International Journal of Medical Informatics 161 (2022) 104724

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Deep Learning-based detection of psychiatric attributes from German mental health records

Sumit Madan [a,b,*], Fabian Julius Zimmer [c], Helena Balabin [a], Sebastian Schaaf [d,1], Holger Fröhlich [a,e], Juliane Fluck [f,g,1], Irene Neuner [c], Klaus Mathiak [c], Martin Hofmann-Apitius [a,e], Pegah Sarkheil [c,h,*]

[a] Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53757 Sankt Augustin, Germany
[b] Institute of Computer Science, University of Bonn, 53113 Bonn, Germany
[c] Department of Psychiatry, Psychotherapy and Psychosomatics, Faculty of Medicine, RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany
[d] HPC and Scientific Computing, German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany
[e] Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, 53113 Bonn, Germany
[f] Knowledge Management, ZB MED – Information Centre for Life Sciences, 50931 Cologne, Germany
[g] The Agricultural Faculty, University of Bonn, 53115 Bonn, Germany
[h] Department of Psychiatry and Psychotherapy, University of Münster, 48149 Münster, Germany

A R T I C L E I N F O

Keywords:
Electrical Health Records
Mental State Examination
Clinical Text Mining
Deep Learning, AMDP

A B S T R A C T

Background: Health care records provide large amounts of data with real-world and longitudinal aspects, which is advantageous for predictive analyses and improvements in personalized medicine. Text-based records are a main source of information in mental health. Therefore, application of text mining to the electronic health records – especially mental state examination – is a key approach for detection of psychiatric disease phenotypes that relate to treatment outcomes.
Methods: We focused on the mental state examination (MSE) in the patients' discharge summaries as the key part of the psychiatric records. We prepared a sample of 150 text documents that we manually annotated for psychiatric attributes and symptoms. These documents were further divided into training and test sets. We designed and implemented a system to detect the psychiatric attributes automatically and linked the pathologically assessed attributes to AMDP terminology. This workflow uses a pre-trained neural network model, which is fine-tuned on the training set, and validated on the independent test set. Furthermore, a traditional NLP and rule-based component linked the recognized mentions to AMDP terminology. In a further step, we applied the system on a larger clinical dataset of 510 patients to extract their symptoms.
Results: The system identified the psychiatric attributes as well as their assessment (normal and pathological) and linked these entities to the AMDP terminology with an F_1-score of 86% and 91% on an independent test set, respectively.
Conclusion: The development of the current text mining system and the results highlight the feasibility of text mining methods applied to MSE in electronic mental health care reports. Our findings pave the way for the secondary use of routine data in the field of mental health, facilitating further clinical data analyses.

1. Introduction

Beside the common advantages of clinical routine data like availability and cost-effectiveness, the use of routine data in mental health research has additional values regarding the longitudinal information. Because mental health conditions are highly dependent on dynamic brain-related processes like developmental, adaptive, and degenerative changes, follow-ups and reevaluations over an extended period can provide essential information in understanding the psychopathological aspects of the diseases. Retrospective studies based on large electronic

* Corresponding authors at: Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing SCAI, Schloss Birlinghoven, 53757 Sankt Augustin, Germany (S. Madan). Department of Psychiatry and Psychotherapy, University of Münster, 48149 Münster, Germany (P. Sarkheil).
E-mail addresses: sumit.madan@scai.fraunhofer.de, sumit.madan@gmx.de (S. Madan).
[1] These authors worked at 1 during conduction of the study.

https://doi.org/10.1016/j.ijmedinf.2022.104724
Received 24 November 2021; Received in revised form 7 February 2022; Accepted 18 February 2022
Available online 22 February 2022

collections of mental health records can be an attractive source of longitudinal information. To enable the use of unstructured clinical routine data in mental health research, considerable work needs to be done to extract structured information, which involves coding of free-text reports through text mining techniques.

Commonly, a psychiatric clinical examination consists of an interview regarding the past and present symptoms and observing the current pathological signs. Mental status exam (MSE) [1] is a standardized form of examination with methods for observing and describing the mental state and behaviors of each patient, based on both objective observations of the clinician and subjective descriptions given by the patient herself. To capture the information related to MSE, a medical expert examines the patient for any possible signs or symptoms of a psychiatric condition and provides documentation as a part of the patient's medical record. The MSE has a great potential to be a main information source to extract phenotype information from the electronic health records [2]. Text documentation of MSE follows a specified form and semantics [3] that intends to facilitate rapid medical communication and training (Fig. 1). To identify the entities that contain the main relevant and classifiable information, advanced text mining approaches are needed as MSE documentation often contains incomplete sentences, abbreviations, acronyms, negations. Also, the documentation is very individual and often contains vague or uncertain expressions.

Mapping the MSE documentation to a reference system is an important step for generation of the comparable results. For relating the mentioned symptoms to the standard psychopathological concepts, we relied on the assessment based on the Association for Methodology and Documentation in Psychiatry (AMDP) system [4], which has been developed for the standardized assessment of mental state. The AMDP system, which has been internationally recognized and translated into many languages (English, French, German, Italian, Portuguese, etc.), represents a terminology of psychopathological symptoms and their rating. It contains a short definition for each symptom, notes on the severity (mild, medium, severe), and a list of distinct examples. The symptoms listed by the AMDP system (in total 140 features) are numbered and grouped together in the AMDP manual. Altogether, AMDP can serve as a terminology for the MSE.

Researchers have proposed rule and machine learning-based text mining methods to extract various kind of information from electronic health records (EHR). Barak-Corren et al. [5] extracted demographic characteristics, diagnostic codes, laboratory results and prescribed medications from English EHRs to predict suicidal behavior. Hazewinkel et al. [6] have analyzed textual data included in notes and reports of Dutch EHRs of patients that were admitted in a psychiatric hospital in The Netherlands. Clinical notes in English language of psychiatry wards from Mayo Clinic were utilized by Sohn et al. [7] to detect drug side effects using a rule-based approach. Named entity recognition (NER) has been particularly popular in mental health care in identification of the relevant concepts from the clinical records [8,9]. It has been already applied to identify predictors of suicide from EHR [8] and social media [9].

Recently, transfer learning-based methods such as Bidirectional Encoder Representations from Transformers (BERT) have gained a lot of attention [10]. Briefly, transfer learning rests on the idea that pre-trained word embeddings [11] learned from large amounts of training data (e.g., Wikipedia articles) using deep learning models already contain a significant amount of information that is relevant for more specific downstream tasks, including NER in clinical documents. Lee et al. [12] published BioBERT which is additionally trained on PubMed and PubMed Central articles, achieving state-of-the-art results in several biomedical natural language processing (NLP) tasks. Similarly, for the clinical domain, Alsentzer et al. [13] have created Clinical BERT embeddings by performing a pre-training on 2 million freely available clinical notes [14] using the BERT architecture [10]. To our knowledge no transformers-based language model has yet been applied to German clinical data. An extended description of the related work is included in Supplementary.

In this work, we introduce a text mining approach to extract key clinical information from MSE documents. As a first step, we extracted MSE containing documents from the clinical information system and prepared a manually annotated dataset. The text mining system was implemented as a two-step procedure for 1) deep learning-based recognition of relevant entities, such as psychological assessments (NER), and 2) mapping to the standard AMDP terminology (entity linking). For NER, we fine-tuned a freely available deep learning-based general language understanding model, so called GermanBERT [15], that is pre-trained on German textual content. Additionally, we evaluated our text mining system thoroughly based on an independent test set and demonstrate the promising prediction performance. Finally, we applied our workflow to identify psychopathological symptoms from an enhanced set of psychiatric patient data. We also self-assess the quality of our medical AI work that employs medical data using the IJMEDI checklist [16] (included in Supplementary).

2. Materials and methods

2.1. Study data

2.1.1. Sample selection

We selected a set of MSE texts from discharge summaries, which are issued when or after the patient leaves the care of the hospital as the primary communication mechanism between hospitals and other healthcare providers. More than 30.000 German documents of this kind are available in the electronic archives of the Department of Psychiatry and Psychotherapy, University Hospital Aachen (UKAachen). For the current study, MSE sections of 660 patients were isolated from the discharge summaries of (pseudo-)randomly selected patients from various ranges of mental disorders, who received treatment in the inpatient services of UKAachen between 2014 and 2019. Patients' identification information (such as names, gender) were removed from all study data. The study data consists of two different datasets – an annotated dataset used for system training and evaluation, and an additional unlabeled dataset for later system application. Note that the additional dataset has no label annotations and thus cannot be used for model evaluation purposes. The dataset for training consisted of 100 documents, which were randomly chosen out of the 150 annotated documents. The remaining 50 annotated documents were used as independent test data.

Table 1 shows the demographics of the patient collective (n = 150) of the dataset for system training and evaluation in which 75 (50%) were

"The patient was conscious. He was oriented to time, place and person. In social contact he was friendly and cooperative. Speech monotonous, no disturbance of attention and memory, concentration disturbed. He exhibited loosening of associations and flight of ideas, no compulsions, no anxiety. Delusions were not observed. He reported auditory hallucinations. Affect labile, mood depressed, partly restless, no agitation, no aggressive behavior. He denied suicidal ideas."

Fig. 1. An exemplary MSE report.

Table 1
Demographics of the patient collective of two different datasets for system training and evaluation, and for system application.

	Dataset for system training and evaluation	Dataset for system application
Total	n = 150 (training = 100 and test = 50)	n = 510
Gender	75 female, 75 male	223 female, 287 male
Age at date of discharge (years)	44.52 (mean) 17.56 (standard deviation)	48.66 (mean) 12.08 (standard deviation)
Retrospective time span	2014 – 2018	2017 – 2019

Table 2
Total number of manual annotations for each class appearing in the annotated set of 150 documents.

Class	Total annotations
Attribute	3,423
NormalAssessment	1,734
PathologicalAssessment	1,302
AMDP concept	1,276

female and 75 (50%) were male. The sample is further characterized by the categories of mental disorders encoded with ICD-10 [17] diagnoses, since various disorders are expected to be differential in MSE outcomes (Supplementary Table S1). Whereas the additional dataset of anonymized, unannotated psychiatric discharge summaries from years 2017 to 2019 was used to predict the patients' symptoms by applying the developed system. In total, we extracted 510 MSEs (170 MSEs for each of the three categories; see Supplementary Table S1) from discharge summaries.

2.2. Data annotation

To build the gold standard, MSEs have been tagged by a medical expert with the following label types: 1) Attribute: assessed components, 2) NormalAssessment (related to a component), and 3) PathologicalAassessment (related to a component) and further verified by a board-certified psychiatrist. Fig. 2 shows exemplary excerpts of two MSEs. Furthermore, the phrases that have been labeled with the type *PathologicalAssessment* in combination with their related attributes were mapped to the AMDP terminology (Fig. 3). The mapping was performed only for this label type, because the AMDP terminology only covers the pathological mental states. The annotation guidelines are included in Supplementary.

2.3. Ethics and data protection

The data used for this retrospective research is considered as "real-world data" collected during the primary mental health care in hospital. The use of patients' data for the current research was approved by "the Medical Ethics Committee" at the RWTH Aachen Faculty of Medicine (EK 349/20).

3. Methodology

In order to automatically detect entities in MSEs, we employed the widely used approach of fine-tuning a pre-trained language model on a given task-specific dataset. One of the most well-known deep learning models used for this type of approach is Bidirectional Encoder Representations from Transformers (BERT) [10]. BERT fully relies on using attention functions [18] to learn token embeddings. The original BERT model is limited to the English language. However, recently, a version for the German language (GermanBERT [15]) has been made publicly available. Following the example of the BERT transfer learning approach, GermanBERT [15] forms the equivalent adaptation of the language model to the German language domain. Using the same hyperparameters as the original $\text{BERT}_{\text{BASE}}$ [10] model, the GermanBERT [15] model was trained from scratch on the German Wikipedia dump, the OpenLegalData [19] dump and German news articles. In this work, we further fine-tuned GermanBERT on our study data to recognize the defined entity classes. Once the model is fine-tuned on the MSEs, it can be used to predict new labels on other unseen examinations. Lastly, the identified entities are mapped / normalized to the best matching AMDP concepts. A summary of the overall workflow is shown in Fig. 4.

3.1. Preprocessing datasets for Fine-Tuning process

For NER, the offsets of the annotated entities, which represent the actual position in text, were first converted into an inside-outside-beginning (IOB) format [20] (see Fig. 5). Overall, there were a total of nine labels present in the token classification setting. Seven of them represented the IOB scheme, namely one outside class, together with the beginning and inside labels for each of the three original annotation classes. Additionally, there were two more labels for padding and tagging sub words of a labelled entity. Afterwards, the documents were segmented into single sentences. Only sentences containing entities were considered for fine-tuning GermanBERT.

3.2. Fine-tuning GermanBERT on study data for entity recognition

Based on the pre-trained GermanBERT model, a tokenizer and a language model were initialized. To adapt the general-purpose language model to the given entity recognition task, a token classification head consisting of a feed-forward and a softmax layer was added on top of the output of the GermanBERT model. More precisely, this final output layer represented each possible token label (resulting in $\text{dim}_{\text{out}} = 9$) and, additionally, was fully connected to the previous layer (of dimension 768).

After tokenization and preprocessing, the training data was used within a 5-fold cross-validation to fit and optimize the parameters of the model. The respective hyperparameters are listed in Supplementary Table S3. For assessing the performance of the fine-tuned models, we used entity-level precision, recall, and F_1-scores, both separately for each class, as well as the micro and macro averages of all classes.

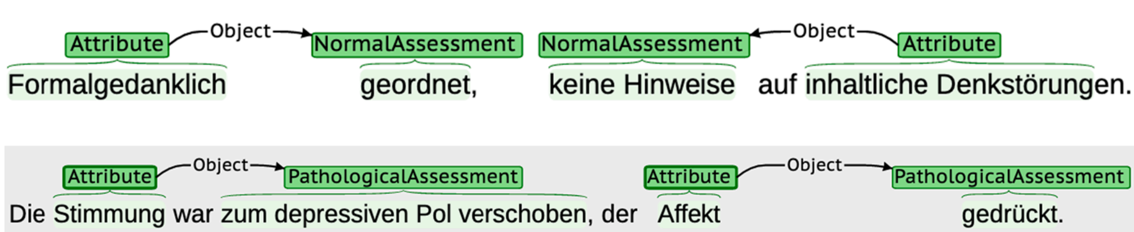

Fig. 2. Exemplary excerpts from the MSE document. The gold standard annotation of the attributes is represented by the green boxes. Each attribute of the type NormalAssessment or PathologicalAssessment is related to an annotation of the type Attribute. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

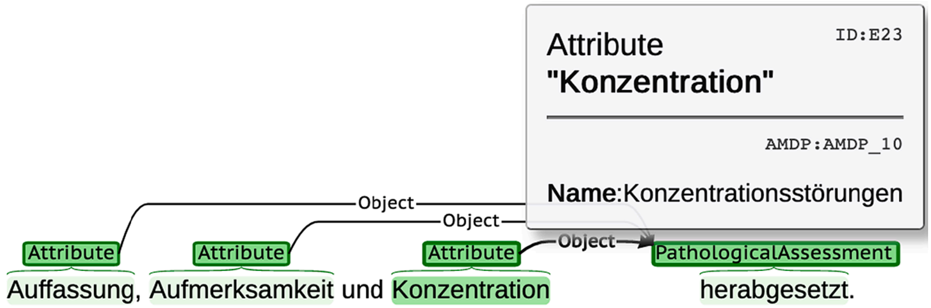

Fig. 3. An annotated sentence that relates the Attribute “Konzentration” (English: concentration) and its pathological assessment “herabgesetzt” (English: reduced) to the AMDP concept “Konzentrationsstörungen” (ID: AMDP:10) (English: concentration disturbance).

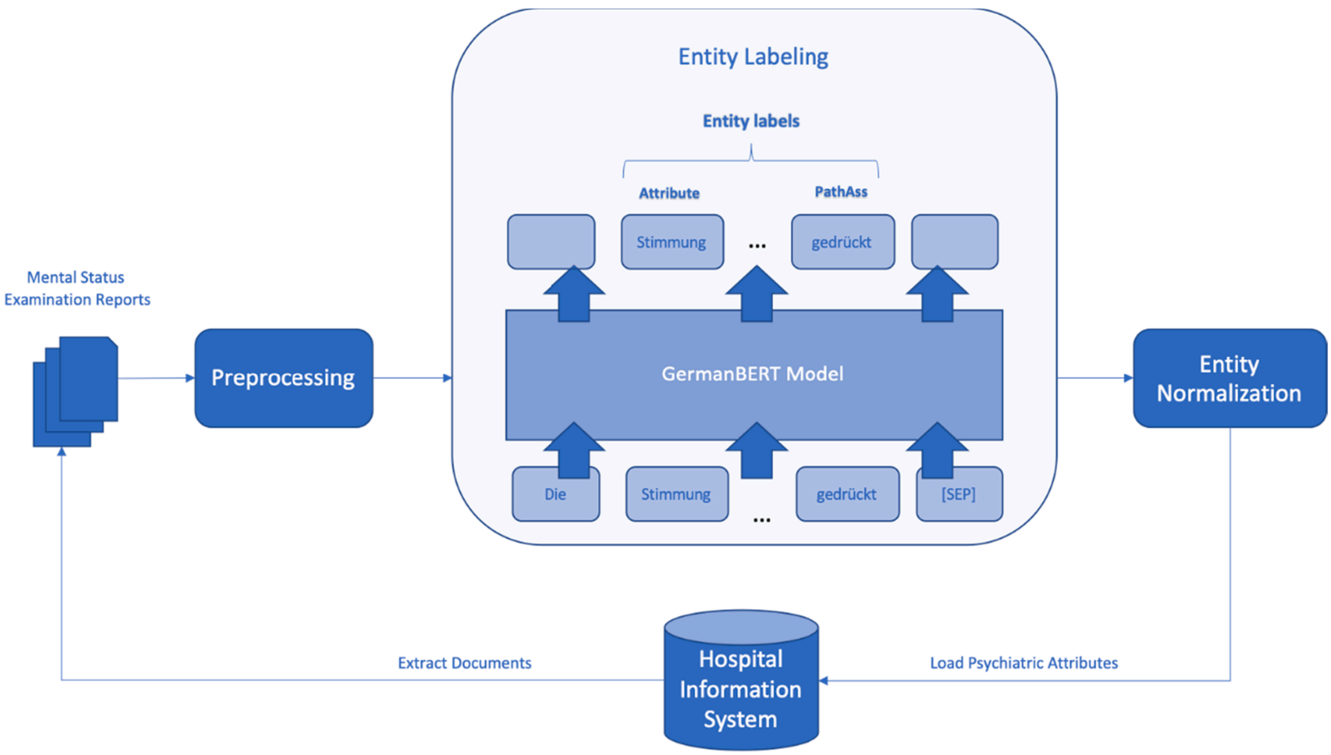

Fig. 4. Outline of the general workflow used for the analysis of MSE reports, consisting of the fine-tuning procedure of the pre-trained GermanBERT model, as well as additional pre- and post-processing steps. Sentences are used as input, as described by the (simplified) input cells in the *entity labeling* box. The entity labels are normalized in the entity normalization procedure to the AMDP terminology. The results are then further loaded in the database of the HIS.

$$\text{Precision} = \frac{\sum \textit{true positive}}{\sum \textit{true positive} + \textit{false positive}}$$

$$\textit{recall} = \frac{\sum \textit{true positive}}{\sum \textit{true positive} + \textit{false negative}}$$

$$F_1 - \textit{score} = 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}$$

3.3. Linking to AMDP terminology

Our initial empirical experiments showed that due to the small size of AMDP annotations, training of a robust machine learning (ML) model to map pathologically assessed attributes to AMDP terminology was not yet feasible. Therefore, we implemented a traditional dictionary-lookup algorithm to link the mentions to AMDP. The dictionary is derived in a multi-step process. First, we gathered all linked mentions of pathologically assessed attributes from the training data. In a second step, we concatenated the mentions of both classes to generate synonyms for the AMDP concepts appearing in training data. Next, we derived further synonyms by also considering the labels of all available AMDP concepts in the terminology themselves. Finally, we merged the synonyms and harmonized (such as lower casing, removal of special characters) them in a post-processing step.

This dictionary was used as the main resource for the matching algorithm that was used to link mentions of attribute and pathological assessment classes to AMDP terminology. After processing the documents with the ML-based named entity recognition, the detected entities were used as an input to the matching algorithm. As a first step, the sentences containing attribute and pathological assessment entities were filtered for further processing. A rule was defined to link the mentions of attributes to pathological assessments. The entities were linked to build pairwise combinations. The mentions of pairwise combinations are concatenated, which are further labeled as queries. Furthermore, a stemming approach, using the German Snowball stemmer, was applied on the synonyms of the dictionary and on the queries. Now, with the exact string match the queries were compared with the synonyms. If no

Die	O
Patienten	O
ist	O
wach	B-ATTR
,	O
bewusstseinsklar	B-ATTR
und	O
zu	B-ATTR
allen	I-ATTR
Qualitäten	I-ATTR
orientiert	B-NORM
.	O

Fig. 5. Exemplary sentence taken from the preprocessed MSE reports. B-ATTR, I-ATTR and B-NORM denote the beginning of an attribute, inside of an attribute, and beginning of a normal assessment entity, respectively. Any word that is not belonging to one of the three classes is labelled with O (meaning outside).

result was found, a fuzzy string match was performed. In our experiments, on the validation set we found a threshold of 91 as the best value for the fuzzy string-matching algorithm (see Section Implementation). This value was further used as a threshold for the final system.

3.4. Implementation

The system has been implemented with Java and Python. We use a modified version of BratReader from the DKPro Core [21] library to read the annotated documents and create a JSON document for further processing. We employed the spaCy library [22] for converting text into IOB format. The deep learning-based model for entity recognition was based on the Framework for Adapting Representation Models (FARM) [23] that internally uses the PyTorch implementation of the pre-trained model in the Transformers [24] library. The machine learning lifecycle was managed with the Mlflow [25] library that allows for logging all training and evaluation experiments as well as metrics and models. For entity normalization, German Snowball stemmer [26] integrated in pystemmer [27] was used to perform stemming of German tokens and fuzzywuzzy [28] was used to execute fuzzy string matching.

3.5. Data availability

The data that support the findings of this study are available on request from the corresponding authors, but restrictions apply to the availability of these data. The data are not publicly available due to data protection and privacy reasons as they represent sensitive patient data.

4. Results

4.1. Datasets

Data was obtained from the health care records of the Dept. of Psychiatry and Psychotherapy, RWTH Aachen University Hospital for training and evaluation of the models. A total of 150 documents were annotated manually with several classes. Table 2 contains the total number of annotations for each class. All the annotations were performed on the sentence level. In the annotated dataset 1,089 Attribute annotations can be found. These attributes have been further linked to mentions that are annotated with two classes NormalAssessment (569 entries) and PathologicalAssessment (386 entries). The pathological assessed attributes are normalized with the AMDP terminology. In total 773 AMDP concepts have been linked to 386 pathological assessed attributes.

Table 3 lists 10 top lower-cased annotations for each of the important classes that appear in the annotated dataset. The most common attributes are *wach, stimmung, konzentration, and antrieb.* The most common pathological AMDP concepts that appear in the dataset are *Dysphorisch* (eng. Dysphoric), *Affektarm (eng. Emotionless), Konzentrationsstörungen (eng. Concentration disorders), Antriebsarm (eng. Less energized).*

4.2. Entity recognition

To detect the mentions of each class, we used the pre-trained GermanBERT model and fine-tuned it on the MSE documents. For this purpose, we initially split the entire dataset into a training dataset (100 documents) and a test dataset (50 documents) at random. To identify the best possible model variant based on the training data we employed 5-fold cross-validation. Based on the performance assessed through cross-validation (detailed results are included under Section Cross-Validation Results in Supplementary), we use the optimized hyperparameters to train the final model on the whole training dataset. The generalization performance of the final model was assessed on the held-out test set of 50 documents. Table 4 presents the classification scores for each class on the test set averaged over five runs (the results of the training performance are included in Supplementary Table S2). We reached a precision of 89.0%, a recall of 87.4%, and an F_1-score of

Table 4
Precision, recall and F_1-score on the test dataset (50 documents) of the NER on MSEs task, averaged over five runs. Support column informs about the total number of instances of each class in the test dataset.

	Precision (%)	Recall (%)	F_1-score (%)	Support
Attribute	89.0 ± 0.0	87.4 ± 1.2	88.6 ± 0.8	795
Pathological Assessment	87.4 ± 2.0	85.0 ± 0.0	86.2 ± 1.0	314
Normal Assessment	86.6 ± 2.0	85.6 ± 1.2	86.2 ± 1.6	282
macro average	88.0 ± 0.0	87.0 ± 0.0	87.0 ± 0.0	1,391
micro average	88.0 ± 0.0	87.0 ± 0.0	87.0 ± 0.0	1,391

Table 3
The most frequent annotations for Attributes, NormalAssessments, PathologicalAssessments, and AMDP concepts appearing in the annotated dataset. Some of the entries (such as “wach”, eng. Alert, or “ängste”, eng. Fears) count as Attributes and normal/pathological assessments at the same time and might appear in different classes.

Attributes	(n)	NormalAssessments	(n)	PathologicalAssessments	(n)	AMDP Concepts	(n)
wach	47	kein	126	reduziert	60	Dysphorisch (AMDP:67)	67
stimmung	47	orientiert	50	gedrückt	21	Affektarm (AMDP:61)	54
konzentration	46	wach	46	ängste	17	Konzentrationsstörungen (AMDP:10)	52
antrieb	46	klar	37	verlangsamt	14	Antriebsarm (AMDP:80)	47
kontakt	41	geordnet	33	distanziert	14	Aufmerksamkeitsstörungen (AMDP:152)	44
aufmerksamkeit	40	kein anhalt	26	herabgesetzt	11	Auffassungsstörungen (AMDP:9)	31
ich-störung	40	freundlich	24	angespannt	11	Motorisch unruhig (AMDP:83)	24
suizidalität	39	gepflegt	21	depressiv	11	Affektstarr (AMDP:79)	22
affekt	36	keine hinweise	20	vermindert	10	Ängste (AMDP:153)	21
bewusstsein	36	ruhig	16	beeinträchtigt	8	Affektlabil (AMDP:77)	20

88.6% for the detection of Attribute class. The PathologicalAssessment and NormalAssessment classes are both detected with an F_1-score of 86.2%. For all three classes, the scores are significantly better than achieved through cross validating the models. In summary, the validation on an unseen test set reveals that the model is quite generalizable and robust in terms of entity detection.

4.3. Extraction of AMDP concepts

One of the goals of the current work is to delineate the pathological attributes as AMDP concepts. For this purpose, the extracted patient attributes are first linked with their assessment. If pathological, they have been mapped to the corresponding concepts from the AMDP terminology. Table 5 shows the results of the extraction of AMDP concepts by using the algorithm mentioned in Section *Normalization to AMDP* on the test set. We reached a precision of 90.0%, a recall of 92.0%, and an F_1-score of 91.0%.

4.4. System application on additional patient records

Next, we applied the developed system to the additional dataset of 510 unannotated MSEs. The system could detect 7,047 Attribute (unique: 183), 3,073 NormalAssessment (unique:67), and 2,254 PathologicalAssessment entities (unique: 157). Furthermore, the mapping to the AMDP terminology retrieved a total of 2,197 AMDP concepts (unique: 44). Table 6 provides an overview of the top 10 annotations with their associated total amounts of the available classes. Most importantly, 7 out of top 10 AMDP concepts in the annotated training dataset are identical to the results of this dataset (Table 6).

5. Discussion

Most of the clinical routine data in the mental health discipline is recorded as text documents. Therefore, text mining techniques are becoming crucial for extraction of relevant information. In this work, we present the first text mining approach to mental health data analytics in the German speaking region. We focused on MSE as the main part of the clinical evaluation of psychiatric patients and a standard for communicating the evaluation results. A pre-trained deep neural network (GermanBERT [15]) have been fine-tuned to identify relevant attributes and psychological assessments in German clinical routine data. In a further step, a method to relate the extracted information to AMDP, a standard terminology for psychopathology, have been implemented. We validated the results of the approach on an independent test set to demonstrate the robustness of the method. Finally, we applied the workflow on a larger clinical dataset, which returns a set of symptom variables for further clinical and research data analyses.

Based on the expert annotation of our dataset, consisting of 150 MSE, 90% of the MSE pathological entities could be referred to the AMDP symptom list, which confirms the efficacy of this system in normalizing the unstructured MSE in routine data. The fine-tuned model that detects various entities such as attributes and their assessment as normal or pathological reached an F_1-score of around 86% on test dataset for all entity classes, which is a quite reasonable performance. Furthermore, the mapping / normalization of the pathological symptoms to AMDP terminology achieved a high F_1-score of 91%. These promising results encourage future efforts towards automatically structuring the clinical notes from the EHRs. Our results revealed that the AMDP concept of dysphoria has been most frequently identified in the MSE reports, suggesting evaluation of dysphoria to be the focus of clinicians. Other frequent AMDP concepts include emotionality, concentration, and drive. Altogether, it can be inferred that the clinicians use the MSE as a tool to observe and assess the patient's current mental state with a focus on affective evaluations. Text mining of MSE reports in EHR might primarily inform about the affective signs and symptoms. Several psychiatric attributes are reported rarely. As for now, we have only analyzed MSE of 660 patients through our information extraction pipeline, which may not comprehensively cover more diverse psychiatric attributes. Therefore, we plan to extend the work with a broader analysis of additional MSE reports that could reveal interesting associations.

Table 5
Precision, recall, and F_1-score on the test dataset of extraction of AMDP concepts. Support column shows the number of total AMDP concepts in the dataset.

	Precision (%)	Recall (%)	F_1-score (%)	Support
AMDP concept	90.0	92.0	91.0	341

The AMDP terminology has been developed to introduce a systematic to the terminology of psychopathology. The purpose for its development was a comparable and reliable documentation of evaluation results in clinical practice and research [4]. The AMDP system offers 100 psychopathological (and 40 somatic) definable symptoms, sorted in main categories as individual entries. We suggest the AMDP system can be used as a normative reference for the identified entities in text mining of MSE documents. That mentioned, the standard clinical terminology SNOMED-CT is quite popular for documenting patient clinical information in many countries. Recently, Germany has become a new member of the SNOMED International consortium and will start to apply this terminology in clinical research and practice. We suggest that in the near future a mapping of AMDP to SNOMED-CT will be required for a consistent harmonization of mental health care data to assure the convergence of clinical interpretations and machine-readable codes. This mapping might be also indicated as SNOMED-CT is likely to code specific items, while AMDP is a comprehensive system that includes the normal findings and unmentioned attributes.

The current workflow is based on costly manual extraction of the MSE section from the discharge summaries for training data creation. To speed up and improve the workflow further approaches are needed to include automatic segmentation of the MSE section in the clinical documents. A further point for future development is identifying the severity of disease symptoms and predicting ICD-10 codes directly from the collection of symptoms. Such approaches offer a great potential for a more cost-effective coding for secondary use of clinical data, for example in scientific research or in the context of insurance claims. Extending the text mining techniques to other clinical text sections like medical history and nursing reports is a useful further step for capturing the data needed for a broad biopsychosocial phenotyping. To test and improve the generalizability of the workflow, further future research is planned by applying the workflow to documents from multiple clinical centers.

6. Conclusion

In this study, we constructed a text analysis system composed of a neural network model and a traditional NLP and rule-based analysis methods extracting mental state attributes from the psychiatric discharge summaries. Routine clinical data was used for training, test, and validation. The proposed approach achieved promising results. Automatized transforming unstructured texts into a structured format, the standard AMDP terminology, enables identification of meaningful patterns and new insights from the clinical routine data in the psychiatric discipline.

CRediT authorship contribution statement

Sumit Madan: Conceptualization, Supervision, Methodology, Software, Visualization, Validation, Writing – original draft. **Fabian Julius Zimmer:** Resources, Data curation, Validation. **Helena Balabin:** Software, Writing – original draft. **Sebastian Schaaf:** Writing – review & editing. **Holger Fröhlich:** Writing – review & editing. **Juliane Fluck:** Conceptualization, Writing – review & editing. **Irene Neuner:** Writing – review & editing. **Klaus Mathiak:** Conceptualization, Writing – review & editing. **Martin Hofmann-Apitius:** Funding acquisition, Supervision,

Table 6
Top 10 annotations of Attribute, NormalAssessment, PathologicalAssessment, and AMDP concept with their associated total amounts appearing in the additional dataset of 510 patients.

Attributes	(n)	NormalAssessments	(n)	PathologicalAssessments	(n)	AMDP Concepts	(n)
antrieb	436	kein hinweis	448	reduziert	481	Dysphorisch (AMDP:67)	334
wach	392	orientiert	402	gedrückt	169	Antriebsarm (AMDP:80)	263
konzentration	384	geordnet	235	unruhig	126	Konzentrationsstörungen (AMDP:10)	212
suizidalität	374	distanziert	227	verlangsamt	110	Affektarm (AMDP:61)	185
stimmung	364	freundlich	226	depressiv	109	Aufmerksamkeitsstörungen (AMDP:152)	167
aufmerksamkeit	359	kein anhalt	224	herabgesetzt	95	Auffassungsstörungen (AMDP:9)	149
auffassung	326	ruhig	184	vermindert	63	Affektstarr (AMDP:79)	145
zwänge	324	kein	136	angespannt	60	Motorisch unruhig (AMDP:83)	107
kontakt	319	regelrecht	130	labil	45	Psychomotorisch verlangsamt (AMDP:156)	96
fremdgefährdung	310	zugewandt	118	gesteigert	42	Verlangsamt (AMDP:16)	72

Project administration, Writing – review & editing. **Pegah Sarkheil:** Funding acquisition, Project administration, Conceptualization, Methodology, Data curation, Validation, Writing – original draft.

Acknowledgements

S.M., F.J.Z, M.HA., and P.S's work was supported and funded from the EU's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2 by the HBP Medical Informatics Platform). S.M., H.B., S.S., and J.F.'s work was performed in context of the project IDSN that is supported by the German Federal Ministry of Education and Research (BMBF) as part of the program "i:Dsem – Integrative Data Semantics in the Systems Medicine", project number 031L0029 [A-C]. We would like to acknowledge the organization AMDP e.V. (Die Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie) for the permission to use AMDP terminology for this research.

Summary Points.

"What was already known on the topic":

- Application of routine data in mental health investigations has been very limited to date.
- The major obstacle has been the unstructured information, mainly as text-based documents.
- Hazewinkel et al. [6] and Sohn et al. [7] have analyzed Dutch and English EHRs from psychiatry to obtain frequently used concepts and drug side effects.
- To our knowledge, extraction of psychiatric attributes and symptoms has not been scientifically explored yet.

"What this study added to our knowledge":

- We propose a text mining system for extraction of the relevant information from the mental health examination records, which we evaluated on an independent dataset.
- A deep-learning approach has been applied to identify the relevant attributes in the mental state examination records.
- We created a system to link the attributes to the AMDP standard terminology to semantically enhance the data and make it interoperable.
- We achieved encouraging results and demonstrated the feasibility of using text mining methods to extract relevant information from patient data, which can be used in future for mental health research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2022.104724.

References

[1] J.J. Silverman, M. Galanter, M. Jackson-Triche, D.G. Jacobs, J.W. Lomax, M. B. Riba, L.D. Tong, K.E. Watkins, L.J. Fochtmann, R.S. Rhoads, Practice Guidelines for the Psychiatric Evaluation of Adults, Third Edition, American Psychiatric Association (2016), https://doi.org/10.1176/appi.books.9780890426760.
[2] D.C. Martin, The Mental Status Examination, in: H.K. Walker, W.D. Hall, J.W. Hurst (Eds.), Clinical Methods: The History, Physical, and Laboratory Examinations, 3rd ed., Butterworths, Boston, 1990. http://www.ncbi.nlm.nih.gov/books/NBK320/ (accessed February 9, 2021).
[3] M. Soltan, J. Girguis, How to approach the mental state examination, BMJ. 357 (2017), j1821, https://doi.org/10.1136/sbmj.j1821.
[4] R.-D. Stieglitz, A. Haug, E. Fähndrich, M. Rösler, W. Trabert, Comprehensive Psychopathological Assessment Based on the Association for Methodology and Documentation in Psychiatry (AMDP) System: Development, Methodological Foundation, Application in Clinical Routine, and Research, Front Psychiatry. 8 (2017), https://doi.org/10.3389/fpsyt.2017.00045.
[5] Y. Barak-Corren, V.M. Castro, S. Javitt, A.G. Hoffnagle, Y. Dai, R.H. Perlis, M. K. Nock, J.W. Smoller, B.Y. Reis, Predicting Suicidal Behavior From Longitudinal Electronic Health Records, AJP. 174 (2) (2017) 154–162, https://doi.org/10.1176/appi.ajp.2016.16010077.
[6] M.C. Hazewinkel, R.F.P. de Winter, R.W. van Est, D. van Hyfte, D. Wijnschenk, N. Miedema, E. Hoencamp, Text Analysis of Electronic Medical Records to Predict Seclusion in Psychiatric Wards: Proof of Concept, Front. Psychiatry. 10 (2019), https://doi.org/10.3389/fpsyt.2019.00188.
[7] S. Sohn, J.-P.-A. Kocher, C.G. Chute, G.K. Savova, Drug side effect extraction from clinical narratives of psychiatry and psychology patients, J Am Med Inform Assoc. 18 (2011) i144–i149, https://doi.org/10.1136/amiajnl-2011-000351.
[8] M. Senior, M. Burghart, R. Yu, A. Kormilitzin, Q. Liu, N. Vaci, A. Nevado-Holgado, S. Pandit, J. Zlodre, S. Fazel, Identifying Predictors of Suicide in Severe Mental Illness: A Feasibility Study of a Clinical Prediction Rule (Oxford Mental Illness and Suicide Tool or OxMIS), Front Psychiatry. 11 (2020) 268, https://doi.org/10.3389/fpsyt.2020.00268.
[9] J. Du, Y. Zhang, J. Luo, Y. Jia, Q. Wei, C. Tao, H. Xu, Extracting psychiatric stressors for suicide from social media using deep learning, BMC Med Inform Decis Mak. 18 (2018) 43, https://doi.org/10.1186/s12911-018-0632-8.
[10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, ArXiv:1810.04805 [Cs]. (2018). http://arxiv.org/abs/1810.04805 (accessed February 11, 2019).
[11] A. Dudchenko, G. Kopanitsa, Comparison of Word Embeddings for Extraction from Medical Records, Int J Environ Res Public Health. 16 (2019) E4360, https://doi.org/10.3390/ijerph16224360.
[12] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, ArXiv: 1901.08746 [Cs]. (2019). http://arxiv.org/abs/1901.08746 (accessed February 6, 2019).
[13] E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, M. McDermott, Publicly Available Clinical BERT Embeddings, in: Proceedings of the 2nd Clinical Natural Language Processing Workshop, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019: pp. 72–78. https://doi.org/10.18653/v1/W19-1909.
[14] A.E.W. Johnson, T.J. Pollard, L. Shen, L.H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, R.G. Mark, MIMIC-III, a freely accessible critical care database, Scientific Data. 3 (2016), 160035, https://doi.org/10.1038/sdata.2016.35.
[15] Branden Chan, Timo Möller, Malte Pietsch, Tanay Soni, Chin Man Yeung, Open Sourcing German BERT, Deepset. (2019). https://deepset.ai/german-bert (accessed September 14, 2019).
[16] F. Cabitza, A. Campagner, The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies, International Journal of Medical Informatics. 153 (2021) 104510, https://doi.org/10.1016/j.ijmedinf.2021.104510.
[17] W.H. Organization, ICD-10: international statistical classification of diseases and related health problems : tenth revision, World Health Organization, 2004 https://apps.who.int/iris/handle/10665/42980 (accessed February 9, 2021.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, in, Advances in Neural Information

Processing Systems (2017) 5998–6008, https://doi.org/10.1017/S0140525X16001837.

[19] Freier Zugang zu juristischen Daten - Open Legal Data, (n.d.). https://de.openlegaldata.io/ (accessed June 21, 2020).

[20] L.A. Ramshaw, M.P. Marcus, Text chunking using transformation-based learning, in: Natural Language Processing Using Very Large Corpora, Springer, 1999: pp. 157–176.

[21] R.E. De Castilho, I. Gurevych, A broad-coverage collection of portable NLP components for building shareable analysis pipelines, in, in: Proceedings of the Workshop on Open Infrastructures and Analysis Frameworks for HLT, 2014, pp. 1–11.

[22] M. Honnibal, I. Montani, S. Van Landeghem, A. Boyd, spaCy: Industrial-strength Natural Language Processing in Python, Zenodo (2020), https://doi.org/10.5281/zenodo.1212303.

[23] deepset.ai, deepset-ai/FARM, deepset, 2021. https://github.com/deepset-ai/FARM (accessed August 20, 2021).

[24] T. Wolf, J. Chaumond, L. Debut, V. Sanh, C. Delangue, A. Moi, P. Cistac, M. Funtowicz, J. Davison, S. Shleifer, Transformers: State-of-the-art natural language processing, in: in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020, pp. 38–45.

[25] MLflow - A platform for the machine learning lifecycle, MLflow. (n.d.). https://mlflow.org/ (accessed August 20, 2021).

[26] German stemming algorithm, (n.d.). http://snowball.tartarus.org/algorithms/german/stemmer.html (accessed August 20, 2021).

[27] PyStemmer, Snowball Stemming language and algorithms, 2021. https://github.com/snowballstem/pystemmer (accessed August 20, 2021).

[28] seatgeek/fuzzywuzzy, SeatGeek, 2021. https://github.com/seatgeek/fuzzywuzzy (accessed August 20, 2021).