Article published online: 2023-05-09

Automated Classification of Free-Text Radiology Reports: Using Different Feature Extraction Methods to Identify Fractures of the Distal Fibula

Automatisierte Klassifizierung von radiologischen Freitext-Befunden: Analyse verschiedener Feature-Extraction-Methoden zur Identifizierung distaler Fibulafrakturen

OPEN ACCESS

Authors

Cornelia L.A. Dewald[1], Alina Balandis[2], Lena S. Becker[1], Jan B. Hinrichs[1], Christian von Falck[1], Frank K. Wacker[1], Hans Laser[2], Svetlana Gerbel[2], Hinrich B. Winther[1], Johanna Apfel-Starke[2]

Affiliations

1 Institute for Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany

2 Centre for Information Management (ZIMt), Hannover Medical School, Hannover, Germany

Key words

ankle, Natural Language Processing, Text Mining, Fibula Fracture, Automatic Classification, Data Set

received 17.10.2022
accepted 18.02.2023
published online 09.05.2023

Bibliography

Fortschr Röntgenstr 2023; 195: 713–719
DOI 10.1055/a-2061-6562
ISSN 1438-9029

© 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence

Frau Dr. Cornelia Lieselotte Angelika Dewald
Radiology, Hannover Medical School, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany
Tel.: +49/5 11/5 32 34 21
dewald.cornelia@mh-hannover.de

ABSTRACT

Purpose Radiology reports mostly contain free-text, which makes it challenging to obtain structured data. Natural language processing (NLP) techniques transform free-text reports into machine-readable document vectors that are important for creating reliable, scalable methods for data analysis. The aim of this study is to classify unstructured radiograph reports according to fractures of the distal fibula and to find the best text mining method.

Materials & Methods We established a novel German language report dataset: a designated search engine was used to identify radiographs of the ankle and the reports were manually labeled according to fractures of the distal fibula. This data was used to establish a machine learning pipeline, which implemented the text representation methods bag-of-words (BOW), term frequency-inverse document frequency (TF-IDF), principal component analysis (PCA), non-negative matrix factorization (NMF), latent Dirichlet allocation (LDA), and document embedding (doc2vec). The extracted document vectors were used to train neural networks (NN), support vector machines (SVM), and logistic regression (LR) to recognize distal fibula fractures. The results were compared via cross-tabulations of the accuracy (acc) and area under the curve (AUC).

Results In total, 3268 radiograph reports were included, of which 1076 described a fracture of the distal fibula. Comparison of the text representation methods showed that BOW achieved the best results (AUC = 0.98; acc = 0.97), followed by TF-IDF (AUC = 0.97; acc = 0.96), NMF (AUC = 0.93; acc = 0.92), PCA (AUC = 0.92; acc = 0.9), LDA (AUC = 0.91; acc = 0.89) and doc2vec (AUC = 0.9; acc = 0.88). When comparing the different classifiers, NN (AUC = 0,91) proved to be superior to SVM (AUC = 0,87) and LR (AUC = 0,85).

Conclusion An automated classification of unstructured reports of radiographs of the ankle can reliably detect findings of fractures of the distal fibula. A particularly suitable feature extraction method is the BOW model.

Key Points:

- The aim was to classify unstructured radiograph reports according to distal fibula fractures.
- Our automated classification system can reliably detect fractures of the distal fibula.
- A particularly suitable feature extraction method is the BOW model.

Citation Format

- Dewald CL, Balandis A, Becker LS et al. Automated Classification of Free-Text Radiology Reports: Using Different Feature Extraction Methods to Identify Fractures of the Distal Fibula. Fortschr Röntgenstr 2023; 195: 713–719

ZUSAMMENFASSUNG

Ziel Radiologische Befundtexte enthalten häufig Freitext, was eine strukturierte Datenauswertung erschwert. Natural language processing (NLP)-Techniken wandeln Freitext in maschinenlesbare Dokumentenvektoren um, die für die Entwicklung zuverlässiger, skalierbarer Methoden zur Datenanalyse wichtig sind. Ziel dieser Studie war es, unstrukturierte Röntgenbefunde nach Frakturen der distalen Fibula zu klassifizieren und die beste Text-Mining-Methode zu finden.

Material & Methoden Zur Erstellung eines eigenen deutschsprachigen Befunddatensatzes wurden mittels einer dedizierten Suchmaschine Sprunggelenks-Röntgenbilder identifiziert und die entsprechenden Befunde manuell nach Frakturen der distalen Fibula sortiert. Anhand der Daten wurde eine Machine-Learning-Pipeline erstellt, die die Textrepräsentationsmethoden Bag-of-Words (BOW), Term Frequency-Inverse Document Frequency (TF-IDF), Principal Component Analysis (PCA), Non-Negative Matrix Factorization (NMF), Latent Dirichlet Allocation (LDA) und Document Embedding (doc2vec) implementierte. Die extrahierten Dokumentvektoren wurden zum Trainieren von neuronalen Netzen (NN), Support Vector Machines (SVM) und logistischer Regression (LR) verwendet, um distale Fibulafrakturen zu erkennen. Die Ergebnisse wurden mittels Kreuztabellen bzgl. der Accuracy (acc) und der area under the curve (AUC) verglichen.

Ergebnisse Insgesamt wurden 3268 Röntgenbefunde inkludiert, von denen 1076 eine distale Fibulafraktur beschrieben. Der Vergleich der Textdarstellungsmethoden zeigte, dass BOW die besten Ergebnisse erzielte (AUC = 0,98; acc = 0,97), gefolgt von TF-IDF (AUC = 0,97; acc = 0,96), NMF (AUC = 0,93; acc = 0,92), PCA (AUC = 0,92; acc = 0,9), LDA (AUC = 0,91; acc = 0,89) und doc2vec (AUC = 0,9; acc = 0,88). Im Vergleich der Klassifikatoren erwiesen sich die NN (AUC = 0,91) gegenüber SVM (AUC = 0,87) und LR (AUC = 0,85) als überlegen.

Schlussfolgerung Durch die automatisierte Klassifikation von unstrukturierten Befunden von Sprunggelenksaufnahmen können Frakturen der distalen Fibula zuverlässig erkannt werden. Eine besonders geeignete Methode zur Feature Extraction ist das BOW-Modell.

Kernaussagen:

- Ziel war die automatisierte Klassifizierung unstrukturierter Röntgenbefunde entsprechend distaler Fibulafrakturen.
- Eine zuverlässige Detektion von distalen Fibulafrakturen ist durch das automatisierte Klassifizierungssystem gewährleistet.
- Eine besonders geeignete Methode zur Feature Extraction ist das BOW-Modell.

Introduction

The analysis of electronic health records (EHRs) lays the basis for a developing healthcare system, as it enables access to large data volumes [1–3], which support research and ultimately can increase patient safety and decrease healthcare costs [4, 5]. Radiological reports are a particularly rich source of compact clinical information within an EHR. These reports document information about the patient's health status and the radiologist's interpretation of medical findings. However, written radiological reports are often unstructured, which poses a challenge for the conversion into a computer-based representation [1, 6].

Machine learning (ML) and natural language processing (NLP) are subsections of artificial intelligence. Classic ML methods can model data, such as radiology reports, using (un-)supervised learning methods [7]. This typically requires pre-processing by means of NLP in order to extract machine-readable features from unstructured texts. In this step, feature extractors transform the raw data into a suitable internal representation. During this feature extraction, uncorrelated or superfluous features may be deleted, which can improve the accuracy of learning algorithms. Nevertheless, the complexity of the natural language used in free-text reports and the variations among the different dictation styles of radiologists can be problematic [8]. Thus, the choice of feature extraction methods during pre-processing of texts is particularly important [9]. In contrast, modern ML methods, such as neural networks (NN), have the capability to perform an end-to-end approach. This includes feature extraction in the training pipeline of the model as one of many tunable hyperparameters, potentially leading to a better adapted model. After the conversion of unstructured free text reports into feature vectors, classifiers can detect, extract, and classify patterns during (un-)supervised learning [6, 10]. Such structured information can, e. g., be the classification of patients into different groups.

NN has become the gold standard for text processing as it can achieve reliable results [11]. The current iteration of NN-based models is derived from large transformer language models, such as BERT [12]. Adaptations for the medical domain include BioBert [13] and ClinicalBERT [14]. BioBERT was mainly trained on 4.5 billion words of PubMed abstracts and 13.5 billion words of PMC articles. ClinicalBERT was trained on nearly 2 million anonymized notes by clinical physicians.

However, classic ML methods such as vector machines (SVM) have also been demonstrated to be suitable for the high dimensional vectors extracted by NLP and are thus used in recent studies [15]. Logistic regression (LR) is a well-established method, providing robust results [16].

Reports of X-ray images of the ankle are a suitable candidate to test a feature extraction/classification system, as fractures of the distal fibula are common (accounting for 70 % of all ankle fractures [17]). Distal fibula fractures can be isolated or combined with distal tibia fractures (bimalleolar or trimalleolar fractures) [18]. Unstable ankle

fractures are usually treated by open reduction and internal fixation [19, 20]. Subsequently, plenty of pre- and postoperative X-ray images of the lower fibula exist in every hospital with a trauma or orthopedic unit. As postoperative complications can potentially lead to long-term impairments [18], further research taking into account the enormous data amounts certainly leads to improved patient safety.

Text mining (commonly used term to denote the task of NLP) [6] techniques for radiological reports have been previously proposed to support the detection and surveillance of various diseases, including bone fractures [5, 21–23]. The aim of this study was to find the best feature extraction method for free-text radiological reports and to classify reports of ankle X-rays by fractures of the distal fibula.

Materials and methods

This retrospective, IRB-approved study was performed between 02/2019 and 01/2020. We assessed de-identified free-text radiological reports of ankle X-rays in two planes of patients treated at Hannover Medical School, between 01/2015 and 09/2019.

Training dataset

Due to a lack of existing data, we established a novel German language report dataset. A designated search engine based on the Enterprise Clinical Research Data Warehouse of the Hannover Medical School comprising pseudonymized clinical data of > 2.3 million patients was used to identify radiographs of the ankle. Data was used exclusively from inpatients who consented to the usage of their data for research purposes. The search was conducted using the search term "OSG in 2 Ebenen" (ankle X-ray in two planes). A radiologist manually assigned class labels to 3268 reports according to whether the report described a fracture of the distal fibula or not. Reports were excluded if no statement about the distal fibula was made. Only texts directly reporting on the presence (e. g., "dislocated fracture of the distal fibula") or absence (e. g., "no fracture of the distal fibula") were included in the training dataset. Reports describing tibial involvement (bi- or trimalleolar fractures), other fractures, and combined reports covering X-rays beyond the ankle were included in the analysis. Another dataset containing 400,000 radiology reports was used to train the Doc2Vec models (see below).

For the freely available dataset (link: https://doi.org/10.26068/mhhrpm/20230208-000), a further de-identification step was manually performed to displace names of patients or doctors and dates, if applicable.

Pre-processing

As classification is performed on numerical data, the first steps of ML on the texts were cleaning, normalizing, and pre-processing the data, which transformed text into machine-readable numerical vectors (▶ **Fig. 1**). We used the nltk stopword list to remove stopwords and a self-programmed script to remove HTML tags. Since stemming of German words and clinically used abbreviations resulted in a different literal sense and thus negatively impacted the AUC, we decided not to use a stemmer. Furthermore, we removed the words "nicht", "viel", and "sehr" (engl. "not", "much", "very") from the stopword list.

The feature extraction methods bag-of-words (BOW), term frequency-inverse document frequency (TF-IDF), principal component analysis (PCA), non-negative matrix factorization (NMF), latent Dirichlet allocation (LDA), and Doc2Vec were used for pre-processing. BOW is the easiest and most commonly used method for text representation [24], but TF-DIF is a robust and common method in pre-processing as well. Since they count the frequency of occurrences in a text, both techniques transform text data into very high-dimensional vectors. NMF, PCA, and LDA are methods for dimensionality reduction. PCA is one of the most commonly used methods in the basic literature [25], leading to solid results. Simply put, PCA reduces a dataset of potentially correlated features to a set of values that are linearly uncorrelated. NMF is an easily interpretable linear technique that is robust for word and vocabulary recognition while compressing original text into smaller data vectors. LDA is popular in topic modeling, where the main topics in a text are extracted and classified [26]. Doc2Vec is a method that uses Deep Learning (a technique based on neural networks (NN)) to train a model that not only transforms text into vectors, it also models how similar these texts are. The various methods were compared by accuracy (acc) and area under the curve (AUC).

Supervised learning

The pre-processed data were randomly divided into training and test datasets, with a validation dataset for the neural network in order to avoid overfitting the data and, subsequently, more reliable results. During training, the algorithms never came into contact with the test data. It was kept separate for the evaluation of the trained algorithms on unknown data. Three different ML algorithms were trained on the resulting feature vectors: NN, SVM, and LR. The algorithms were optimized for AUC and evaluated with 10-fold cross-validation on the training dataset.

Results

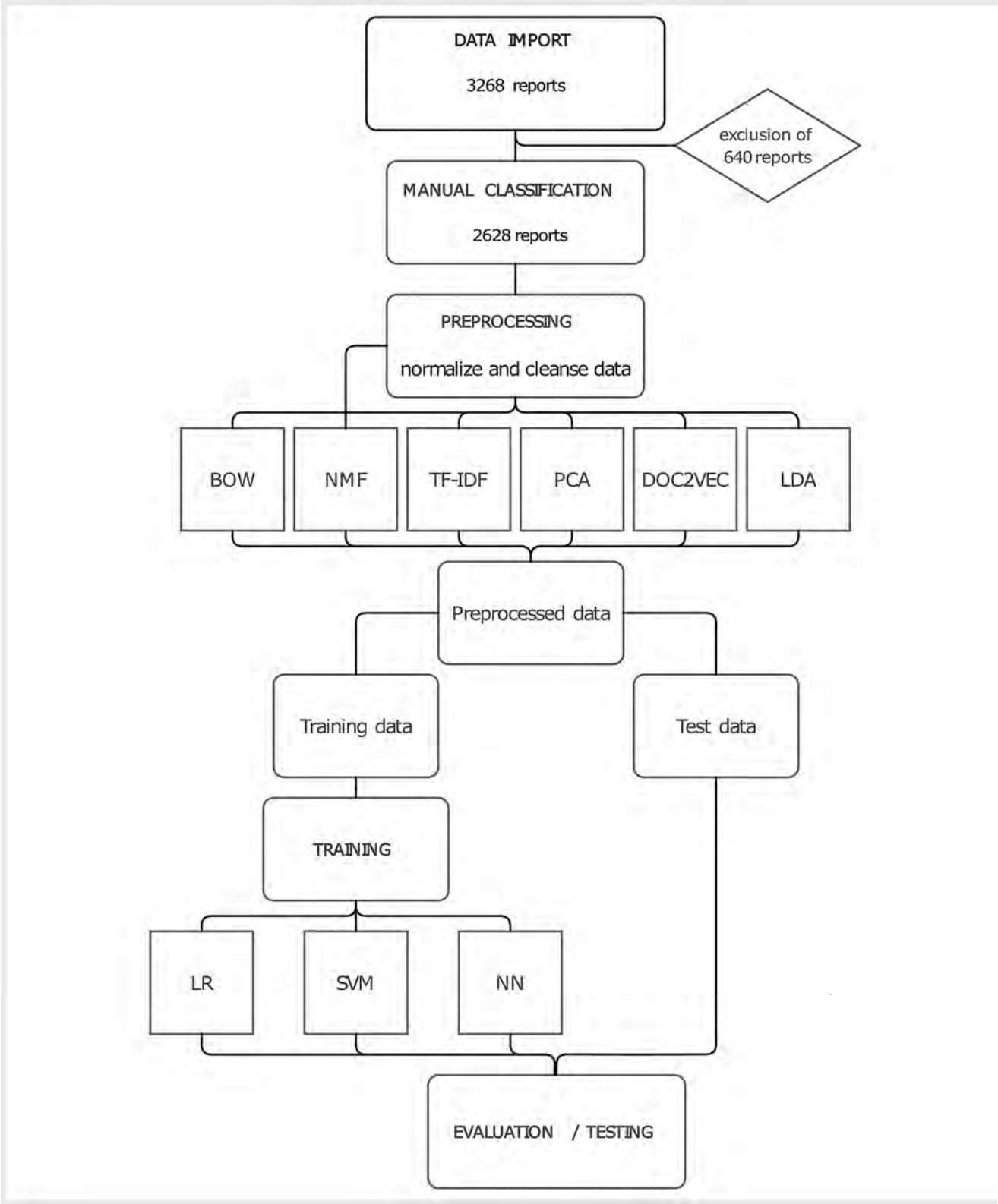
Training dataset

We assessed 3268 unstructured radiological reports of two-plane ankle X-rays. 640 reports were excluded, as they did not directly report on the distal fibula, thus it could not be defined whether a distal fibula fracture is present or not. The remaining 2628 free-text reports were included in the training dataset. Of those, 41 % (1076) described a fracture of the distal fibula. 59 % of the reports (1552) stated that no fracture of the distal fibula was present. The free-text reports were short in length, containing a median of 646 (interquartile range (IQR) 514–824) characters.

Due to the open data initiative for research transparency, the dataset is published under the following link: https://doi.org/10.26068/mhhrpm/20230208-000.

Machine Learning

Six feature extraction methods (BOW, TF-IDF, PCA, NMF, LDA, Doc2Vec) were used to train three different ML algorithms (NN, SVM, and LR) and optimized for the AUC. The trained networks



▶ **Fig. 1** Machine learning workflow in this study. BOW: bag-of-words; NMF: non-negative matrix factorization; TF-IDF: term frequency-inverse document frequency; PCA: principal component analysis; LDA: latent Dirichlet allocation; LR: logistic regression; SVM: vector machines; NN: neuronal networks.

were used to predict the label of the test data and reached the AUC. The BOW model achieved the best results (AUC: NN 0.99; SVM 0.97; LR 0.97), closely followed by the TF-IDF (AUC: NN 0.99; SVM 0.96; LR 0.96). In combination with NN, NMF achieved similar results (AUC 0.98). For details, refer to ▶ **Table 1** (AUC data) and ▶ **Table 2** (Accuracy data).

Discussion

In this manuscript, we describe our approach to classify unstructured radiograph reports according to fractures of the distal fibula. Special attention was paid to various feature extraction methods for pre-processing. To do so, we created a manually labeled novel German language report dataset, which is not yet available across the German medical NLP landscape in this format and is specifically based on radiological findings. We invite other groups to use our dataset, which is available as open data (link: https://doi.org/10.26068/mhhrpm/20230208-000).

▶ **Table 1** Overview table of AUC values of various feature extraction methods used to train different ML algorithms and evaluated with 10-fold cross-validation on the training dataset.
BOW: bag of words; LDA: Latent Dirichlet allocation; LR: Logistic regression; NMF: Non-negative matrix factorization; NN: Neural network; PCA: Principal component analysis; SVM: Support vector machine; TF-IDF: Term frequency-inverse document frequency.

	NN	SVM	LR	Average AUC
Dummy	0.5			
BOW	0.99	0.97	0.97	0.977
TF-IDF	0.99	0.96	0.96	0.970
NMF	0.98	0.9	0.9	0.927
PCA	0.95	0.91	0.9	0.920
LDA	0.94	0.89	0.88	0.903
Doc2Vec	0.94	0.9	0.85	0.897

▶ **Table 2** Overview table of accuracy values of various feature extraction methods used to train different ML algorithms and evaluated with 10-fold cross-validation on the training dataset.
BOW: bag of words; LDA: Latent Dirichlet allocation; LR: Logistic regression; NMF: Non-negative matrix factorization; NN: Neural network; PCA: Principal component analysis; SVM: Support vector machine; TF-IDF: Term frequency-inverse document frequency.

	NN	SVM	LR	Average Accuracy
Dummy	0.5			
BOW	0.96	0.97	0.97	0.967
TF-IDF	0.95	0.96	0.97	0.96
NMF	0.94	0.91	0.9	0.917
PCA	0.91	0.9	0.9	0.903
LDA	0.88	0.89	0.88	0.883
Doc2Vec	0.87	0.9	0.86	0.877

Our automated classification pipeline was able to reliably detect findings of fractures of the distal fibula. BOW was the most reliable feature extraction method for the tested models in combination with the aforementioned dataset. TF-IDF achieved AUC values very similar to BOW. TF-IDF is characterized by a lower number of dimensions. However, this does not confer a relevant advantage, as the employed models (especially neural networks) can reliably compute high dimensional data as provided by methods like BOW. Non-negative matrix factorization (NMF) proved to be a solid alternative method for producing vectors with lower dimensions. In conjunction with the supervised learning method NN, the results of NMF achieved AUC values similar to BOW and TF-IDF. The selection of an appropriate feature extraction method for pre-processing significantly impacted the results of the machine learning model – meaning that, in our tests, the best classification method could not compensate for an ill-suited feature extraction method. In this study, the choice of document representation for pre-processing of the data might be more important than the classifiers for ML-part.

In various studies, open-source datasets in English were used to compare innovative feature extraction methods to established techniques. Kim et al. e. g., performed a comparison of BOW, doc2vec, TF-IDF, and a self-made text representation method (bag-of-concepts). Contrary to our results, doc2vec showed the best results, and TF-IDF outscored BOW [27]. In contrast to our study, Kim et al. classified non-medical texts. Similar results were presented in a study comparing TF-IDF, LDA, and doc2vec for several datasets, of which one was EHR-based [28]. Doc2vec showed the best results, LDA and TF-IDF were on par. However, there is limited comparability to our study, as medical and non-medical texts were not separately analyzed. Furthermore, in our study, Doc2vec was trained on a specific sort of medical texts (radiologic reports), which might lead to a lack of diversity of informational content. This might imply that text representation methods need to be designated to the type of text. However, further research is necessary to substantiate this hypothesis.

For further studies, it could be interesting to evaluate the impact of the inclusion of various medical texts on the results. A suitable dataset to validate (or refute) our results in future studies might be a German preprint dataset published by Borchert et al. [29], which was not available at the time of our analysis.

Large transformer-based language models for the medical domain, such as BioBERT and ClinicalBERT, did not apply to our dataset, as they target the English language specifically. Currently, this type of model is not publicly available for the German language in the radiological domain. However, we see the potential of this development and are contributing our anonymized dataset of German clinical notes as open access.

Conclusion

The future of improved patient care relies on the utilization of big data. The health sector has experienced widespread digitalization during the last years, which has led to a continuously growing

amount of patient data. As radiology was among the first specialties for which computerization became obligatory for daily work, it is widely digitized. Therefore, a significant amount of data is digitally stored in radiologic reports. Unfortunately, they mostly contain unstructured text. This is a major obstacle for rapid extraction and subsequent use of information by clinicians and researchers [6]. As a result, radiology reports are often used only once by the clinician who ordered the study and are rarely used again [8].

ML information extraction techniques provide an effective method to automatically identify and classify free-text radiology reports, which can be useful in various clinical and non-clinical settings. An automated classification can support diagnostic surveillance, e. g., assist in the management of cases that require follow-up or even monitor public health-related trends such as increases in disease activity in a hospital or on a population level. Moreover, it can support cohort building for epidemiologic studies and also provide query-based case retrieval.

This study shows that automated classification of unstructured reports of radiographs of the ankle can reliably detect findings of fractures of the distal fibula. Special attention was paid to various methods for pre-processing, and it was shown that a particularly suitable feature extraction method is the BOW model for our setting. This automated classification system can serve as a reference for future studies as well as decision-support systems, which might prospectively improve clinical management and patient safety.

Limitations

It needs to be emphasized that the comparability between the mentioned studies is limited due to the varying pipeline setups and the used datasets. Contrary to the discussed studies, our dataset was German, which might impact the results. Furthermore, this project was narrowly focused on extracting a single type of information – presence or absence of a fracture of the distal fibula. Information on other fractures or pathologies was not extracted. We set up a binary classification system, which did not classify the fractures into different subclasses. Furthermore, it needs to be assessed whether the classification system can reliably be used for other radiology reports.

Regarding the dataset, although the exam description should be "OSG in 2 Ebenen", we cannot guarantee that the search term is exhaustive. Lastly, the achieved results might be over-adapted to the training dataset, which is a common problem in ML. To rule this out, the system will be validated with an unknown dataset.

CLINICAL RELEVANCE

- Text mining techniques have the potential to support the detection and surveillance of diseases.
- In this manuscript, we describe our approach to automatically classify unstructured radiograph reports according to fractures of the distal fibula.
- Our automated classification system as well as the enclosed dataset might serve as a reference for future studies as well as decision-support systems, which could potentially improve clinical management and patient safety.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Hersh WR, Weiner MG, Embi PJ et al. Caveats for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research. Med Care 2013; 51 (8): S30–S37

[2] Smith M, Saunders R, Stuckhardt L et al. Best care at lower cost. National Academies Press; 2014.

[3] Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med 2010; 2 (57): 57cm29

[4] Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. New England Journal of Medicine 2010; 363 (6): 501–504

[5] Grundmeier RW, Masino AJ, Casper TC et al. Identification of long bone fractures in radiology reports using natural language processing to support healthcare quality improvement. Applied clinical informatics 2016; 7 (4): 1051

[6] Pons E, Braun LM, Hunink MM et al. Natural language processing in radiology: a systematic review. Radiology 2016; 279 (2): 329–343

[7] Gerbel S, Laser H, Schönfeld N et al. The Hannover Medical School Enterprise Clinical Research Data Warehouse: 5 Years of Experience. In: International Conference on Data Integration in the Life Sciences. Springer; 2018: 182–194

[8] Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. Artificial intelligence in medicine 2016; 66: 29–39

[9] Reddy CK, Aggarwal CC. Healthcare data analytics. Vol. 36. CRC Press; 2015.

[10] Hearst MA. Untangling text data mining. In: Proceedings of the 37th Annual meeting of the Association for Computational Linguistics.1999: 3–10

[11] Rajkomar A, Oren E, Chen K et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine 2018; 1 (1): 1–10

[12] Devlin J, Chang MW, Lee K et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 2018. [cited 2022 Oct 17]; Available from: https://arxiv.org/abs/1810.04805

[13] Lee J, Yoon W, Kim S et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Wren J, editor. Bioinformatics. 2019 Sep 10;btz682.

[14] Huang K, Altosaar J, Ranganath R. ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. 2019. [cited 2022 Oct 17]; Available from: https://arxiv.org/abs/1904.05342

[15] Yamamoto Y, Saito A, Tateishi A et al. Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach. Scientific reports 2017; 7 (1): 1–12

[16] Christodoulou E, Ma J, Collins GS et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology 2019; 110: 12–22

[17] Gougoulias N, Sakellariou A. Ankle Fractures. In: Bentley G European Surgical Orthopaedics and Traumatology: The EFORT Textbook [Internet]. Berlin, Heidelberg: Springer; 2014: 3735–3765 [cited 2021 Mar 19]. Available from: doi:10.1007/978-3-642-34746-7_152

[18] Hasselman CT, Vogt MT, Stone KL et al. Foot and Ankle Fractures in Elderly White Women: Incidence and Risk Factors. JBJS 2003; 85 (5): 820–824

[19] Knutsen AR, Sangiorgio SN, Liu C et al. Distal fibula fracture fixation: Biomechanical evaluation of three different fixation implants. Foot and Ankle Surgery 2016; 22 (4): 278–285

[20] Neumann MV, Strohm PC, Reising K et al. Complications after surgical management of distal lower leg fractures. Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine 2016; 24 (1): 146

[21] Zuccon G, Wagholikar AS, Nguyen AN et al. Automatic classification of free-text radiology reports to identify limb fractures using machine learning and the snomed ct ontology. AMIA Summits on Translational Science Proceedings 2013; 2013: 300

[22] de Bruijn B, Cranney A, O'Donnell S et al. Identifying wrist fracture patients with high accuracy by automatic categorization of X-ray reports. Journal of the American Medical Informatics Association 2006; 13 (6): 696–698

[23] Do BH, Wu AS, Maley J et al. Automatic retrieval of bone fracture knowledge using natural language processing. Journal of digital imaging 2013; 26 (4): 709–713

[24] Zhixiang X, Chen M, Weinberger K et al. An alternative text representation to TF-IDF and Bag-of-Words [Internet]. arXiv; 2013 [cited 2023 Jan 22]. Available from: http://arxiv.org/abs/1301.6770

[25] Deisenroth MP, Faisal AA, Ong CS. Dimensionality Reduction and Principal Component Analysis. Math. Mach. Learn. Vol. 80 2018: 314–344

[26] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research 2003; 3: 993–1022

[27] Kim HK, Kim H, Cho S. Bag-of-concepts: Comprehending document representation through clustering words in distributed representation. Neurocomputing 2017; 266: 336–352

[28] Kim D, Seo D, Cho S et al. Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. Information Sciences 2019; 477: 15–29

[29] Borchert F, Lohr C, Modersohn L et al. GGPONC: A Corpus of German Medical Text with Rich Metadata Based on Clinical Practice Guidelines [Internet]. arXiv; 2020 [cited 2023 Jan 22]. Available from: http://arxiv.org/abs/2007.06400