Single and Cross-Disorder Detection for Autism and Schizophrenia

Aleksander Wawer¹ · Izabela Chojnicka³ · Lukasz Okruszek² · Justyna Sarzynska-Wawer²

Received: 25 June 2020 / Accepted: 12 January 2021 / Published online: 27 January 2021
© The Author(s) 2021

Abstract

Detection of mental disorders from textual input is an emerging field for applied machine and deep learning methods. Here, we explore the limits of automated detection of autism spectrum disorder (ASD) and schizophrenia (SCZ). We compared the performance of: (1) dedicated diagnostic tools that involve collecting textual data, (2) automated methods applied to the data gathered by these tools, and (3) psychiatrists. Our article tests the effectiveness of several baseline approaches, such as bag of words and dictionary-based vectors, followed by a machine learning model. We employed two more refined Sentic text representations using affective features and concept-level analysis on texts. Further, we applied selected state-of-the-art deep learning methods for text representation and inference, as well as experimented with transfer and zero-shot learning. Finally, we also explored few-shot methods dedicated to low data size scenarios, which is a typical problem for the clinical setting. The best breed of automated methods outperformed human raters (psychiatrists). Cross-dataset approaches turned out to be useful (only from SCZ to ASD) despite different data types. The few-shot learning methods revealed promising results on the SCZ dataset. However, more effort is needed to explore the approaches to efficiently training models, given the very limited amounts of labeled clinical data. Psychiatry is one of the few medical fields in which the diagnosis of most disorders is based on the subjective assessment of a psychiatrist. Therefore, the introduction of objective tools supporting diagnostics seems to be pivotal. This paper is a step in this direction.

Keywords Zero-shot learning · Few-shot learning · Transfer learning · Sentic computing · Schizophrenia detection · Autism spectrum disorder detection

Introduction

The goal of this paper is to explore the automated means of disorder detection from textual data. Our focus is on classic methods, using various types of word frequencies as well

as state-of-the-art approaches based on deep neural text encoders. In other words, the paper is a review of current standard automated methods applicable to the detection of psychiatric disorders.

The methods and features tested in our paper are mostly universal, applicable to multiple other text classification tasks. They mostly do not explore language characteristics that are known to be linked to the disorders.

All data used in our experiments come from two Polish language clinical studies on schizophrenia and autism spectrum disorder. Utterances originate from clinically approved diagnostic tools. The groups are relatively large for a clinical study, as they include 94 subjects and controls for schizophrenia and 74 for autism spectrum disorder. The groups are balanced: each consists of the same amount of healthy controls and diagnosed patients. The type of data is mixed. In the case of schizophrenia, it consists of open-ended answers to six standardized questions. In the case of autism spectrum disorder, it consists of dialogues carried out using a set of stimuli (a picture book).

✉ Aleksander Wawer
axw@ipipan.waw.pl

Izabela Chojnicka
izabela.chojnicka@psych.uw.edu.pl

Lukasz Okruszek
lokruszek@psych.pan.pl

Justyna Sarzynska-Wawer
jsarzynska@psych.pan.pl

¹ Institute of Computer Science, Polish Academy of Sciences, Jana Kazimierza 5, Warszawa 01-248, Poland

² Institute of Psychology, Polish Academy of Sciences, Jaracza 1, Warszawa 00-378, Poland

³ Faculty of Psychology, University of Warsaw, Stawki 5/7, 00-183 Warszawa, Poland

We also used the related character (compatibility) of both investigated disorders to use methods focused on transfer learning in a cross-dataset manner. Using data from one disorder to help predict other disorders is not always possible, but due to data scarcity and other problems it may be an important research direction. In addition to potentially improving the quality of predictions, it can also be used to better understand the relationship between disorders. Our work is probably the first to explore the relationships between autism spectrum disorder and schizophrenia using transfer learning in a cross-dataset setting.

Another area explored in our paper is few-shot learning, a set of techniques to train models with minimal training data, contrary to the standard practice of using a large amount of data. This direction is a key one in the context of automated clinical psychiatry, where the datasets are typically tens in size. Few-shot techniques were first applied to computer vision. Our paper is among the relatively few to apply these techniques in natural language processing and, to our knowledge, the first in disorder detection from textual data.

One of the main goals of this work is also the application of selected Sentic tools to the clinical textual data and comparisons with other means of text representation, such as deep neural text encoders and dictionaries.

The results of the experiments conducted in our article may prove important to support clinical decision-making processes in psychiatry and develop next generation solutions, such as deep neural text representation models combined with machine learning to support diagnosis.

Mental Disorders

Psychiatric disorders are among the most costly non-communicable diseases reported by the World Health Organization's (WHO) disease burden and mortality estimates, and are recognized as a severe public health problem [1]. In recent decades, dynamic biomedical progress has not translated into the same progress in diagnosing and treating mental health problems. Currently, the diagnosis of a mental disorder is based on behavioral criteria described in classification systems, namely the WHO's International Classification of Diseases (ICD-10 and ICD-11, [2, 3]) and the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-5, [4]). Diagnostic assessment is based on behavioral presentation and self-reported symptoms, not biological or etiological information. An attempt to re-conceptualize mental disorders is underway within the Research Domain Criteria (RDoC) framework. RDoC integrates several levels of biological and clinical information, from genomics and neurobiology to behavior and self-report [5]. RDoC is not meant to replace current

diagnostic classifications, but to describe the full range of human behavior from typical to atypical, from mental health to illness.

One of the behavioral elements among the RDoC's cognitive systems domain is language, defined as a system of shared symbolic representations of the world, the self and abstract concepts, which support thought and communication. Interestingly, several psychiatric phenotypes, including schizophrenia and autism, display a range of communicative and linguistic difficulties that may be directly associated with diagnostic criteria, but not necessarily.

Schizophrenia

Schizophrenia (SCZ) is a severe neuropsychiatric disorder that impairs many aspects of functioning, including cognitive, affective, and social aspects of behavior. Symptoms listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) include delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms such as diminished emotional expression [4]. A systematic review of epidemiological data indicates that the lifetime prevalence is between 0.30 and 0.66% [6], and in 2017, there were nearly 20 million cases globally [7]. Even though much effort has been directed toward creating a biopsychosocial model of SCZ and utilizing neurobiological markers for its diagnostic process, SCZ is still formally diagnosed primarily based on the patient's subjective report and observation.

Participants with SCZ

Our sample included 47 patients with SCZ (SCZ group; 21 females, mean age = 34.3 (SD = 7.15)) and 47 demographically matched healthy controls. The inclusion criteria for our participants were: (1) age > 18 years, (2) Polish as first and primary language, and (3) no intellectual disabilities or history of neurological or psychiatric disorders, drug abuse, or any other disorders. For the SCZ group, the criteria included (4) the clinical diagnosis of SCZ as determined by a psychiatrist based on the ICD-10 diagnostic criteria, and (5) a stable clinical condition—no change in treatment and no significant changes in the severity of symptoms for a total of four weeks preceding the examination. Patients who participated in the study were recruited from outpatients being treated at the Institute of Psychiatry and Neurology in Warsaw, Poland. Control group participants were recruited from healthy volunteers who responded to online advertisements. They were paired head to head with patients with SCZ based on their sex, age, and parental education.

Scale for the Assessment of Thought, Language, and Communication (TLC)

The TLC scale [8] contains six questions, four of which, respectively, concern the patient and their family, the person closest to the patient, and their interests and childhood. Two of the questions are more abstract—they ask why people get sick and why people believe in God. The responses were recorded and then transcribed. Patients completed the TLC scale as a part of the initial clinical assessment performed by a qualified psychiatrist. Two variants of the transcriptions were prepared for the patient group—a full version and an edited one (fragments were removed in which patients explicitly admitted to being diagnosed with SCZ). In the case of the individuals from the control group, the TLC scale was performed by an experimenter at the Institute of Psychology of the Polish Academy of Sciences.

Autism Spectrum Disorder

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting communication, social interactions, and behavior. Symptoms listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) include persistent deficits in social–emotional reciprocity, pragmatic (social) language, nonverbal communicative behaviors used for social interaction, and deficits in developing, maintaining, and understanding social relationships [4]. In addition to impairments in social functioning, restricted, repetitive, and stereotyped patterns of behavior and interests (RRBs) are among the core symptoms of ASD.

Although a contributing genetic variation can be found in 5–30% of cases [9], ASD is behaviorally defined, and there is currently no objective way (e.g., a blood test) to detect this disorder. Hence, diagnostic assessment is based on behavioral presentation. The American Academy of Pediatrics [10] has recommended universal screening for ASD at ages 18 and 24 months. Screening tools need to address heterogeneity in clinical presentation among individuals with ASD at different developmental levels and consider variability in age, cognitive abilities, and sex [11].

Participants with ASD

Our sample included 37 individuals with idiopathic ASD (ASD group) and 37 controls with typical development (TD). The inclusion criteria for participants were: (1) age ≥ 6 years, (2) nonverbal IQ ≥ 75 , (3) Polish as a first and primary language, and (4) no hearing, sight, and mobility impairments. For the ASD group, clinical diagnosis of ASD was (5) determined by a psychiatrist based on the ICD-10 diagnostic criteria and meeting the criteria for ASD on the ADOS-2 and SCQ. The exclusion criteria for controls

included a personal or family history of ASD, a history of developmental disorders, and neurological or psychiatric conditions. The intellectual functioning of all participants was tested using the Wechsler Intelligence Scale for Children-Revised for verbal children and adolescents aged 6.0–16.11, and the Wechsler Adult Intelligence Scale for verbal participants older than 16.11.

Recruitment was conducted in four cities in Poland. Participants were contacted through diagnostic and therapeutic centers specialized in diagnosing ASD and other disorders, as well as foundations and associations supporting individuals with developmental disabilities, and public schools and universities.

The ADOS-2 Assessment and the Picture Book Task

The ADOS-2 is a semi-structured, standardized, observational assessment of social communication skills and restricted and repetitive behaviors associated with a diagnosis of ASD [12, 13]. The instrument includes several play-based and conversational activities. The ADOS-2 consists of five modules, each of which is appropriate for individuals of differing ages and language developmental levels, from children of at least 12 months of age to adults. Each module provides some unstructured and structured situations and tasks that allow for observation of behaviors and interest deficits. Modules 3 and 4 are intended for fluently speaking participants having the expressive language skills of a typical four-year-old child. Participants assessed with Module 3 or 4 should be able to produce a range of sentence types and grammatical forms as well as provide information outside the context of the ADOS-2 assessment. The Polish version of the ADOS-2 demonstrates good psychometric characteristics [14].

For this study, the ADOS-2 assessments were administered by qualified examiners trained in the instrument for scientific purposes with the supervision of an independent ADOS-2 trainer.

Language samples were derived from the Modules 3 and 4 ADOS-2 *Telling a Story from a Book* task. To evoke narratives, examiners used the picture book “Tuesday” by David Wiesner. The book depicts the adventures of a group of frogs, which one Tuesday evening start to float on lily pads to a nearby suburb. The pictures show mysterious, unreal, and humorous situations, as well as the mental and emotional states of several characters. Following the ADOS-2 instructions, the examiner introduced the story on pages 1 and 2: “Let’s look at this book. It tells a story in pictures. See, it starts on Tuesday, around eight o’clock in the evening. A turtle is sitting on a log. He sees something. Can you tell me the story as we go along?” The participant continued the story. According to the ADOS-2 manual, an examiner may interact with the participant to keep rapport

and naturally elicit narratives. An examiner is not allowed to model labeling emotions. In our experiment, the examiners did not take turns in telling a story.

The participants' narratives during the ADOS-2 *Telling a Story from a Book* task were video recorded and then transcribed by two experienced transcribers. The transcribers were blind to the diagnosis.

Schizophrenia and Autism Spectrum Disorder

Until 1980, autism was not officially recognized in diagnostic classifications. In the DSM-II, autism was placed under the diagnostic umbrella of SCZ. However, there came a growing need to differentiate between mental disorders that developed during the infancy period (i.e., autism) and the psychoses arising mostly in adolescence and later (i.e., SCZ). The latter encompasses a loss of the sense of reality in individuals who were sufficiently socially adapted and functioned independently before the first symptoms appeared. On the other hand, impairments in autism are present from the early stages of child development [15]. In the next edition of the American Psychiatric Association classification, the DSM-III, the term infantile autism was included in a new category, Pervasive Developmental Disorders, emphasizing this condition's developmental aspects. In the DSM-5, autism spectrum disorder replaced the terms childhood autism and pervasive developmental disorders, reflecting the heterogeneity and dimensional characteristics of the disorder [16]. Nevertheless, ASD and SCZ appear to share genetic risk factors, etiopathogenesis pathways, as well as potential links in the specific clinical characteristics of these two disorders [17].

SCZ and ASD are usually perceived as two opposite ends of one continuum of modes of cognition, including typically developing cognition. Many features of these disorders seem to have opposite phenotypes—in one disorder they are overgrowth, while in the other undergrowth. Such diametric patterns are observed in features related to social brain development, such as features of gaze, agency, mentalizing (deficient in ASD, excessive in SCZ), as well as in verbal communication and language. Concerning language, differences between ASD and SCZ are related to formal thought disorder, which refers to an impaired capacity to sustain coherent discourse. Formal thought disorders, like symptoms of SCZ in general, can also be divided into positive and negative. Positive thought disorders are characterized by the pressure of speech, derailment, tangentiality, incoherence, and illogicality, whereas negative thought disorders are associated with problems with the poverty of speech and speech content. Patients with SCZ more often exhibit positive symptoms; however, poverty of speech or speech content may also occur. In patients with ASD, negative thought disorders dominate. In a study

conducted by Rumsey [18], adults with ASD, most of whom were high functioning, showed a high frequency of poverty of speech, poverty of speech content, and perseverations. They did not differ from the healthy control group on any of the features of positive thought disorder. Patients with SCZ, on the other hand, had more manifestations of the pressure of speech, derailment, and illogicality in their statements. Similar results were obtained by Dykens, Volkmar, and Glick [19] on young adults and older adolescents. Other differences regarding language deficits appear to result from memory impairments. According to Eigsti and colleagues [20], individuals with ASD integrate semantic information during the interpretation of syntactic structures and consolidate semantic knowledge from such structures differently from controls with TD. Other linguistic impairments include problems with relative clauses, wh-questions, raising, and passives [21]. These impairments indicate a deficit in procedural memory [22]. According to the same authors, patients with SCZ exhibit language impairments related rather with semantic memory, working memory, and executive functions, such as problems with speech perception (auditory verbal hallucinations), abnormal speech production (formal thought disorder), and production of abnormal linguistic content (delusions) [23].

Among other difficulties described in SCZ patients are word-finding difficulties and verbosity. As a result, the narration of a person with SCZ is vague and ambiguous [24]. Similarly, individuals with ASD produce narrations that are less causally connected and less coherent than those produced by controls [25]. Individuals with SCZ may have problems with the pragmatic use of language [24], which is one of the most noticeable characteristics of ASD. Both ASD and SCZ individuals display problems with clausal embedding and relative clauses, although individuals with ASD have a number of additional syntactic limitations. Additionally, patients with ASD and SCZ differ from healthy people in terms of sentiment. The sentiment of the patients with ASD has lower valence, indicative of poorer moods than healthy people [26]. Patients with SCZ use significantly more words related to negative and less related to positive emotions [27].

Computer Methods in Health Care

To date, several attempts have been made to use automatic text analysis and various computer models to detect psychiatric disorders and diseases.

Most studies that concern SCZ focus on one of its symptoms: speech coherence disorder. Bedi et al. [28] used latent semantic analysis (LSA) and obtained 100% accuracy in predicting later psychosis onset in high-risk youths. In another study, Elvevaag et al. [29] identified a reduction in semantic coherence in patients with SCZ and observed a difference in LSA scores

between patients with severe and mild disorders. A follow-up study [30] examined speech differences between patients with SCZ, their first-degree relatives, and unrelated healthy people, and observed that LSA combined with structural speech analysis differentiates first-degree relatives of patients with SCZ from unrelated healthy individuals. Another approach to coherence-based SCZ prediction was described in [31].

In the case of ASD, several research groups have applied machine learning approaches on different types of data. Artificial neural networks were already used to model specific cognitive problems such as attentional impairments [32], functional disconnections [33] or alterations of the precision of predictions, and sensory information processing [34]. Recently, machine learning models have been tested for differentiating between children with ASD and with TD using feature tagging of home videos [35]. Thabtah and Peebles proposed a new automated approach to using data from screening questionnaires for ASD classification [36]. However, to our knowledge, there are no studies in which computer models were used to distinguish patients with ASD from healthy controls on the basis of language features. However, differences in ASD patients and healthy people's languages have already been proved [37].

Some efforts have been directed at other disorders, such as Parkinson's disease. In [38], two transfer learning strategies were implemented (layer freezing and fine-tuning) for the classification of Parkinson's patients using handwriting signals. The authors showed that the proposed strategies improved the accuracy up to 92.3%.

Sentic Computing and Health Care

Health-related applications of Sentic computing include distilling knowledge from reviews of health services [39]. The authors employed a language visualization and analysis system, an emotion categorization model, opinion mining based on a web ontology language, and techniques for finding and defining topic-dependent concepts.

In another study, Sentic methods were used to process tweets in the contexts of the 2014 Ebola and 2016 Zika outbreaks [40]. Researchers applied the word embedding approach to classifying unstructured text data, using PubMed abstracts for training purposes.

The Sentic research that is closest to ours is [41], which focuses on classifying suicidal ideation and other mental disorders. The method uses text representation enhanced with lexicon-based sentiment scores and latent topics, and relation networks with an attention mechanism.

The most recent version of SenticNet [42], a commonsense knowledge base for sentiment analysis, integrates logical reasoning within deep learning

architectures. However, for health care and diagnostics, the more relevant resource seems to be AffectiveSpace [43].

Method

In this section, we describe methods to detect both disorders. They include both automated approaches and human raters.

In automated approaches, we used three types of text representation: bag of words, dictionary vectors (POS + LCM + Sentiment), and utterance embedding vectors obtained from the Universal Sentence Encoder (USE) [44]. We then used those representations with multiple types of algorithms to classify texts:

- XGBoost [45].
- Linear kernel SVM, the LinearSVC implementation from scikit-learn [46] with default settings.
- Neural network dense layer of 64 neurons with *tanh* activation, a dropout layer of 0.5, followed by a classification layer with *sigmoid* activation.

Bag of Words (BOW)

The performance of this well-known type of text representation reflects the amount of information useful for detecting mental disorders that is present in word occurrences. In this approach, we used unigram word-level vectors with TF-IDF transformation. We used the implementation from scikit-learn [46] with default hyperparameters for each class.

Dictionary Vectors (POS + LCM + Sentiment)

The second baseline approach is to learn disorders not from word occurrences, but 21 general language categories. The list of categories contains:

- Syntactic word classes (POS or part of speech).
- The Linguistic Category Model or LCM [47], a tool to measure the level of language abstraction.
- Sentiment (word polarity).

We scaled the word counts to eliminate the influence of varying document lengths.

Part of Speech

To obtain part-of-speech occurrences, we used the Spacy-PL module¹, which internally uses a bi-LSTM neural network to label the tokens [48]. This tool categorizes tokens into 15 categories.

¹ <https://github.com/ipipan/spacy-pl>

Linguistic Category Model (LCM)

The LCM [47] distinguishes among four categories of words that are used to describe interpersonal behavior and communication. These four categories include three types of verbs and adjectives. The most abstract category in LCM is adjectives (ADJ), such as smart or nice, which are usually used to describe highly conceptual dispositions or personality traits. The most abstract verb is state verbs (SV) that refer to mental (e.g., to think, to understand) and emotional (e.g., to love, to admire) states or changes therein. More concrete than SVs in the LCM are interpretative action verbs (IAV), which are used to describe a general class of behaviors without identifying the specific behavior to which they refer in a given context (e.g., help). The most concrete type of verb is the descriptive action verb (DAV). DAVs refer to verbs that describe a single and observable event defined by at least one physically invariant feature and a clear beginning and end (e.g., a call). In our study, we used the Polish LCM dictionary (LCM-PL; [49]), which contains the 6,000 most frequent Polish verbs. For each subject, the number of tokens (an individual occurrence of a linguistic unit) of a given word type (DAVs, IAVs, SVs, ADJs) in the whole statement was added up. We calculated the level of abstraction according to the weighted summation formula recommended by the LCM's authors ($DAV + IAV * 2 + SV * 3 + ADJ * 4$). The LCM part of the vector contributes four variables.

Sentiment

We used a dictionary of 5421 positive and negative words, compiled from two sources. The dictionary was created manually and represents the sentiment of the most frequent sense of a word, and therefore requires no word sense disambiguation. The first source was 32 thousand lexical units in plWordNet, manually annotated with sentiment labels [50]. Sentiment labels were re-assigned and verified to represent the most frequent sense only. The second source is a manually labeled dictionary of 1774 negative and 1493 positive words, pulled from a rule-based sentiment analyzer [51]. The sentiment part of the vector contains two variables: the number of positive and negative words.

Sentic Representations

Another two approaches to text representation investigated in our study became possible after translating the data into English. Texts were translated automatically using Google's API and then verified manually to fix possible issues.

The most promising Sentic resources in the context of automating diagnosis are AffectNet and AffectiveSpace, described in more detail below. They enable analysis at the

concept level as well as the creation of multidimensional affective vectors to represent texts.

In each of the methods, vector generation begins with finding concepts in the analyzed texts. On average, 59 concepts were found in each of the ASD texts and 168 concepts in each of the SCZ texts.

AffectNet

The first English language method to represent texts is based on vectors computed using AffectNet [52]: an affective semantic network, in which common-sense concepts are linked to a hierarchy of affective domain labels. The entire knowledge repository is a large matrix, with every known concept of some statement being a row and every known semantic feature (relationship+concept) being a column.

AffectNet vectors were calculated as the sum of the values for each semantic feature of each concept detected in the text. The vectors, converted to dense representation, consisted of 40195 features.

AffectiveSpace

The second English language method uses AffectiveSpace vectors [43]. Due to a large number of concepts and semantic features, the AffectNet matrix may be too multidimensional and contain too much information. AffectiveSpace, a low-rank approximation obtained from AffectNet using matrix decomposition, is an attempt to solve this problem.

AffectiveSpace vectors were calculated in a manner similar to max pooling operation used in neural networks. In this method, the n th position of a text vector is the maximum value that appears at a n th position among the vectors of all concepts present in a text. The vectors consisted of 100 features.

Universal Sentence Encoder (USE) Text Representations

One of the methods used to obtain text representations was the USE [44]. This deep neural network text encoder supports 16 languages, among them Polish. The input to the network is variable-length text in any of the 16 supported languages and the output is a 512-dimensional vector. The model is intended to be used as a universal utterance representation suitable for text clustering, semantic textual similarity retrieval, cross-lingual text retrieval, and many other tasks. It serves as an accurate but general-purpose vector representation of textual utterances for text classification.

The multitask training setup of USE was based on the technique described in [53]. In this technique, representations are obtained using a dual-encoder-based

Table 1 Results obtained by automated methods applied to detect schizophrenia (SCZ) and autism (ASD). Reported measures include accuracy (acc.), specificity (spec.), and sensitivity (sens.). The upper part SINGLE-DOMAIN METHODS refers to experiments carried out within the same disorder data. All experiments, except those marked (EN), were carried out on Polish language texts. Reported measures, in this case, are averages in leave-one-out cross-validation. The lower part TRANSFER LEARNING METHODS refers to the results of cross-dataset experiments described in Section 2.5. We report the results achieved on the target dataset

		SCZ			ASD		
SINGLE-DOMAIN METHODS							
Representation	classifier	acc.	spec.	sens.	acc.	spec.	sens.
AffectNet (EN)	SVM	0.627	0.617	0.638	0.486	0.486	0.486
	XG Boost	0.617	0.574	0.660	0.419	0.459	0.378
AffectiveSpace (EN)	SVM	0.585	0.617	0.553	0.554	0.622	0.486
	XG Boost	0.532	0.532	0.532	0.310	0.378	0.243
Bag of words	SVM	0.766	0.745	0.788	0.622	0.568	0.676
	XG Boost	0.670	0.681	0.660	0.466	0.540	0.351
POS + LCM +Sentiment	SVM	0.606	0.425	0.787	0.459	0.351	0.568
	XG Boost	0.684	0.659	0.702	0.513	0.540	0.486
USE	SVM	0.808	0.830	0.787	0.635	0.649	0.622
	XG Boost	0.713	0.723	0.568	0.568	0.568	0.568
USE	Dense	0.758	0.729	0.813	0.582	0.443	0.769
TRANSFER LEARNING METHODS							
USE	Dense [0-shot]	0.521	0.085	0.957	0.608	0.730	0.486
USE	Dense [pre-train]	0.741	0.776	0.731	0.526	0.482	0.630

model trained to maximize the representational similarity between sentence pairs drawn from parallel data. The representations are enhanced using multitask training and unsupervised monolingual corpora.

In our experiments, we used the USE with two types of heads. In a single-domain scenario, with machine learning algorithms XGBoost [45] and linear kernel SVM, and in the transfer learning scenario with a neural network dense layer. The choice of a dense layer was dictated by the ability to train the model for a number of epochs on one (source) dataset and continue training on another (target) dataset.

Transfer Learning

Transfer learning is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. We use this framework to train models on one disorder's data (for example, train on SCZ) and apply to another disorder's data (for example, test on ASD).

This is closely related to zero-shot learning, a problem setup, where at test time, a learner observes samples from classes that were not observed during training. We apply zero shot in a sense similar to that of transfer learning: Predict using no labeled examples of a given disorder, basing solely on another disorder's dataset.

We have conducted the following experiments:

- **USE with zero shot** ([0-shot] in Table 1). In this scenario, we trained a dense classifier on USE vectors from one dataset and test on another dataset. In other words, only SCZ data are used to detect ASD and vice versa.

- **USE with pre-train** ([pre-train] in Table 1). In this experiment, we began just like in the previous experiment, by training a dense classifier on USE vectors from one dataset (source). Unlike in the zero-shot scenario, we further trained it on the target dataset. More precisely, we trained on the entire source dataset, then randomly divided the target data into two subsets, one for further training (70%) and the other for testing (30%). To avoid overfitting, we trained on the source data only as long the loss decreased on the target training subset (early stopping set to 1 epoch, max 20 epochs). Then, we continued to train on the target training subset for 10 epochs. Finally, we calculated the quality of the predictions on the target testing subset. In our experiments, we repeated this procedure 10 times and averaged the resulting scores.

Human Raters

To assess the classification effectiveness of clinicians, we asked clinicians experienced in the diagnosis of ASD and SCZ to classify participants based on their statements. One psychiatrist (Rater 1) and one psychologist (Rater 2) rated the SCZ group, while two psychiatrists (Rater 3 and 4) rated the ASD group. The psychiatrists were asked to read utterances produced by a participant (without examiners' statements in the case of the ASD group; a cut version of transcripts in the case of the SCZ group) and assign the participant to one of the two groups: SCZ or TD (Raters 1 and 2) and ASD or TD (Raters 3 and 4). The psychiatrists were not guided by any predetermined criteria. They assessed persons based on their knowledge and experience. Rater 2 (qualified psychologist) divided the subjects into two

Fig. 1 Accuracy of selected computer-based methods and human baselines

groups, patients and healthy controls, based on their TLC scores (Thought, Language and Communication Scale). The scale includes 18 language phonemes whose presence is marked on a scale of 0 to 3 or 4 points (depending on the item). These include poverty of speech and speech content, illogicality, incoherence, clanging, neologisms, word approximations, pressure of speech, distractible speech, tangentiality, derailment, stilted speech, echolalia, self-reference, circumstantiality, loss of goal, perseveration, and blocking. A general result was obtained by summing up points from all of the listed items, and this was used as the criteria that assigned a given person to either a patient group or a control group. The most accurate results were obtained by calculating a median for grouped data. The results obtained by the experts are shown in Fig. 1.

Results

Results of Automated Methods

Table 1 summarizes the results obtained by automated methods. Results for detecting SCZ were much higher than those for ASD, with the difference as large as 20% in the case of accuracy. The best results for both disorders in terms of accuracy were obtained by the SVM classifier with USE embeddings in a single-domain scenario. The BOW representation yielded relatively good results. One should

also consider the performance of the POS+LCM+Sentiment text representation on SCZ a relative success, especially with XG Boost, as it consists of only 21 features. Among the Sentic text representations that work with English-language texts, AffectNet vectors with SVM turned out to be the best in terms of accuracy for SCZ, and AffectiveSpace vectors with SVM for ASD.

The results of transfer learning experiments are not straightforward. When predicting ASD by models trained on SCZ, one should note the high results achieved by USE in a zero-shot scenario, with very good accuracy and the best observed specificity. However, this holds only in one direction, as the results of predicting SCZ by models trained on ASD are marginally close to random.

The results of transfer learning in the scenario with pre-training ([pre-train] in Table 1) reveal that it does not help in reaching state-of-the-art results. A dense layer classifier without pre-training manages to achieve slightly but noticeably better results than the variant with pre-training.

However, this does not mean that there is no valuable information that one could use, and there appears to be some relevant ASD-predicting information in SCZ data. This can be confirmed not only by the relative success of the zero-shot approach. In the pre-training ([pre-train]) setting, the training on the source domain was carried as long as it improved the loss on the dev set consisting of target domain data only. In most cases, the training continued for 3 to 6 epochs, on average. Since the training did not stop after one epoch, it indicates that the first epochs of training improved the loss of not just one, but both domain data.

Table 2 Accuracy of Prototypical Networks few-shot learning and baseline dense network for schizophrenia (SCZ) and autism spectrum disorder (ASD). The nt parameter refers to the number of training

examples, qt to the number of observations in the query set. * For ASD, nt=30 and qt=8, the actual nt value was 25 due to data size limits

	SCZ			ASD		
	Prototypical Networks		Dense baseline	Prototypical networks		Dense baseline
nt	qt=5	qt=8	N/A	qt=4	qt=8	N/A
2	0.53	0.525	0.505	0.497	0.512	0.545
10	0.586	0.561	0.545	0.476	0.463	0.502
20	0.609	0.591	0.557	0.468	0.453	0.510
30	0.646	0.624	0.596	0.5	0.448*	0.518

Results of Human and Automated Methods

Figure 1 illustrates the results of both humans and the best computer-based approaches to the detection of ASD and SCZ.

The best computer-based methods outperform the best human baselines: Raters 1 and 2 diagnosing SCZ, and Raters 3 and 4 diagnosing ASD. This observation is more significant in the case of SCZ, but also holds for ASD. Transfer learning in the zero-shot scenario yields good results from SCZ to ASD, but not in reverse (here, the results approximate the random baseline of 0.5). Top performers are USE embeddings with SVM classifier.

Few-Shot Learning

Deep learning methods typically rely on large-scale datasets with at least thousands of labeled examples. A few-shot learning approach is proposed to tackle the problem of low quantities of training data, and efficiently learn from few examples.

We modified the code in one of the publicly available few-shot repositories² and adapted it for use with text data. Instead of convolutional encoders, we introduced dense layers, acting on USE embeddings as input instead of picture pixels.

Out of the several tested solutions, the one that yielded the best results is Prototypical Networks [54]. It is based on the assumption that there exists an embedding that points cluster around, called a prototype. The algorithm learns a metric space in which classification can be performed by computing distances to the prototype representations of each class. After sampling the support and the query examples, the prototypes are computed as the mean of the embeddings for the support set (labeled samples). The probability for a query point to belong to a given class is equal to the softmax over the distances to the prototypes. We set the number of

epochs to 30, the size of the query dataset as the qt parameter. The setting is a 2-way (k-way) one, with an n-parameter (n-shot) determined by the nt as in the Table 2, which presents the results of the Prototypical Networks approach on both datasets, averaged from 5 experiments. The table also includes a baseline setting: averaged results of an experiment repeated 10 times with default neural network training, based on the USE representation and the dense classifier described at the beginning of Section 2. The size of randomly selected training sets was nt cases, and test sets consisted of 20 random observations. The number of epochs was set to 10.

The results indicate that the only dataset where the few-shot learning method is applicable comes from patients with SCZ. In the case of ASD, the values do not exceed random baselines. In the case of SCZ, there is a clear relationship between the increasing accuracy and the number of training observations. Quite possibly, the best values could be achieved by even larger training sets than the tested ones. It is also clear that the Prototypical Networks few-shot solution has the edge over the baseline of typical neural network training.

Discussion

Prior work has shown that automated machine learning classification in psychology and psychiatry can reach high levels of accuracy [28, 35, for instance]. In this study, we used classic methods as well as state-of-the-art deep learning approaches to test their effectiveness in the automated detection of ASD and SCZ from textual data. We compared the effectiveness of: (1) blinded classification by psychiatrists, (2) dedicated standardized psychological instruments involving textual data collection, and (3) several automated means applied to the data collected by these standardized tools. They included multiple text representations such as baseline methods (bag of words, dictionary vectors) and the USE based on a deep neural network. We experimented with several model training

² <https://github.com/oscarknagg/few-shot>

regimes, including transfer learning, zero-shot, and few-shot scenarios. We also investigated the possibility of knowledge transfer between datasets associated with both disorders.

Our study only investigated predictions from textual utterances, which is not a standard diagnostic practice. It is worth noting that psychiatrists make a diagnosis based on tools that take into account complex behavioral qualities such as communication skills, social functioning, and patterns of behaviors and interests. The effectiveness of such holistic tools used by psychiatrists is very high, as for example, ADOS-2 achieved an accuracy of 95% on our ASD dataset. However, the reported results of computer models are still very promising and more accurate than expert assessments based solely on text.

The results indicate that training the model on the statements of patients with SCZ improves the accuracy of recognizing people with ASD, while training it on the statements of people with ASD does not improve the accuracy of recognizing people with SCZ. This is probably because both patients with SCZ and with ASD exhibit symptoms of negative thought disorder such as poverty of speech or poverty speech content, which enables the model to learn to detect these types of symptoms. On the other hand, people with ASD rarely exhibit the positive symptoms (e.g., the pressure of speech, derailment) that dominate the statements of people with SCZ, and therefore, training the model on the statements of individuals with ASD does not make it easier to recognize these symptoms and, as a result, does not affect the effectiveness of distinguishing people with SCZ from healthy people. This notion is further substantiated by a recent study [55] that observed high rates of patients with SCZ misclassified as individuals with ASD (43.59 percent) and low specificity (56.41 percent) for distinguishing both clinical groups with ADOS-2, which focuses on ASD-specific symptomatology.

Furthermore, participants assessed with the TLC Scale (both SCZ and TD) produced longer sentences in comparison to participants (both ASD and TD) assessed using the ADOS-2 picture book task. Research on narration in ASD indicates that narrative context matters and participants with ASD exhibit more difficulties in less structured tasks, i.e., narratives of personal experiences (as in the case of the TLC scale) or pictures from the Thematic Apperception Test than storybook tasks [56]. This indicates the possible importance of proper task selection as the basis for computational linguistics analyses.

Limitations

We had access to large groups by clinical standards, but they were of different ages (adults suffering from SCZ and children suffering from ASD), which seems to be of great

importance in terms of language and linguistic competences that develop with age. Thus, the more complex utterances produced by adults from the SCZ sample may have provided information absent in the ASD group. Besides, these groups were examined using various tools. The use of methods based on transfer learning will undoubtedly be more effective when using similar data—subjects of similar age, responding to the same topics, and finally, the same text type (not answers to questions as in TLC and dialogues as in ADOS-2).

The limitation of experiments carried out using Sentic resources, as reported in our paper, is the necessity to translate texts into English. Despite promising results, the optimal solution would be to use natively Polish solutions.

Future Work

One promising direction of future work will be based on improving few-shot techniques, adapting to textual input. Currently, the methods are being developed mostly in the area of computer vision, while low-data text applications such as clinical text datasets are an important and promising area, as indicated by our results.

The experiments described in this paper could improve screening and diagnostic systems. In the case of ASD, one solution of this kind has recently been proposed by [57]. Our results provide input for the text and language processing component of an automated screening solution.

Another direction of research is the exploration of the language properties characteristic of patients with SCZ and ASD. This can be achieved with AffectNet and AffectiveSpace.

Conclusions

Detection of mental disorders from textual input is an emerging field for applied machine and deep learning methods. In this article, we explored the limits of automated detection of ASD and SCZ: two distinct disorders with unique characteristics, but also certain similarities. We compared the classification accuracy of: (1) dedicated diagnostic tools that involve collecting textual data, (2) automated methods applied to the data gathered by these tools, and (3) psychiatrists.

Our article tested the effectiveness of several baseline approaches, such as bag of words and dictionary-based vectors, followed by a machine learning model. We used two Sentic text representations, made possible by translating the utterances into English. We also applied select state-of-the-art deep learning methods for text representation and inference. Finally, we experimented with the transfer, zero-shot, and few-shot learning methods.

Our results revealed that the best breed of automated methods outperforms human raters (psychiatrists), with the highest scores obtained by deep neural network text encoders with machine learning heads. Cross-dataset approaches (transfer learning) turned out to be useful despite different data types, possibly due to the common background behind both impairments. However, the benefits were demonstrated only from SCZ to ASD. Few-shot learning methods revealed promising results on the SCZ dataset. However, more work is needed to explore the approaches to efficiently train models, given minimal amounts of labeled clinical data.

Our work is a contribution to the field of automated diagnosis, which in the future could provide input for automated psychiatric assessments. The introduction of objective tools supporting diagnostics seems to be pivotal. Such tools can be based on automated text analysis that achieves increasingly better accuracy. This paper is a step in this direction.

Acknowledgements The SCZ research was supported by The National Science Centre (Poland) grant number UMO-2016/23/D/HS6/02947 awarded to Łukasz Okruszek. The ASD study was supported by the State Fund for Rehabilitation of Disabled Persons (Poland) grant number BEA/000020/BF/D awarded to Izabela Chojnicka with a contribution from the University of Warsaw.

Compliance with Ethical Standards

Conflicts of Interest The authors declare that they have no conflict of interest.

Ethical Standard All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kassebaum NJ, Arora M, Barber RM, Bhutta ZA, Brown J, Carter A, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1603–58.
- World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. World Health Organization; 2007.
- World Health Organization. International Statistical Classification of Diseases and Related Health Problems. World Health Organization; 2020. Available from: <https://icd.who.int/browse11/l-m/en> of subordinate document. Accessed 06 June 2020
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). Arlington, VA: American Psychiatric Association; 2013.
- Kelly J, Clarke G, Cryan J, Dinan T. Dimensional thinking in psychiatry in the era of the Research Domain Criteria (RDoC). *Ir J Psychol Med*. 2018;35(2):89–94.
- Van Os J, Kapur S. Schizophrenia. *Lancet*. 2009;374(9690):635–45.
- James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1789–858.
- Andreasen NC. Scale for the assessment of thought, language, and communication (TLC). *Schizophr Bull*. 1986;12(3):473–82.
- Schaaf CP, Betancur C, Yuen RK, Parr JR, Skuse DH, Gallagher L, et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nat Rev Genet*. 2020;21:367–76.
- Johnson CP, Myers SM, Lipkin PH, Cartwright JD, Desch LW, Duby JC, et al. Identification and evaluation of children with autism spectrum disorders. *Pediatrics*. 2007;120(5):1183–215.
- Lai MC, Lombardo MV, Baron-Cohen S. Autism. *Lancet*. 2014;383(9920):896–910.
- Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1–4. Torrance, CA: Western Psychological Services; 2012.
- Lord C, Luyster RJ, Gotham K, Guthrie W. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module. Torrance, CA: Western Psychological Services; 2012.
- Chojnicka I, Pisula E. Adaptation and validation of the ADOS-2. Polish version. *Front Psychol*. 2017;8:1916.
- Schopler E, Mesibov GB. Diagnosis and Assessment in Autism. Current Issues in Autism. New York: Springer Science & Business Media; 2013.
- McDougle CJ. Autism Spectrum Disorder. Primer on: Oxford University Press; 2016.
- De Crescenzo F, Postorino V, Siracusano M, Riccioni A, Armando M, Curatolo P, et al. Autistic Symptoms in Schizophrenia Spectrum Disorders: a Systematic Review and Meta-analysis. *Front Psychiatr*. 2019;10:78.
- Rumsey JM, Andreasen NC, Rapoport JL. Thought, language, communication, and affective flattening in autistic adults. *Arch Gen Psychiatr*. 1986;43(8):771–7.
- Dykens E, Volkmar F, Glick M. Thought disorder in high-functioning autistic adults. *J Autism Dev Disord*. 1991;21(3):291–301.
- Eigsti IM, de Marchena AB, Schuh JM, Kelley E. Language acquisition in autism spectrum disorders: A developmental review. *Res Autism Spectr Disord*. 2011;5(2):681–91.
- Perovic A, Janke V. Issues in the acquisition of binding and control in high-functioning children with autism. *UCL Working Papers in Linguistics*. 2013;25:131–143.

22. Benítez-Burraco A, Murphy E. The oscillopathic nature of language deficits in autism: from genes to language evolution. *Front Hum Neurosci*. 2016;10:120.
23. Murphy E, Benítez-Burraco A. Bridging the gap between genes and language deficits in schizophrenia: an oscillopathic approach. *Front Hum Neurosci*. 2016;10:422.
24. Marini A, Spoletini I, Rubino IA, Ciuffa M, Bria P, Martinotti G, et al. The language of schizophrenia: An analysis of micro and macrolinguistic abilities and their neuropsychological correlates. *Schizophr Res*. 2008;105(1-3):144–155.
25. Diehl JJ, Bennetto L, Young EC. Story recall and narrative coherence of high-functioning children with autism spectrum disorders. *J Abnorm Child Psychol*. 2006;34(1):83–98.
26. Nguyen T, Duong T, Phung D, Venkatesh S. Affective, linguistic and topic patterns in online autism communities. In: *International Conference on Web Information Systems Engineering*. Springer; 2014. p. 474–488.
27. Mitchell M, Hollingshead K, Coppersmith G. Quantifying the language of schizophrenia in social media. In: *Proceedings of the 2nd workshop on Computational linguistics and clinical psychology: From linguistic signal to clinical reality*; 2015. p. 11–20.
28. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophr*. 2015;1(1):1–7.
29. Elvevåg B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. *Schizophr Res*. 2007;93(1-3):304–16.
30. Elvevaag B, Foltz PW, Rosenstein M, DeLisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *J Neurolinguistics*. 2010;23(3):270–84.
31. Iter D, Yoon J, Jurafsky D. Automatic detection of incoherent speech for diagnosing schizophrenia. In: *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*. New Orleans, LA: Association for Computational Linguistics; 2018. p. 136–146.
32. Gustafsson L, Papliński AP. Self-organization of an artificial neural network subjected to attention shift impairments and familiarity preference, characteristics studied in autism. *J Autism Dev Disord*. 2004;34(2):189–98.
33. Park J, Ichinose K, Kawai Y, Suzuki J, Asada M, Mori H. Macroscopic cluster organizations change the complexity of neural activity. *Entropy*. 2019;21(2):214.
34. Philippsen A, Nagai Y. Understanding the cognitive mechanisms underlying autistic behavior: a recurrent neural network study. In: *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE; 2018. p. 84–90.
35. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS Med*. 2018;15(11):1–20.
36. Thabtah F, Peebles D. A new machine learning model based on induction of rules for autism detection. *Health Informat J*. 2019;26(1):264–286.
37. Chojnicka I, Wawer A. Social language in autism spectrum disorder: A computational analysis of sentiment and linguistic abstraction. *PLoS One*. 2020;15(3):e0229985.
38. Naseer A, Rani M, Naz S, Razzak MI, Imran M, Xu G. Refining Parkinson's neurological disorder identification through deep transfer learning. *Neural Comput Appl*. 2020;32(3):839–54.
39. Cambria E, Hussain A, Durrani T, Havasi C, Eckl C, Munro J. Sentic computing for patient centered applications. In: *IEEE 10th International Conference on Signal Processing Proceedings*; 2010. p. 1279–1282.
40. Khatua A, Khatua A, Cambria E. A tale of two epidemics: contextual word2vec for classifying twitter streams during outbreaks. *Inform Process Manag*. 2019;56(1):247–57.
41. Ji S, Li X, Huang Z, Cambria E. Suicidal ideation and mental disorder detection with attentive relation networks. *arXiv preprint arXiv:200407601*. 2020.
42. Cambria E, Li Y, Xing FZ, Poria S, Kwok K. SenticNet 6: ensemble application of symbolic and subsymbolic AI for sentiment analysis. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management. CIKM '20*. New York, NY, USA: Association for Computing Machinery; 2020. p. 105–114.
43. Cambria E, Fu J, Bisio F, Poria S. AffectiveSpace 2: enabling affective intuition for concept-level sentiment analysis. In: *AAAI*; 2015. p. 508–514.
44. Yang Y, Cer D, Ahmad A, Guo M, Law J, Constant N, et al. Multilingual Universal Sentence Encoder for semantic retrieval. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Online: Association for Computational Linguistics; 2020. p. 87–94.
45. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16*. New York, NY, USA: ACM; 2016. p. 785–794.
46. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
47. Semin GR, Fiedler K. The cognitive functions of linguistic categories in describing persons: Social cognition and language. *J Pers Soc Psychol*. 1988;54(4):558–68.
48. Krasnowska-Kiera K. Morphosyntactic disambiguation for Polish with Bi-LSTM neural networks. In: Vetulani Z, Paroubek P, editors. *Proceedings of the 8th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*. Pozna, Poland: Fundacja Uniwersytetu im. Adama Mickiewicza w Poznaniu; 2017. p. 367–371.
49. Wawer A, Sarzyńska J. The Linguistic Category Model in Polish (LCM-PL). In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: European Language Resources Association (ELRA); 2018. p. 4398–4402.
50. Zaśko-Zielińska M, Piasecki M, Szpakowicz S. A large wordnet-based sentiment lexicon for Polish. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing*. Hissar, Bulgaria: INCOMA Ltd. Shoumen, BULGARIA; 2015. p. 721–730.
51. Buczyński A, Wawer A. Shallow parsing in sentiment analysis of product reviews. In: *Proceedings of the Partial Parsing workshop at LREC*; 2008. p. 14–18.
52. Cambria E, Hussain A. Sentic computing: a common-sense-based framework for concept-level sentiment analysis. 1st ed. Incorporated: Springer Publishing Company; 2015.
53. Chidambaram M, Yang Y, Cer D, Yuan S, Sung Y, Strophe B, et al. Learning cross-lingual sentence representations via a multi-task dual-encoder model. In: *Proceedings of the 4th Workshop on Representation Learning for NLP (Repl4NLP-2019)*. Florence, Italy: Association for Computational Linguistics; 2019. p. 250–259.
54. Snell J, Swersky K, Zemel R. Prototypical Networks for Few-shot Learning. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. *Advances in Neural Information Processing Systems*. Curran Associates: Inc; 2017. p. 4077–4087.
55. Trevisan DA, Foss-Feig JH, Naples AJ, Srihari V, Anticevic A, McPartland JC. Autism spectrum disorder and schizophrenia are better differentiated by positive symptoms than negative symptoms. *Front Psychiatr*. 2020;11:548.

56. Losh M, Capps L. Narrative ability in high-functioning children with autism or Asperger's syndrome. *J Autism Dev Disord*. 2003;33(3):239–251.
57. Shahamiri SR, Thabtah F. Autism AI: a new autism screening system based on Artificial Intelligence. *Cogn Comput*. 2020;12(4):766–77.