higher importance during the next automated learning process. Additionally, 100 further radiological reports were drawn from the main sample and were classified manually. The automated classification based on machine learning was repeated (see step 3) using the enlarged test data (n = 304) enriched with information from the manual verification.

Step 5 Following the automated classification from step 3, the 200 classifications with the lowest confidence values were reviewed manually and reassigned if needed as described above. Additionally 200 further radiological reports were drawn and classified manually. The enriched and enlarged test data (n = 504) was again classified through the AEP as described in step 3.

Step 6 All remaining radiological reports were classified manually as well. The automated classification based on machine learning was repeated (see step 3) using the enlarged and enriched test data (n = 994).

Step 7 The risk groups G2 and G4 proved not to be very distinct and tended to overlap because both groups contain large numbers of healthy people. Thus, both groups were combined and the automated classification with AEP was repeated as described in step 3.

Step 8 In this step 160 classifications with the lowest confidence values were selected and searched for relevant keywords. These keywords were flagged as of higher importance for the automated learning process and directly entered into the

AEP. The optimized automated classification was repeated with the whole test data using all 4 risk groups.

Step 9 For comparison, the complete main sample was additionally classified by a physician (rater 2).

Step 10 The results of the second manual classification by rater 2 was then used to repeat the automated classification with AEP on the whole test data (n = 994) including the manual revisions and the weighted keywords.

Statistical analysis

Statistical analysis was performed using SAS v9.4 (SAS Institute Inc., Cary, NC, USA). The reliability of the AEP was determined by calculating the inter-rater reliability between the last automated classification and each of the two manual classifications.

The response of the manual raters 1 and 2 were compared with those of the AEP. Cohen's Kappa Statistic was used to determine the level of agreement between each rater and the AEP and 95 % confidence intervals were calculated for the obtained kappa. The Fleiss Kappa Statistic was used to determine the overall mean kappa rating between subgroups of raters (rater 1, rater 2) and the automatic rating of the AEP.

Results

The hospital dataset contained 7,102 eligible radiological reports. From these, a random sample of 1,000 reports was drawn. 6 reports were missing and therefore excluded, leaving 994 reports for inclusion in the analyses.

The manual classification (Step 1) showed that classification using ICD-10 is not feasible. From the 100 radiological reports in test data 1, only 50 % could be assigned to an ICD-10 code. In 39 %, an assignment was not possible, and in 11 % an assignment was judged unreasonable. The poor results were caused by the inappropriate structure and content. In many cases, no diagnosis and hence no diseases were given. For example, for a tumor exclusion, no prevalent disease is present. The radiological findings contained medical terminology, as well as general and clinic-specific abbreviations, guesses, negations, compounds, or measurements. Therefore, further information, which was not included in some reports, would be necessary to fully understand the data. In addition, typing errors occurred. In most cases, the radiologist is not provided with any further patient data apart from the indication. Therefore, the radiological findings are mostly only descriptions of normal morphology or – if present - its pathological changes.

In step 1 the classification by risk groups (n = 204) showed that 44 % of all CTs were performed in relation to a (potential or confirmed) cancer (G1 = 27 % and G2 = 17 %). 27 % had a higher risk of mortality (G3) and 29 % had no elevated risk of cancer or mortality (G4).

The keyword search (step 2) could insufficiently reproduce findings from the manual classification. In case of the 11 % where a reasonable assignment was not possible, there was no distinct disease term that could be extracted from the radiological reports. The structure of the radiological reports made the planned use of software-supported keyword searches using RegEx unreliable. The ICD-10 and the ruled-based assignment to risk groups also proved not to be suitable. Most problems were caused through the heterogeneity of the radiological reports. For example, the reports used differing terms in the ICD-10 descriptions or information were only available with information relevant to an understanding of the radiological report (contextual information).

The initial validity (step 3) was 0.69 (▶ **Table 1**). Logically, the validity could only be slightly improved (1 %) by revising the assignments with high confidence values (step 4). By revising the 200 assignments with low confidence values (step 5), the overall validity was increased by 5 %. The enlargement to the whole test data (step 6) led to an increase of only 1 %. In comparison to step 6, the use of weighted keywords (step 8) improved the validity by 10 %, thus resulting in an overall validity of 0.86. The use of different training data sets led only to marginally changed validity.

The inter-rater-reliability between the two manual classifications was good. They coincided in 93 % of all radiological reports. The kappa coefficient was 0.89 [95 % confidence interval (CI): 0.87–0.92]. The automated classifications coincided in 86 % of all radiological reports with the second manual classification (step 10). The kappa coefficient was 0.79 [95 % CI: 0.75–0.81].

The choice of the training data set only marginally influenced the final results. The concordance of the automated classification with the second manual classification was only 2 % lower (84 %) when using the second manual classification than the first manual classification, which was used for initial training. The kappa coefficient was 0.76 [95 % CI: 0.72–0.79]. The disagreement of the two manual classifications did not follow a systematic pattern and spread across all risk groups. The misclassification of the AEP, however, followed a systematic pattern: some reports were incorrectly classified as G4 (no elevated risk). Furthermore, some reports from G2 (not verified suspicion of cancer) were classified as G1 (diseased with cancer) (▶ **Table 2**, ▶ **Fig. 1**).

Total agreement between the two raters and the AEP were seen in 812 radiology reports. Agreement between rater 1 and rater 2 differed in 71 reports. In 160 instances rater 1 and the AEP and in 141 instances rater 2 and the AEP differed. The agreement is stron-

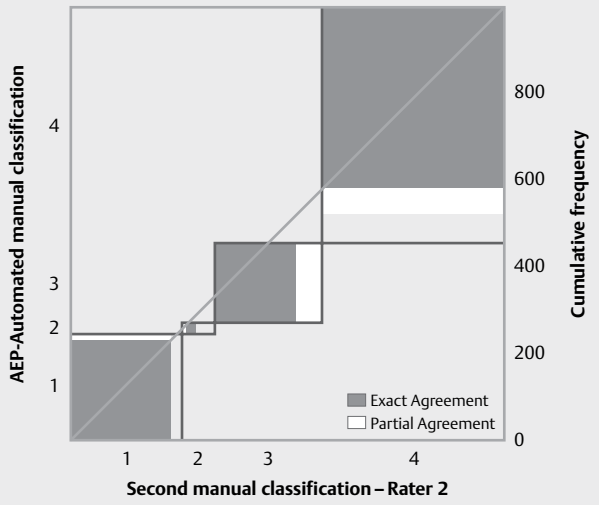
▶ **Table 1** Average validity of the automated, machine learning classification.

Step	Observations	Description of training dataset	Validity
3	204	All 204 randomly selected manually classified radiological reports.	0.69
4	304	Manual classification of 100 results with high confidence values.	0.70
5	504	Manual classification of 200 results with low confidence values.	0.75
6	994	Manual classification of the 490 remaining radiological reports.	0.76
7	994	Combination of G2 and G4 risk groups.	0.81
8	994	Weighting of manually identified key words and usage of all four risk groups.	0.86
9	994	Two manual classifications	0.93
10	994	Using the second manual classification as training set	0.84

▶ **Table 2** Agreement data and Kappa of the automated and manual classifications.

2nd rater vs. AEP						1st rater vs. 2nd rater					
Rater 2	**AEP[1]**					**Rater 1**	**Rater 2**				
	G1	**G2**	**G3**	**G4**	**Sum**		**G1**	**G2**	**G3**	**G4**	**Sum**
G1	232	0	0	23	255	G1	237	2	0	1	240
G2	11	22	1	44	78	G2	11	72	1	12	96
G3	3	0	183	58	244	G3	5	1	218	8	232
G4	0	1	0	416	417	G4	2	3	25	396	426
Sum	246	23	184	541	994	Sum	255	78	244	417	994
Kappa	0.79 [CI 0.75–0.81]					**Kappa**	0.89 [CI 0.87–0.92]				
Concordance	86 %					**Concordance**	93 %				

[1] Averbis Extraction Platform



▶ **Fig. 1** Agreement plot between automated manual classification and rater 2.

▶ **Table 3** Level of agreement of automated and manual classifications.

Rater				
Risk Group	**Rater 1**	**Rater 2**	**AEP[1]**	**Kappa (SD)**
G1	240	255	246	0.91 (SD 0.02)
G2	96	78	23	0.57 (SD 0.02)
G3	232	244	184	0.83 (SD 0.02)
G4	426	417	541	0.79 (SD 0.02)
Overall	994	994	994	0.81 (SD 0.01)

[1] Averbis Extraction Platform

gest in risk group 1 (prevalent cancer, ▶ **Table 3**). The Fleiss kappa coefficient between both raters and the AEP was 0.81 ($z = 68.82$; $p < 0.001$).

Discussion and conclusion

This study described a method to make the information of free medical texts accessible for epidemiological research. Interactively we assessed the feasibility and validity of manual classifications, data extraction by means of classical keyword searches, and finally intelligent text mining tools capable of automated learning. The manual classification proved to be the most reliable but also most resource expensive method. Conventional keyword searches failed to detect the relevant, hidden contextual information in the text. The text analysis tool AEP however provided promising results suitable for large scale research projects using medical routine data.

In our case we extracted medical conditions that appear to be more frequently associated with the use of CT in pediatrics. It showed that in pediatric CT patients the majority of scans were unrelated to cancer suspicion (Group 3 and 4: 66 %). According to the results of the first manual classifications 24 % of the findings were indicating cancer (Group 1) and 10 % had an a priori higher risk of cancer development (Group 2). This distribution was like estimates of experts which were interviewed during the early research phase.

The ICD-10 classification is not suitable for classifying radiological reports. A manual classification is possible but very labor-intensive. On average, the manual classification of only one result took 2 to 3 min. This is not feasible if large numbers of reports are to be classified.

Conventional software-based methods, such as keyword search and rule-based classifications, were also not satisfying in our case. Given the nature of free text, the relevant information was mostly hidden in the semantical context and thus accessible to intelligent approaches only. To account for this, the composition of the contextual information must be translated into complex rules which detect the combined occurrence of specific terms. However, due to the heterogeneity of the report texts, the effort needed to program these rules easily outweighs the efforts of manual classification.

Finally, we opted for a more elaborate and complex approach. In this approach, state-of-the-art classification methods based on statistical learning algorithms were applied to the problem. As input, these approaches require sufficiently large sets of manually classified examples (training data) based on which they derive an optimal statistical model to automatically classify unseen examples. The combination of manual classification, machine learning and semantic text analysis programs is possible. The validity of this approach can be improved with the revision of a fraction of automated classification with low confidence value and the weighting of keywords.

The combination of 2 risk groups led to a small improvement and a loss of information and was therefore not deemed useful. However, in other research settings, with more distinct groups, this might lead to improved results.

Overall, a validity of 86 % was achieved by the machine-learning approach. The inter-rater reliability between the AEP and the 2 manual classifications was good. In our case the misclassification of the software followed a systematic pattern: text components indicating existing cancer risks sometimes remained undetected and the findings have been mistakenly assigned to Group 4 (no risk). In the overarching research setting, this would lead to an underestimation of the true risk of childhood cancer associated with diagnostic CT.

Overall, in order to achieve the stated validity, the revision and keyword analysis of a subsample consisting of classification with low confidence values is theoretically sufficient. Thus, this method is very well suited to classify the findings in a large cohort study or similar research settings. Further improvements in the validity of the method are possible through the concatenation of several automated classification methods (boosting). However, implementation is rather complicated.

A weakness of our approach was the relatively small sample in combination with partly overlapping risk groups. In an exploratory analysis the classification model derived from step 6 was applied to a large dataset of three hospitals [3, 5]. The validity for each individual clinic could be improved. This is likely due to the more heterogeneous training data and the absolute number of datasets per clinic. However, more work is needed to better understand conditions improving the validity in larger datasets from different sources. In the statistical analysis plan we planned the reproduction of the search in other countries of the European cohort study to assess if there are different results according to language. Results from a British study suggest that a semantic search in radiologist reports showed a satisfactory performance [10]. The authors also validated a subsample by a pediatric radiologist but they did not report inter-rater-reliability with the automatic coding procedure.

Radiological reports have changed significantly, especially in recent years. The discussion about whether and how a report should be structured to serve its different purposes has been ongoing for several years [11]. In future, studies in radiology, which use current data, will be able to use structured report texts [12].

Meanwhile, there are promising approaches in the field of deep learning also for text classification tasks. We have implemented document classification tasks based on convolutional neural networks in another domain (on patents) and had approximately 5–7 % better accuracy than with previous machine-learning methods [13]. Other topics to mention are boosting [14] or word embedding [15].

In conclusion, this study described a method to extract certain information from complex free texts in radiological reports. In our case we extracted medical conditions that appear more frequently associated with the use of CT in pediatrics. Such information provides insights for clinical practice and epidemiological studies. The method is potentially transferable to any other research, which plans to utilize information from free texts; however, further research on various features of the automated approach is needed to improve applicability.

Conflict of Interest

Philipp Daumke is the CEO of the Averbis GmbH. The AEP was not provided free of Charge, however the second human classification was provided free of Charge. The Averbis GmbH was not involved in the Analysis and the judgement of the results of the different approaches. Besides that there are no conflicts of interest for any of the authors of this manuscript.

References

[1] Swart E, Gothe H, Geyer S et al. Gute Praxis Sekundärdatenanalyse (GPS) Leitlinien und Empfehlungen. Gesundheitswesen 2015; 77: 120–126. doi:10.1055/s-0034-1396815

[2] Swart E, Bitzer EM, Gothe H et al. A Consensus German Reporting Standard for Secondary Data Analyses, Version 2 (STROSA-STandardisierte BerichtsROutine für Sekundärdaten Analysen). Gesundheitswesen 2016; 78: e145–e160. doi:10.1055/s-0042-108647

[3] Krille L, Dreger S, Schindel R et al. Risk of cancer incidence before the age of 15 years after exposure to ionising radiation from computed tomography: results from a German cohort study. Radiat Environ Biophys 2015; 54: 1–12. doi:10.1007/s00411-014-0580-3

[4] Bosch de Basea M, Pearce MS, Kesminiene A et al. EPI-CT: Design, challenges and epidemiological methods of an international study on cancer risk after paediatric and young adult CT. J Radiol Prot Radiol Prot 2015; 35: 611–628. doi:10.1088/0952-4746/35/3/611

[5] Krille L, Jahnen A, Mildenberger P et al. Computed tomography in children: multicenter cohort study design for the evaluation of cancer risk. Eur J Epidemiol 2011; 26: 249–250. doi:10.1007/s10654-011-9549-6

[6] Hammer GP, Seidenbusch MC, Schneider K et al. A cohort study of childhood cancer incidence after postnatal diagnostic X-ray exposure. Radiat Res 2009; 171: 504–512. doi:10.1667/RR1575.1

[7] Daumke P, Enders F, Simon K et al. Semantic annotation of clinical text – the averbis annotation editor. Proceedings of the GMDS 2010 Mannheim

[8] Langlotz CP. RadLex: a new method for indexing online educational materials. Radiographics 2007; 27: 62. doi:10.1148/rg.266065168

[9] Marwede D, Daumke P, Marko K et al. RadLex – German version: a radiological lexicon for indexing image and report information. Rofo 2009; 181: 38–44. doi:10.1055/s-2008-1027895

[10] Journy NM, McHugh K, Harbron RW et al. Medical conditions associated with the use of CT in children and young adults, Great Britain, 1995-2008. Br J Radiol 2016; 89: 1995–2008. doi:10.1259/bjr.20160532

[11] Boland GW, Duszak R Jr. Structured reporting and communication. J Am Coll Radiol 2014; 12: 1286–1288. doi:10.1016/j.jacr.2015.08.001

[12] Pinto dos Santos D, Scheibl S, Arhold G et al. A proof of concept for epidemiological research using structured reporting with pulmonary embolism as a use case. Br J Radiol 2018, Jun 5. doi:10.1259/bjr.20170564

[13] Kim Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014. Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics. 10.3115/v1/D14-1181

[14] Bloehdorn S, Hotho A. Boosting for Text Classification with Semantic Features. In: Mobasher B, Nasraoui O, Liu B, Masand B.eds. Advances in Web Mining and Web Usage Analysis. WebKDD 2004. Lecture Notes in Computer Science. 2006: 3932: 149–166 Springer; Berlin, Heidelberg: doi:10.1007/11899402_10

[15] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: ICLR Workshop Papers 2013

Hinweis
Dieser Artikel wurde gemäß des Erratums vom 20.1.2020 geändert.

Erratum
Im oben genannten Artikel wurde der Name des Autors Lucian Le Cornet korrigiert.