Fuzzy Contrast Set Based Deep Attention Network for Lexical Analysis and Mental Health Treatment

USMAN AHMED, Western Norway University of Applied Sciences, Norway
JERRY CHUN-WEI LIN*, Western Norway University of Applied Sciences, Norway
GAUTAM SRIVASTAVA, Brandon University, Canada; China Medical University, Taiwan

Internet-Delivered Psychological Treatments (IDPT) consider the mental problems based on Internet interaction. As the increases of the pandemic, more online tools are then widely used to result in evidence-based mental health serves. This increase helps to cover more population by using fewer resources for mental health treatments. Adaptivity and customization for the remedy routine can help to solve mental health issues speedily. In this research, we propose a fuzzy contrast-based model that uses the attention network for positional weighted words, classifies mental patient authored text into distinct symptoms. After that, the trained embedding is then used to label the mental data. Then attention network expands its lexicons to adapt to the usage of transfer learning techniques. The proposed model uses similarity and contrast sets to classify the weighted attention words. The fuzzy model then uses the sets to classify the mental health data into distinct classes. The method is compared with non-embedding and traditional techniques to demonstrate the proposed model. From the experiments, the feature vector can achieve a high ROC-Curve of 0.82 with nine symptoms problems.

Additional Key Words and Phrases: Fuzzy System, Deep learning, Constraint sets, Human intervention.

1 INTRODUCTION

The COVID-19 epidemic in 93% of nations throughout the world has interrupted mental health services, according to a report by the new World Health Organization (WHO). On the other hand, mental health demand has grown as a result of the lockdown of impacted regions as a preventative precaution. Physiological anxiety factors such as dread of sickness and anxiety about the future increase during any confinement [37]. Isolation, lack of school-wide relationships, jobs add to mental stress and lead to a generally bad psychiatric treatment for the community. The absence of protection equipment, social isolation and a high-stress environment exacerbates anxiety and symptoms of depression of health professionals at the frontline level. There was a significant degree of anxiousness during a lockdown [13]. It is a reaction to life at the beginning that no one could anticipate. Since there is so much uncertainty about what causes depression, several elements are often associated with their research. Among the various facts that a wide and growing literature provides are several studies on how anxiety is dealt with. Due to the conflicting accounts, it is still difficult to extract meaningful information.

A mix of current circumstances causes depression and long-term and personal factors rather than a single immediate problem or occurrence [27]. It is not always possible to evaluate the cause or the amendment in harsh conditions [18]. It is imperative to recognize early signs and symptoms of depression and obtain assistance as quickly as possible. Many Internet fora and social media sites now allow individuals to interact anonymously

*Corresponding author.
Authors' addresses: Usman Ahmed, Western Norway University of Applied Sciences, Bergen, Norway, usman.ahmed@hvl.no; Jerry Chun-Wei Lin*, Western Norway University of Applied Sciences, Bergen, Norway, jerrylin@ieee.org; Gautam Srivastava, Brandon University, Brandon, Canada; and China Medical University, Taichung, Taiwan, SRIVASTAVAG@brandonu.ca.

2375-4699/2022/1-ART $15.00
https://doi.org/10.1145/3506701

and to talk about their misery, bereavement, and future treatment options [19]. People from all around the world can freely express their opinions and feelings [26]. Online surveillance can be a proactive and promising way to identify those issues at high risk. It can prompt mediation and improve general well-being [31].

Anxiety is one of the world's most debilitating disorders, according to the World Health Organization[1]. With over 264 million individuals afflicted globally, it has become a prevalent condition [12]. Untreated depression has the potential to worsen and cause lifetime pain [22]. Of the worst kind, anxiety can lead to suicidal thoughts. For example, there is about 800,000 people die each year by suicide according to the World Health Organization. In those aged 15 to 29 years, suicide is the second-largest cause of death. The reason is that between 76% and 85% of mentally ill people in low- and middle-income countries are untreated. Lack of financial assistant and support, lack of trained practitioners, inaccurate assessment, and society's mental stigma are an obstacle to good treatment [12]. Negative tautism, shyness and fear of disclosures are the primary obstacles preventing people from seeking treatment. People are frequently embarrassed, humiliated, and fearful of having their psychological anguish examined in-depth [29]. Because of these factors, people may be hesitant to admit that they are sad or seek mental treatment and therapy. The prevention and treatment of mental health disorders have become a global concern in the healthcare systems.

To develop an adaptive system that would reduce waiting times and deliver intervention at a reduced cost, the overburdened health care system is under pressure from economic and technological considerations. IDPT can aid a broad population with physical and psychological suffering while using fewer resources The Psychological Treatment [28]. Most solutions already in place are tunnel-based, inflexible and incompatible [27]. Existing models lack the adaptive behaviour, leading to less user adhesion and more losses [14]. Therapies should take into consideration of the several methods that consumers can get well treatment. The implementation may make this user adopt an IDPT system to consider the user's behaviour. Users express their preferences and requests according to their conditions and psychological symptoms [27].

This research aims to obtain depressed data from the written language of the patient. We then identify and visualize the results by utilizing a deep attention-based approach. The dialogue expresses the worries of a patient with mental health in the majority of cases. We next analyze the extraction of the factors causing symptoms of depression based on the statements of the patient. By employing interactive internet technology (ICT), we wish to deliver contextual information and visualization for mental health. We then analyze the removal of factors that produce symptoms associated with depression based on the comments of the patient. In addition, the designed model can help deliver the preventive steps using ICTs that provide contextual knowledge and visualization for mental health. We also utilize the NLP technique and in-depth learning to extract symptoms of anxiety and sadness from mental health therapy. After that, the semantic vectors are then used to expand the synonym to identify mental health issues. Our method helps to make the learning system generalized by minimizing the data entry tasks. From the experiments, the proposed method reached 0.82 ROC, showing that semantic vectors for synonymous expansion enhance the accuracy of training without sacrificing results.

The rest of the paper is arranged as follows. Section 2 describes the related works. Section 3 outlines the fundamental strategy for the experimentation, data collection and model development. Section 6 discusses the outcomes and findings. Section 7 concludes by providing a summary and additional work recommendations.

2 RELATED WORK

Numerous attempts to enhance depression diagnosis using computer-aided techniques have indeed been made. Fliege *et al.* [16] discussed how to measure depressive symptoms using the Item Response Theory Test (IRT) Depression-CAT and D-CAT. They developed an application to analyze depression symptoms using actual patient data, thus increasing measurement accuracy and minimizing responsive loads. Instead of using a static

[1] https://www.who.int/

questionnaire, an adaptive questionnaire was used to measure the progress [16]. The earlier data responses to queries have been utilized to select the next best questions. By asking the most pertinent questions for each patient's CAT, it was feasible to incorporate fewer items while achieving greater measurement precision throughout the whole construct range.

Lehrman *et al.* [10] discussed the method based on key linguistic text features. This work focuses on supervised classification methods and text features, which can only be used to detect states of mental effect by utilizing a limited dataset in short texts [41]. This approach, which also has a difficulty with binary classification, classifies brief sentences as disturbed or not. At this fine level, there are four textual classifications: severe pain, mild discomfort, response and pleasure. Any post that indicates an active desire to hurt anybody or yourself has been classified as high distress in the annotated set of succinctly written messages, whereas those which have only noticed negative sentiments have been labelled as moderate. In the numerous public online forums on mental well-being, the researchers evaluated the data set of 200 comments. This dataset was used for further deep learning methods including Naive Bayes, Maximum Entropy, and Decision Tree.

Online community teens stressed out utilizing layered widespread models were examined by Dinakar *et al.* [9]. They trained a set of base models for predicting labels including linear kernel support (SVM-L), a radial base kernel (SVM-R), and stochastic gradient boosted decision trees (GBDT). To form these models into categories of text in 23 different topics, text categorization was employed. For each code, a meta-function set was coupled with the SMV-L, SVM-R, and GBDT [9]. The features for the basic classifier have been used by the chi-squared feature selection and hand-coded features that include unigrams, lexicons, and part-of-speech bigrams, among others. The ratings for the decision function of each prediction were then translated into meta-functions for meta-learners and also the topic distribution for the L-LDA model. They looked at 7,147 individual tales on a prominent teenage aid website posted by concerned teens.

The behaviour of Twitter users and adolescents, in general, was studied by De Choudhury *et al.* [8] to decide whether or not they had been depressed. It seeks to create a machine learning model, thus identifying and predicting the beginnings of fear or sadness in specific individuals with a range of social media signals. The authors addressed the problem of generating an overview of the fundamental truth. Amazon Turkish mechanical annotators were born and required to complete the centre of epidemiological studies of the depression scale. There were other inquiries regarding their history and present situations, which are distressing. The Turkish Mechanics who completed the questionnaire were asked for details about their Twitter log-in used to pull the feed from Twitter. On depressed/non-depressed data, a machine learning classifier was constructed by utilizing both tweets and network features; the feature of many followers is also included. A highly favourable relationship was discovered with the statistics on anxiety control centres when the classification was applied to a large sample of US geo-located Twitter data. Research that examined over 2 million tweets among 476 users to predict depression was published. The most effective results have been obtained using the SVM classification utilizing a collection of conduction properties, Tweet responses, as well as time and frequency of postings, such as pronouns, cursing and sad phrases.

Another research used public Twitter data to explore psychological problems [5]. In addition to indications of anxiety, bipolar disorder and seasonal affective disorder, they gathered data for several mental illnesses quickly and affordably. Researchers have utilized LIWC research to evaluate how much each disease group differs from a control group. They have duplicated prior severe depression outcomes and have added new bipolar PTSD results. Two language models have been used: (1) a standard LM unigram to check every full word's likelihood and (2) a 5-g LM character in sequences up to five characters. The classifier was established to separate each group from the control group by showing the corresponding signal in the language of each group [23]. Throughout the analysis and classification, the correlations are then analyzed to identify connections and obtain insights into quantifiable and significant Twitter psychological signals. To recognize stress, Lin *et al.* [17] employed a deep neural network (DDN) to solve the limitations. Data from four microblogs have been studied, and the authors analyze the effects of their suggested four-layered DNNs. Examples of machine learning algorithms are Random Forest, SVM and Naive Bayes. They utilized three pooling techniques for each model to evaluate performance: maximum pooling, the medium over time and medium over time. Each model performed well or poorly depending on the grouping technique. DNN, by using average overtime pooling, on the other hand, achieved the best results. Neuman *et al.* [30] developed an additional methodology named "Pedesis", which employed the NLP Dependency Parsing method to crack websites that incorporate anxiety and extract improved conceptual domains into metaphorical connections. The domain knowledge is then utilized for defining words or sentences used for depression metaphors. Based on these facts, human experts have developed a "depression lexicon", which contains synonyms of the first and second grades. The vocabulary is used to autonomously evaluate the quantity of text depression and if the content deals with the subject of depression, hidden patterns and large functions are often utilized to help the neural network build a unique depiction of the area [33]. The trained network then utilizes the knowledgeable characteristics to predict the conditional input vector distribution. Indeed, for domain-specific applications, several neural network topologies are proposed. The multi-layered perception architecture is one of the major notions. Every hidden layer utilizes average output layers in this network to calculate input and weights from the previous layer. The nonlinear activation function is used on the final/output layer of the network. Thus, they modify the gradient-dependent component and the loss function. The network is needed to lower the loss of supervised education, a nonlinear issue for optimization. The weight and bias parameters are used for maximizing the loss. Most of the approaches are based on the descent process. The gradient-based techniques start with random points for each input vector. It then performs several rounds for a set of cases (batches). The loss is determined for the loss values and gradient by a trainer using the nonlinear objective function. The weights are then modified to decrease the loss function [33]. The loss is gradually reduced to the minimal level or the convergence point.

Hidden layers and the framework of architecture provide their predictive capacity to neural networks. The correct selection of many layers, architectures, layers and hyperparameters helps to tune the networks. To train the tuned network can help the input features higher-order representation of the vector [2, 7]. Higher representation of features is taught to generalize and enhance prediction ability. The network with the lowest computer complexity and best prediction capability is chosen in modern neural network research. The number of architectural ideas has increased during the previous two decades. Vijayakumar and others are the most important distinctions between the concealed layers, layer kinds, shapes, and the linkages between the layers [36]. Wainberg *et al.* demonstrated how higher-dimensional features might be extracted from tabular data through techniques of machine learning [38]. Pattern embedding accumulates from the image pixels in the neural network of convolution (CNN). The pixel information and the variation between them increase the learning and prediction capabilities of the network. The translation-invariable pixel here aids the network. The recurrent RNN architecture was designed and used for sequential data in the area of natural language processes, including machine translation, language generation and time series analysis [39]. Either an encoder or a decoder is made up of the RNN model, with the encoder sequencing the input and decoding it all into a vector with a fixed length. The model uses separate portals that depend upon the loss function to process the input attributes. Another challenge with the RNN encoder and decoder design is the alignment of input and output vectors. The sequence is dependent on the values of the neighbour. The development of a new network known as the attention mechanism is another RNN version [2, 20]. However, it uses the attention approach of the input vector by allocating weights selectively to selected inputs. The decoder may utilize the context vector orientation and related weights for the greater representation of the characteristics, which are dependent on the priority significance and position of the relevant information. For predictions, the weights of the RNN model are either learnt by the architecture and feature representation, including care weight and the context vector [20]. There are many variants of the network, including a soft, hard and global design, for such an attention mechanism. They have created the soft attention paradigm to minimize contextual information [3]. The context vector was built with the average hidden status of the model. The approach helps to understand how the input feature is concealed and to decrease the loss. Xu *et al.* [40] build the context vector under close attention using hidden state sampling. Due to the difficulties of architectural convergence, however, hard attention reduces the cost of computation. Local and global attention are other differences indicated by Luong *et al.* [21]. Global attention is the central ground between softer and harder attention. The model chooses the focal point for each input batch, which contributes to having quick convergence. This is used to learn the position of the attention vector in the local attention model with a prediction function. Techniques anticipate the place of attention. Domain-specific data analyses are needed to make two quite distinct attention on the local and global level computer-efficient.

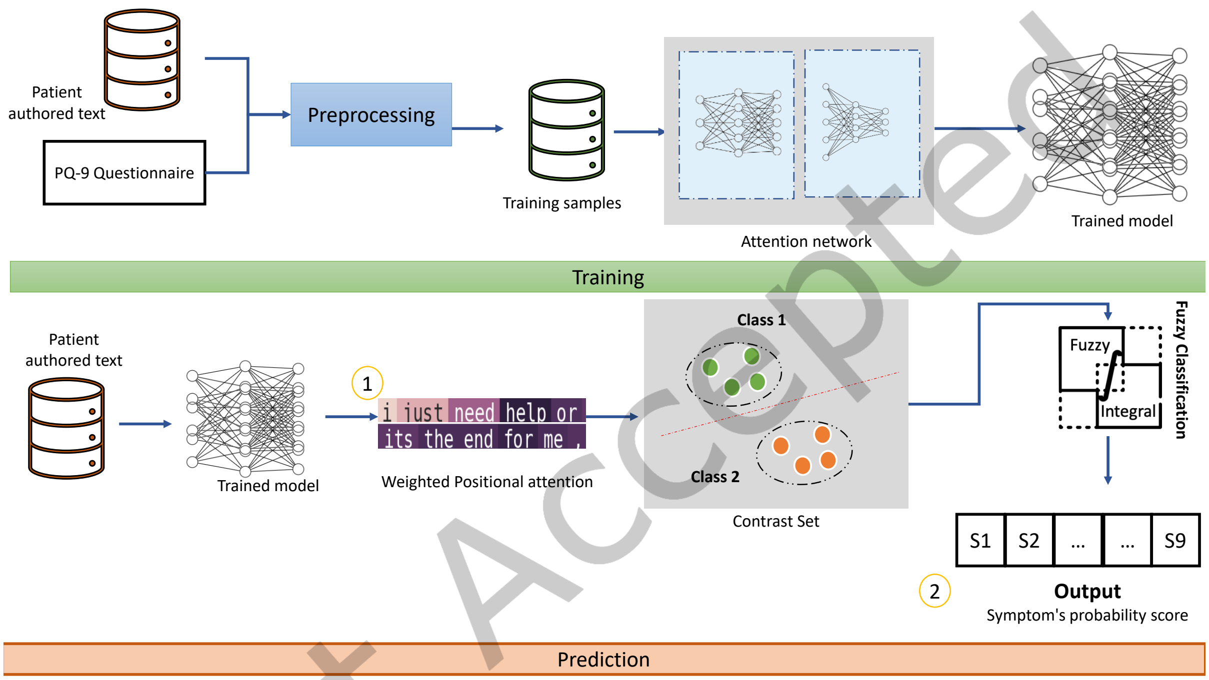

Fig. 1. Methodology of the proposed method and two output visualization with the symptom's probability score are shown in the prediction phase.

3 THE DESIGNED METHODOLOGY

As mentioned in Fig. 1, this paper proposed the embedding training approach for developing a depression symptom identification model. We applied the cosine similarity to the PHQ-9 symptoms score in this technique, as illustrated in Figs. 1 and 2. To increase knowledge and embed word size for similarity, the trained lexical enhanced approach has been proposed. The suggested approach for extracting depression symptoms from just a patient's authored text is described. The sample dataset are used from the Mukhiya *et el.* and Ahmed *et el.* work [2, 28], where data labeling is discussed in section 3.5, from sample dataset an anonymous user provides an example of a patient as follows.

"I am in a really poor spot right now. Even my melancholy and anxiety are severe, and I am unable to function or hold down a job or do anything else, so I spend my days eating junk food at home. Each day is tedious and difficult to

get it through, yet I am unable to operate in society due to my anxiety and sadness."

It is difficult to identify mental health diseases with *ICD10* classification [34]. The dynamic nature and intensity of symptoms change based on a patient treated at a certain period for a certain condition. Psychiatrists, therefore, listen to the contours of the patient and collect further essential information during the whole evaluation procedure for mental health. The method of the psychiatrist involves using a typical analytical questionnaire such as the PHQ-9 and a supported test to examine the diagnostic reliability of each evaluation against the mental health problems of the participants. In order to determine the intensity, the schemes in the survey consist of symptomatic categories, and the frequency is to offer a score based on a certain threshold. For example, nine separate questionnaires are reflected in each of the symptoms, the frequency of which the process of creation is defined as light, moderate, or severe. The method is known as the "Elicitation Process Clinical Symptom" (CSEP) [34]. One of the major aims of this study is to automate the procedure by using active learning. Each set of symptoms is categorized according to the participant's text periodicity, and clinical cumulative anxiety and depression are determined.

Table 1. Nine PHQ-9 questionnaire.

Symptons	PHQ-9
S1	Little interest or pleasure in doing things
S2	felling down depressed or hopeless
S3	trouble faling or staying asleep or sleeping too much
S4	feeling tired or having little energy
S5	poor appetite or over eating
S6	feeling bad about yourself or that you are a failure or have let yourself or your family down
S7	trouble concentrating on things such as reading the newspaper or watching television
S8	moving or speaking so slowly that other people could have noticed or the opposite being or restless that you hvae been moving around a lot more than ususal
S9	thoughts that you would be better off dead or of hurt yourself

3.1 Psychometric questionnaire (PQ)

Several additional PHQ-9 anxiety assessments are available, and PHQ-9 is among the most frequently utilized questions as proposed by Kroenke *et al.* [15]. The proposed technique for patient-authored content uses the standard PHQ-9 questionnaires [1, 2, 27]. Assessing depressed symptoms is a frequent procedure. Perhaps the psychiatrist, as part of standard CSEP practice, asks each category's inquiry and evaluates the patient's response to add the frequency to the class. As shown in Table 1, these nine symptoms can be classified into a variety of diseases including falling asleep, interest, concentrating, and eating problems. The psychiatrist determines the evaluation score after completing all of the question-based assessments. The patient's depression level is indicated by the evaluation score.

3.2 Seed Term Generation

Throughout this study, seed term generation is used for keywords found in the PHQ-9 questionnaire. This section describes how the phrase anxiety lexicon was formed (the word list of symptoms of depression). It frequently comprises words forms of emotions, i.e., anxiety and sadness, as seen in Table 1. Seed lexicons are chosen manually, and the associated hypernyms, hyponyms, and antonyms are found using Wordnet for each sentence [25]. Wordnet, an English lexicon database repository, is managed and created by Princeton University. Maybe for each word category, the database stores names, measures, descriptions and adjectives. There is a collection of synsets in each category word that are then utilized for expressing unusual notions. Synsets are divided into lexicon-based and semantic categories. Words that are part of the same synset are synonyms, for example. The top five terms are helpful and connected to the key symptom phrases, according to empirical research. In addition, only the Wordnet technique is used to expand the seed of the word in Table 1. Different categorization systems have various lists of symptoms of depression [27]. These lists use clinical or informal complaints vocabulary depending on whether the survey is a patient or physician questionnaire. Major classification systems of chronic depression, including DSM-V [2] and [34], are extensively used, and have been integrated in a fine list of symptoms [27].

3.3 Pre-processing step

The pre-processing is required to be implemented, as it requires to structure of the text. Each patient authored text follows the following procedure.

(1) All texts are processed and formatted by the UTF-8 encoding standard. One such assist in the preservation of consistency.
(2) Modify each word's capitalization to lowercase.
(3) Eliminate any tabs or spaces that may have been used to separate text.
(4) Erase all non-valued unique characters (#, +, -, *, =, HTTP, HTTPS).
(5) Substitute text-based phrases with full words, for example, can't with can't.

3.4 Lexicon embedding

A wide range of strategies for detecting emotions has been documented in the extensive NLP literature. Emotional knowledge-based (EKB) systems, on the other hand, have received much attention. EKB is made up of a vocabulary of phrase senses and a learned variety of context anchoring. Affective knowledge consists of words that convey context and emotions. We then provide an embedding strategy that uses contextually changeable words from the depression lexicon (based on word meaning) and emotional input from internet forums. We used a 300-dimensional pre-trained model for global vector for word representation (Glove) [35]. The word embedding for each word token in the patient text. The context is projected in vector space via glove-based vector embedding. Here, the embedded part represents the learned sentence structure. Perhaps the restored embedding has captured the semantic structure of the text. Each word vector is spread according to Charles's 2000 notion that *you shall recognize a word by the company it maintains* [4]. Linguistic patterns are also used to calculate the co-occurrence rates of vector representation terms. There is a comparable word nearby. As a result, the psychological analysis does not require a *pre-trained model*. We expand the dataset by training the custom mental health model with a word sense model and a transfer learning approach. This is true since a large portion of the embedding is based on open-source data (Wikipedia texts) and sentiment expertise (Twitter data). Emotions are expressed using the phrases sad and joyful. These terms, on the other hand, allude to a certain mental state. As a result, word sense must be used to broaden the embedding. The emotional lexicon, which is based on word sense, aids in demonstrating potential consequences. Custom embedding for the categorization of various symptoms can indeed be used to accomplish fine-grain classification. The words that include the part of speech are retrieved using part of speech tagging *noun, verb, adverb, and adjective*. For each retrieved part of speech, we used WordNet to extract synonyms, hyponyms, morphemes, and physical meaning from the corpus, which consists of a series of texts. As a consequence, for each document, we receive emotional words. The W set used to train the model is then utilized to build vocabulary. The learnt vector is the word vector dimension. For every nine symptoms from the PHQ-9 questionnaire, lexicons are converted into a vector using the trained model. The cosine similarity technique is to compute the similarity between patient authored text embeddings and symptoms mentioned in Table 1. Another similarity value ranging from 0 to 1 exists for each of the nine symptoms. We utilize trained to embed to convert the texts into semantically aware vectors.

[2]https://www.psychiatry.org/psychiatrists/practice/dsm

3.5 Dataset

The dataset was obtained via a discussion on the webpage and social media platforms [27]. The 500 texts were labelled using the Amazon Mechanical Turk5 service [27]. To record the remaining data, the proposed method is used. The labelling is made by the technique of PHQ-9 rating, as mentioned below.

a) score 0: not at all,
b) score 1: several days,
c) score 2: more than half the days, and
d) score 3: nearly every day.

For each symptom, we transform the annotation into a binary classifier with score 0: not at all, and score 1, 2, 3: if any symptom is present.

3.6 Deep learning methods

The LSTM network can retain the necessary information in the cell memory. The last step from the hidden state in the output layer can be delivered in an LSTM unidirectional design. During the empirical research, we noticed that the elementally average method to total time stages was superior. We utilized a bidirectional LSTM design, which received token lists from beginning to end and forward LSTM unrolling.

The proposed attention approach makes use of the text's word significance [42]. In addition to the LSTM layer, we then introduced the attention technique. This feature aids in the extraction of useful terms. The dropout layer receives the attention output vector as input. For the training of large networks, supervised learning conventionally requires a large labelled dataset. We used the transfer learning method to expand the lexical analysis and labelled the dataset.

4 CONTRAST SET

In this research, we build the lexicon by the above-mentioned method. However, using the extended lexicon is not enough for learning contextual information. We proposed attention-based contrast sets to map the continuous vector concerning the associated labelling, i.e., symptoms for patient authored texts. The contrast set helps to map the context and word concerning its labelled data points [32]. We used the concept of the support difference and co-occurrence context. We used the attention weights for each labelled data to find the classification of the patient authored texts. The attention learning methods are used for the contrast set generation and then classified the fuzzy-based model in symptoms extraction, detection and classification.

For a given attention-based lexicon, a contrast set pattern contains the user-defined frequency corresponding to different labels across the dataset. To formally define an attention-based contrast set, we proposed the pattern, support and difference definition.

Definition 1. *Attention lexicon dataset: Let $\mathcal{L} = \{l_1, l_2, \ldots, l_n\}$ be a set of lexicons and $Class = \{c_1, c_2, \ldots, c_m\}$ set of distinct labels. A attention lexicon dataset $\mathcal{D}$ contains the lexicons set that have the positional weights for the distinct $Class$, $Instances = \{(\boldsymbol{l}_w, Class_w)\}_{l=1}^{w}$, where $\boldsymbol{l}_w \subseteq Class$ is a instances contain set of attention lexicons, Class is the distinct label for $\boldsymbol{I}_w$ and w represents positional weighted words for the sentences used in the patient authored texts.*

Definition 2. *Pattern: a pattern for the lexicon X is a set of emotional lexicons $X \subseteq \mathcal{L}$ that contains the positional weighted lexicons for the distinct $Class$ i.e., symptoms.*

Definition 3. *Support: the support for the pattern X with respect to distinct label $Class$ is the percentage of the instances in L, which $Class$ contains X.*

Table 2. A set of lexicon instances with distinct classes. − represents that item is not presented in that instances.

Instances#	**Class**	$Item_1$	$Item_2$	$Item_3$	$Item_4$	$Item_5$
I_1	c1	Sad	Unhappy	depressed	-	-
I_2	c1	-	Unhappy	depressed	Helpful	-
I_3	c2	Sad	-	-	Helpful	-
I_4	c1	-	Unhappy	depressed	Helpful	devotion
I_5	c2	Sad	-	depressed	Helpful	-

Table 3. Contrast set with minimum threshold of 0.60, a set of lexicon instances with distinct classes. − represents that item is not presented within instances.

Contrast set	**Count**	**Support**	**Class-1**	**Class-2**	$Sup(instance, Class_1)$	$Sup(instance, Class_2)$	**Support-Difference**
sad	3	0.60	1	2	0.33	0.67	0.33
unhappy	3	0.60	3	0	1	0	1
unhappy, depressed	3	0.60	3	0	1	0	1
depressed, helpful	3	0.60	2	1	0.67	0.33	0.33

$$Sup(X, Class) = \frac{|\,\mathrm{Sup}(X, Class)|}{|\,\mathrm{Sup}(Class, \mathcal{L})|}, \tag{1}$$

where $\mathrm{Sup}(Class, \mathcal{L}) = \{(\boldsymbol{I}_w, Class_w) \in \mathcal{L} \mid l_i = c\}_{i=1}^{n}$ *is the set of lexicons labeled with Class, representing count of X for distinct Class. Also,* $\mathrm{sup}(X, class) = \{(\boldsymbol{I}_w, Class_w) \in \mathcal{L} \mid \boldsymbol{I}_w \in \mathrm{sup}(Class, \mathcal{L})$ *and* $X \subseteq \boldsymbol{I}_w\}_{w=1}^{n}$ *is the set of instances in* $\mathcal{L}$*, containing both pattern X and label Class. The number of distinct Classes j in* $\mathcal{L}$ *represented by the* $\sum_{i=1}^{j} |\mathrm{Sup}\ (Class_w, \mathcal{L})| = |\mathcal{L}|, j.$

Definition 4. *Support-Difference: the support difference for the pattern X with respect to label Class represented as follows:*

$$Diff(X, Class) = MAX\ \{Sup\ (X, Class_i)\}_{i=1}^{j} - MIN\ \{\mathrm{sup}\ (X, Class_i)\}_{i=1}^{j} \tag{2}$$

Example: Consider the dataset mentioned in the Table 2, this contains five instances and two *Class* i.e., c_1 and c_2 respectively. The instance I_w, $\{unhappy, depressed\}$ appears in three instances $I_1,\ I_2,\ and\ I_4$. Therefore, the support for the $\{unhappy, depressed\}$ in dataset $\mathcal{L}$ is $sup(\{unhappy, depressed\}, L) = 3/5 = 0.6$. The support for the set $\{unhappy, depressed\}$ with respect to distinct class c_1 is $sup(\{unhappy, depressed\}, c_1) = 3/3 = 1$, since three instances (i.e., the instance $\{I_1, I_2, I_4\}$) with a class c_1 contains $\{unhappy, depressed\}$. In same way, the support of $\{unhappy, depressed\}$ with respect to the second class c_2 is $sup(\{unhappy, depressed\}, c_2) = 3/0 = 0$. The detailed calculation is mentioned in the Table 3. The final contrast set are mentioned in the Table 5.

If minimum support and difference is set to $\delta = 0.6$, then $\{unhappy, depressed\}$ is said to be contrast set as $sup(\{unhappy, depressed\}, L) = 0.6 \geq \sigma$ and $\mathrm{Diff}(\{unhappy, depressed\}, L) = 1 \geq \delta$

As mentioned in Table 5, consider the contrast set $\{unhappy\}$, we have the similarity context $\mathcal{N}_S(b) = \{unhappydepressed\}$ as both words share the emotional meaning of unhappy. The co-occurrence metric for first instance is $(Sad,\ unhappy, (unhappy, depressed))$ since *unhappy* co-occurs with *sad* in instance I_1, *unhappy* co-occurs with $\{unhappy,\ depressed\}$ in instance $I_{1,2,4}$, and *unhappy* co-occurs with $\{depressed,\ helpful\}$ in instance $I_{2,4}$. The similarity and co-occurrence help to measure the contrast sets as accurate based on the co-occurrence and their individual instance lexicon sets.

With the above example, we describe how the co-occurrence context helps capture the contrast sets. We now discuss the usefulness of the fuzzy sets as mentioned in Table 4.

Table 4. The contrast set with the contexts.

Instances#	Class	Items			
I_1	c_1	Sad	Unhappy	unhappy, depressed	
I_2	c_1	Unhappy		Unhappy, depressed	depressed, helpful
I_3	c_2	Sad			
I_4	c_1	Unhappy		Unhappy, depressed	depressed, helpful
I_5	c_2	Sad		depressed, helpful	

Algorithm 1 Contrast set fuzzy classification

Input: Lexicon attention based model position weighted words.
Output: Contrast set with fuzzy classification rules.
1: $Lexicons \leftarrow Lexicon(attention_network)$;
2: $Sets \leftarrow Contrast_{set}(Lexicons)$;
3: **Fuzzification:** Convert the support difference values (crisp input) into fuzzy data;
4: **Fuzzy Rules generation:** Generate rules by using the fuzzy data;
5: **Defuzzification:** Convert the fuzzy rules into crips rules;
6: Apply contrast set method to improve the fuzzy inferences;
7: Crossvalidate the model using testing data;
8: Perform statistical test to validate contrast set fuzzy inference rules;
9: **Return:** Inference rules.

5 FUZZY INFERENCE SYSTEM

A fuzzy knowledge supports the `if-then` rules to denote the relationship of input and output. The methods involve developing fuzzy rules, fuzzification, inference rules generation and defuzzification to crisp outputs. The membership values convert the data into membership degrees among 0 and 1. We used the Triangular membership function for modifying the contrast set support difference values to fuzzification [1]. The method is mentioned in the following equation 3.

$$f(x) = \begin{cases} 0 & \text{if } x \leq= i \\ \frac{x-i}{j-i} & \text{if } i \leq x \leq j \\ \frac{k-x}{k-j} & \text{if } j \leq x \leq k \end{cases} \tag{3}$$

5.0.1 Fuzzy rule generation. The rules from contrast set and fuzzification values are very important as they help classify the attention positional weighted elements. The itemset after contrast set generation can help to classify into linguistic rules. The lexicon and assign classes help to create many rules.

5.0.2 Defuzzification. The testing data is fed to the fuzzification steps. Based on the membership function values, the fuzzified input is matched with the inference rules. The inferences rules are obtained from the linguistic values, converted into a fuzzy score using the weighted method. From the fuzzy score, a classification decision is produced. The flow of the operation is mentioned in Algorithm 1

6 EXPERIMENTAL RESULT AND ANALYSIS

The patient authored text is pre-processed and converted into an emotional-based lexicon. Then we trained different networks. We used the pre-trained network Glove for the transfer learning task. The trained embedding

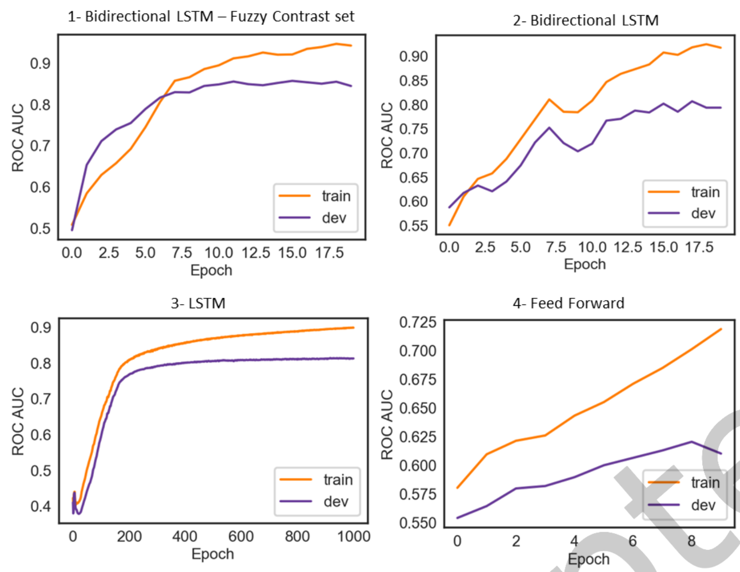

Fig. 2. Training and development set ROX AUC Curve comparisons.

is expanded with a new lexicon. Then the text model is converted into the nine symptom vector lexicon. After that, vectors for the patient questionnaire and patient authored text are used to label the unlabeled data. The labelled data is then trained and compared with different architectures. We used the ROC curve, precision, recall and F-measure performance metrics. We used the Adam optimizer to reduce the training loss. Figs. 2 and 3 showed the performance of attention network [42] with fuzzy contrast set, bidirectional LSTM [24], LSTM [11] and feed-forward network [2]. For deep neural architecture, we used to change the cell type and hidden size. In addition to the LSTM layer, we used the contrast set with a fuzzy classifier to improve model performance. During empirical analysis, models showed the overfitting issue on the development and testing set. To handle these issues, we perform the model for a longer time, i.e., 1000 epochs. We also employ the early stopping method to save and tune model processing. We also used the clipping method to avoid gradient issues [6].

Figs. 2 and 3 showed the performance of the feed-forward network where the training loss reached 0.41 and the testing loss 0.59. The model tends to overfit and close to the upper left corner. The architecture did not perform well as the sequential data do not preserve the sequence in the simple network. The sequential models can handle the sequential data and achieve good performance. Thus, we performed the LSTM network and we have reached the ROC value of 0.78. The recall of LSTM model is mentioned in Fig. 4 and the precision is mentioned in Fig. 5. The model has suffered from vanish gradient issues. The cell becomes complex to the complex gates. The architecture required to be more tuner. The instances with symptoms related to questionnaire two performed well.

In bidirectional with contrast fuzzy set and bidirectional LSTM with attention, they can have better performance. The model used the two-directional approach from a backward and forward pass. Both hidden gates help to preserve sequential order. The attention layer helps to weigh the positional words. As a result, the trained model results in the lowest error. The recall curve at the top corner represents that model has a low false-positive and false-negative rate. The attention layer weights positional words that help with contrast set and fuzzy classifier. The model has high performance with the lowest development set error.

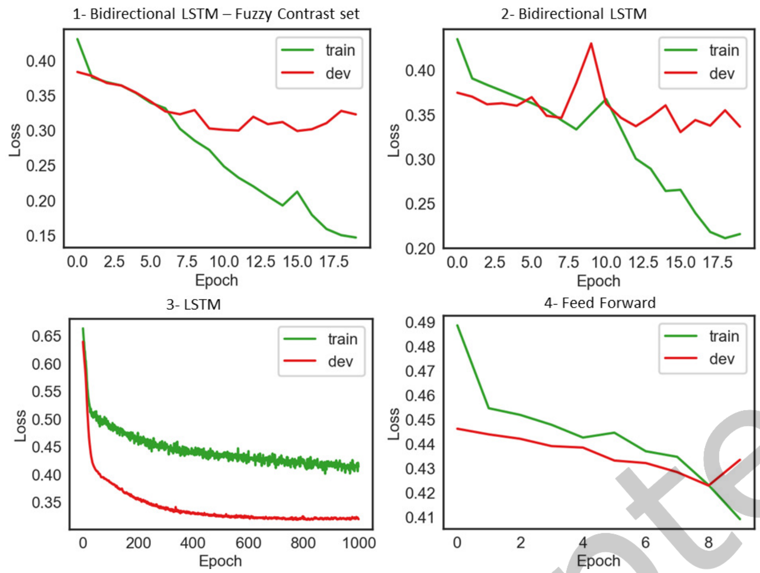

Fig. 3. The Loss analysis of the training and development set.

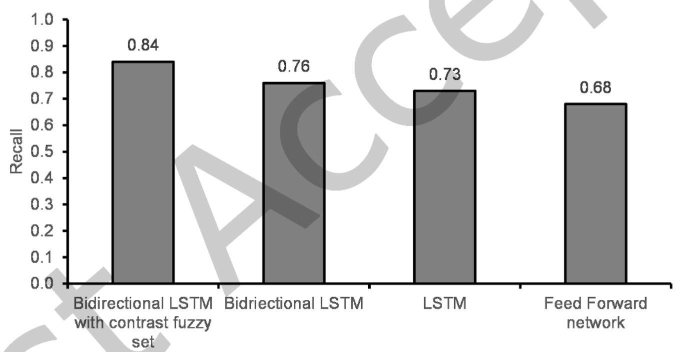

Fig. 4. Recall based comparison with different architectures.

The model achieved ROC 0.91 of the training set and 0.82 development set. The high performance indicates that the model results in a high positive rate. The results support the existence of important words. This helps contrast set to generate distinct boundaries. The model recognizes the target word in the task, and it learned the symptoms of the patient's authored text.

In Figure 7 and Table 5, the proposed model is able to visualise contrast set weights for the quoted words in sentences of patient authored text. The visualised symptom and probability score *[s4: 046, s8: 0.45, s2: 0.39 and s1: 0.26]* reflects the context and patient triggering point. The patient is struggling with *feeling tired or having little energy* and *moving or speaking so slowly that other people . . . . a lot more than ususal* that indicates that the source of the contrast set is the words highlighted dark i.e. *tedious and difficult to get it through* and *my anxiety and sadness*. Therefore, the model can find the contrast set words, highlight them according to probability and symptoms scores.

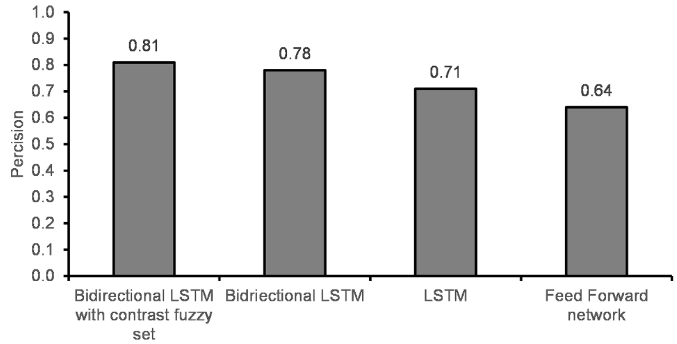

Fig. 5. Precision based comparison with different architectures.

Fig. 6. F1-measure based comparison with different architectures.

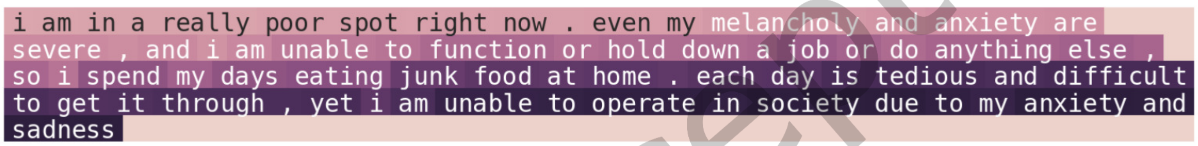

Fig. 7. A explainable depression symptoms extracted by proposed approach.

Table 5. The explain-ability results.

Sorted Sympton	Description	Probability
s4	feeling tired or having little energy	**0.46**
s8	moving or speaking so slowly that other people could have noticed or the opposite being or restless that you hvae been moving around a lot more than ususal	**0.45**
s2	felling down depressed or hopeless	**0.39**
s1	Little interest or pleasure in doing things	**0.26**
s5	poor appetite or over eating	0.15
s3	trouble faling or staying asleep or sleeping too much	0.13
s7	trouble concentrating on things such as reading the newspaper or watching television	0.12
s6	feeling bad about yourself or that you are a failure or have let yourself or your family down	0.02
s9	thoughts that you would be better off dead or of hurt yourself	0.0097

7 CONCLUSION

The tool for NLP and deep learning for health care intervention has been introduced recently. The pandemic era forced treatment of psychological patients by using the online medium. A limited study works on the mental health symptoms. However, the adoption of the model for mental health is not well discussed. This paper used the contrast set with the fuzzy model to classify mental health patients into nine distinct classes. We proposed the support difference contrast set lexicon analysis. The attention network uses that to fuzzify the input. Then contrast inference rules are used to classify the mental health treatment test. The fuzzy rules can be used for the labelling and visualization task. This tool can help psychiatrists to make the customization and appropriate program for the remedy. The computer-aided system helps in highlighting the keywords and helps to adapt and give visualization. The LSTM model with attention and contrast is set to achieve the highest accuracy. The model achieved 0.82 ROC and helped to visualize the weighted words. The weighted words can help to understand the patient's issues. In future, we will try to implement a more adaptive algorithm to classify the text and reduce overfitting issues.

REFERENCES

[1] Usman Ahmed, Jerry Chun-Wei Lin, and Gautam Srivastava. 2021. Fuzzy Explainable Attention-based Deep Active Learning on Mental-Health Data. In *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 1–6.

[2] Usman Ahmed, Suresh Kumar Mukhiya, Gautam Srivastava, Yngve Lamo, and Jerry Chun-Wei Lin. 2021. Attention-Based Deep Entropy Active Learning Using Lexical Algorithm for Mental Health Treatment. *Frontiers in Psychology* 12 (2021), 471.

[3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *The International Conference on Learning Representations*, Yoshua Bengio and Yann LeCun (Eds.). ICLR, 24–34.

[4] Walter G. Charles. 2000. Contextual correlates of meaning. *Applied Psycholinguistics* 21, 4 (2000), 505–524.

[5] Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance* 6, 2 (2020), e19273.

[6] Xiangyi Chen, Steven Z Wu, and Mingyi Hong. 2020. Understanding gradient clipping in private SGD: A geometric perspective. *Advances in Neural Information Processing Systems* 33 (2020), 1–10.

[7] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *The Conference on Empirical Methods in Natural Language Processing*, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). EMNLP, 1724–1734.

[8] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media. In *Proceedings of the Seventh International Conference on Weblogs and Social Media.* AAAI, 23–34.

[9] Karthik Dinakar, Emily Weinstein, Henry Lieberman, and Robert Louis Selman. 2014. Stacked Generalization Learning to Analyze Teenage Distress. In *Proceedings of the Eighth International Conference on Weblogs and Social Media.* AAAI, 23–34.

[10] Herbert Fliege, Janine Becker, Otto B Walter, Jakob B Bjorner, Burghard F Klapp, and Matthias Rose. 2005. Development of a computer-adaptive test for depression (D-CAT). *Quality of life Research* 14, 10 (2005), 2277.

[11] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. 2016. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28, 10 (2016), 2222–2232.

[12] Spencer L James, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. 2018. 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. *The Lancet* 392, 10159 (2018), 1789–1858.

[13] Christian Karmen, Robert C. Hsiung, and Thomas Wetter. 2015. Screening Internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods. *Computer Methods and Programs in Biomedicine* 120, 1 (2015), 27–36.

[14] Artie Konrad, Victoria Bellotti, Nicole Crenshaw, Simon Tucker, Les Nelson, Honglu Du, Peter Pirolli, and Steve Whittaker. 2015. Finding the Adaptive Sweet Spot. In *Proceedings of the 33rd Annual Conference on Human Factors in Computing Systems.* ACM, 3829–3838.

[15] Kurt Kroenke, Robert L. Spitzer, and Janet B.W. Williams. 2001. The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine* 16, 9 (2001), 606–613.

[16] Tim M. H. Li, Michael Chau, Paul W. C. Wong, and Paul S. F. Yip. 2012. A Hybrid System for Online Detection of Emotional Distress. In *Intelligence and Security Informatics.* Springer, 73–80.

[17] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. 2014. User-level psychological stress detection from social media using deep neural network. In *Proceedings of the 22nd ACM international conference on Multimedia.* ACM, 507–516.

[18] David E. Losada and Pablo Gamallo. 2018. Evaluating and improving lexical resources for detecting signs of depression in text. *Language Resources and Evaluation* 54, 1 (2018), 1–24.

[19] Daniel M Low, Laurie Rumker, Tanya Talkar, John Torous, Guillermo Cecchi, and Satrajit S Ghosh. 2020. Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of medical Internet research* 22, 10 (2020), e22635.

[20] Jiasen Lu, Jianwei Yang, Dhruv Batra, and Devi Parikh. 2016. Hierarchical Question-Image Co-Attention for Visual Question Answering. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems*, Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.). NIPS, 289–297.

[21] Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *The Conference on Empirical Methods in Natural Language Processing*, Lluís Màrquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton (Eds.). Stanford, 1412–1421.

[22] Mario Gennaro Mazza, Rebecca De Lorenzo, Caterina Conte, Sara Poletti, Benedetta Vai, Irene Bollettini, Elisa Maria Teresa Melloni, Roberto Furlan, Fabio Ciceri, Patrizia Rovere-Querini, and Francesco Benedetti. 2020. Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors. *Brain, Behavior, and Immunity* 89 (2020), 594–600.

[23] Michelle McDonnell, Jason Edward Owen, and Erin O'Carroll Bantum. 2020. Identification of Emotional Expression With Cancer Survivors: Validation of Linguistic Inquiry and Word Count. *JMIR Formative Research* 4, 10 (2020), e18246.

[24] Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning generic context embedding with bidirectional lstm. In *Proceedings of the 20th SIGNLL conference on computational natural language learning*. ACL, 51–61.

[25] George Miller, Christiane Fellbaum, Judy Kegl, and Katherine Miller. 2009. WordNet: An Electronic Lexical Reference System Based on Theories of Lexical Memory. *Revue québécoise de linguistique* 17, 2 (2009), 181–212.

[26] Julia Mühleck, Sigrid Borse, Eva Wunderer, Bernhard Strauß, and Uwe Berger. 2019. Online-Befragung zur Bekanntheit von Angeboten zur Aufklärung, Prävention, Beratung und Nachsorge bei Essstörungen. *Prävention und Gesundheitsförderung* 15, 1 (2019), 73–79.

[27] Suresh Kumar Mukhiya, Usman Ahmed, Fazle Rabbi, Ka I Pun, and Yngve Lamo. 2020. Adaptation of IDPT System Based on Patient-Authored Text Data using NLP. In *IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 27–36.

[28] S. K. Mukhiya, J. D. Wake, Y. Inal, and Y. Lamo. 2020. Adaptive Systems for Internet-Delivered Psychological Treatments. *IEEE Access* 8 (2020), 112220–112236.

[29] Suresh Kumar Mukhiya, Jo Dugstad Wake, Yavuz Inal, Ka I Pun, and Yngve Lamo. 2020. Adaptive Elements in Internet-Delivered Psychological Treatment Systems: Systematic Review. *Journal of Medical Internet Research* 22, 11 (2020), e21066.

[30] Yair Neuman, Yohai Cohen, Dan Assaf, and Gabbi Kedma. 2012. Proactive screening for depression through metaphorical and automatic text analysis. *Artificial Intelligence in Medicine* 56, 1 (2012), 19–25.

[31] Antoine Neuraz, Ivan Lerner, William Digan, Nicolas Paris, Rosy Tsopra, Alice Rogier, David Baudoin, Kevin Bretonnel Cohen, Anita Burgun, Nicolas Garcelon, et al. 2020. Natural language processing for rapid response to emergent diseases: Case study of calcium channel blockers and hypertension in the COVID-19 pandemic. *Journal of medical Internet research* 22, 8 (2020), e20773.

[32] Dang Nguyen, Wei Luo, Bay Vo, Loan TT Nguyen, and Witold Pedrycz. 2021. Con2Vec: Learning embedding representations for contrast sets. *Knowledge-Based Systems* 229 (2021), 107382.

[33] Giang Nguyen, Stefan Dlugolinsky, Martin Bobák, Viet D. Tran, Álvaro López García, Ignacio Heredia, Peter Malík, and Ladislav Hluchý. 2019. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review* 52, 1 (2019), 77–124.

[34] World Health Organization et al. 1993. *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. Vol. 2. World Health Organization.

[35] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *The Conference on Empirical Methods in Natural Language Processing*. ACL, 1532–1543.

[36] Vivienne Sze, Yu Hsin Chen, Tien Ju Yang, and Joel S. Emer. 2017. Efficient processing of deep neural networks: a tutorial and survey. *Proc. IEEE* 105, 12 (2017), 2295–2329.

[37] Emily A. Troyer, Jordan N. Kohn, and Suzi Hong. 2020. Are we facing a crashing wave of neuropsychiatric sequelae of COVID-19? Neuropsychiatric symptoms and potential immunologic mechanisms. *Brain, Behavior, and Immunity* 87 (2020), 34–39.

[38] Michael Wainberg, Daniele Merico, Andrew Delong, and Brendan J Frey. 2018. Deep learning in biomedicine. *Nature biotechnology* 36, 9 (2018), 829–838.

[39] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, et al. 2016. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *CoRR* abs/1609.08144 (2016), 1609.08144.

[40] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. 2015. Show, attend and tell: neural image caption generation with visual attention. In *The International Conference on Machine Learning (JMLR Workshop and Conference Proceedings, Vol. 37)*, Francis R. Bach and David M. Blei (Eds.). MLR, 2048–2057.

[41] Lingling Xu, Ruyi Jin, Feifei Huang, Yanhui Zhou, Zonglong Li, and Minqiang Zhang. 2020. Development of Computerized Adaptive Testing for Emotion Regulation. *Frontiers in Psychology* 11 (2020), 3340.

[42] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, 1480–1489.