Transactions on Asian and Low-Resource Language Information Processing

Explainable Deep Attention Active Learning for Sentimental Analytics of Mental Disorder

Journal:	*Transactions on Asian and Low-Resource Language Information Processing*
Manuscript ID	TALLIP-22-0135
Manuscript Type:	SI on New Trends in Machine Translation and Technology for Low-Resource Languages
Date Submitted by the Author:	04-Feb-2022
Complete List of Authors:	Ahmed, Usman; Western Norway University of Applied Sciences Jhaveri, Rutvij H.; Pandit Deendayal Energy University Srivastava, Gautam; Brandon University, Lin, Jerry Chun-Wei; Western Norway University of Applied Sciences,
Keywords:	Internet delivered interventions, adaptive treatments, word sense identification, sentimental analysis, text clustering

Explainable Deep Attention Active Learning for Sentimental Analytics of Mental Disorder

USMAN AHMED, Western Norway University of Applied Sciences, Norway
RUTVIJ H. JHAVERI, Pandit Deendayal Energy University, India
GAUTAM SRIVASTAVA, Brandon University, Canada; China Medical University, Taiwan
JERRY CHUN-WEI LIN*, Western Norway University of Applied Sciences, Norway

With the increasing use of the online medium, Internet delivered psychological treatments (IDPs) are becoming an important tool for improving mental disorders. Online based health therapies can help a large segment of the population with little resource investment. The task is greatly complicated by the overlapping emotions for certain mental health. Early adoption of a deep learning system presented serious difficulties, including ethical and legal considerations that contributed to a lack of trust. Modern models required highly interpretable, intuitive explanations that humans could understand. To achieve this, we present a deep attention model based on fuzzy classification that uses the linguistic features of patient texts to build emotional lexicons. In medical applications, a diversified dataset generates work. Active learning techniques are used to gradually extend fuzzy rules as well as the learned dataset. From this, the model is able to gain reduction in labeling efforts in mental health applications. In this way, difficulties such as the amount of vocabulary per class, method of generation, the source of data, as well as the baseline for human performance level can be solved. Moreover, this work illustrates fuzzy explainability by using weighted terms. The proposed method incorporates a subset of unstructured data into the set for training as well as uses a similarity based approach. The approach then updates the model training using the new training points in the subsequent cycle of the active learning mechanism. The cycle is repeated until the optimal solution is found. At this point, all unlabeled text is converted into the set for training. The experimental results show that the emotion based enhancement improves test accuracy as well as helps in developing quality criteria. In the blind test, the bidirectional LSTM architecture with an attention mechanism as well as fuzzy classification achieved an F1 score of 0.89.

Additional Key Words and Phrases: Internet delivered interventions, adaptive treatments, word sense identification, sentimental analysis, text clustering

ACM Reference Format:
Usman Ahmed, Rutvij H. Jhaveri, Gautam Srivastava, and Jerry Chun-Wei Lin*. 2021. Explainable Deep Attention Active Learning for Sentimental Analytics of Mental Disorder. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, Article 1, 17 pages. https://doi.org/10.1145/3506701

1 INTRODUCTION

Artificial Intelligence in Medicine (AIM) has the potential to improve the practice of medicine by enabling physicians as well as other medical professionals to analyze various data more naturally as well as accurately [40]. Remote healthcare entered the era of the Internet of things (IoT) applications. Tracking as well as analysing healthcare's content helps to facilitate the adoption of high quality healthcare services. Assessing patients health with IoT devices is a need of time

*Corresponding author.

as well as its play a vital role in managing health conditions. Biomedical devices can help to improve the informatics research community to solve complex healthcare as well as improve human health. Modern data driven technologies can help clinicians interpret large volumes of diverse medical data as well as make accurate diagnoses as well as treatment decisions for patients [18].

With the advert of the Covid 19, research on the internet delivered healthcare facility was carried out on a commercial basis. The idea of the research is to monitor healthcare without direct contact with the physicians. Thus, this helps the pandemic situation as well as reduced the loaded health facilities. This tends to increase the requirement of mobile healthcare that required real time systems to assist the physician during the mobile healthcare solutions. Artificial intelligence (Artificial Intelligence) tools are helping as well as may be able to reduce physician workload during the current pandemic [21]. Through machine learning, Artificial Intelligence attempts to replicate human behaviors using labeled data (ML). AIM is integrated with the electronic health record (EHR) system used by the majority of healthcare organizations worldwide [7]. Data collected by EHR systems enable the development of more practical Artificial Intelligence applications. In general, an electronic health record includes both structured as well as unstructured data (e.g., patient demographics, diagnoses, as well as procedures) as well as unstructured data (e.g., physician notes as well as clinical reports). This makes it difficult to interpret EHR based data because the data is inconsistent.

Despite promising research efforts at AIM, clinical application is hampered by interpretability issues. Numerous studies have shown that healthcare professionals lack confidence in AIM [10]. Major causes of trust issues include recognized limitations on access to medical data, a lack of expertise needed to integrate clinical workflows, regulatory limitations, as well as development and deployment hurdles that AIM systems have experienced internationally. The Defense Advanced Research Projects Agency (DARPA) announced its Explainable Artificial Intelligence (XAI) research program in 2017 [13]. According to DARPA, modern artificial intelligence systems have limitations in terms of interpretation as well as communication. This is because humans are unable to understand how Artificial Intelligence systems work, i.e., why an Artificial Intelligence can be able to make for the particular strategy and decision.

Researchers also argue that physicians can easily rely on pharmaceuticals like aspirin, despite actually understanding the drugs interaction mechanisms fully [46]. If an Artificial Intelligence system started explaining the decision making by compromising the performance, will physicians accept that system? On the other hand, drug regulatory authorities ensure that any drug undergoes a rigorously designed, randomized conducted clinical trial before being permitted for public use (COVID 19 Vaccines as a recent example). Additionally, regulatory agencies also perform post marketing surveillance, such as the Food and Drug Administration in the United States of America (USA). The FDA's main purpose for pharmaceuticals is to quickly withdraw drugs from the market in case of unexpected side effects, fatality being the worst of these. Artificial Intelligence systems built under limited datasets may have potential biases due to generalization issues new samples. Therefore, Artificial Intelligence systems do not have a similar mechanism to take full responsibility for safety issues as well as efficacy. XAI can help to understand whether AIM decisions are valid as well as get consensus from medical experts. XAI promotes trust in AIM by acting as a decision support systems [15]. The explainability plays a vital role in supporting the implementation of Artificial Intelligence in clinical decision making [15, 44].

There are several data science applications for fuzzy based systems, including human reasoning as well as decision models. By combining cooperative fuzzy principles with system intelligence, fuzzy modeling can contribute to the understanding of Artificial Intelligence decisions, specifically in the area of explainability as well as interpretability for the accuracy performance of machine learning models. By combining machine learning as well as fuzzy modeling approaches, researchers have begun to address the challenge of making artificial intelligence models explainable to humans. The goal of Artificial Intelligence systems is to maximize accuracy while providing users with interpretable conclusions as well as results. With various unstructured data available, Artificial Intelligence systems can be optimized for better performance. However, fuzzy based systems retain their fundamental nature of comprehensible intelligibility, as they can enhance the performance of Artificial Intelligence models as well as help construct explainable artificial intelligence for humans.

There are several data science applications for fuzzy based systems, including human reasoning as well as decision models. By combining cooperative fuzzy principles with system intelligence, fuzzy modeling can contribute to the understanding of Artificial Intelligence decisions, specifically in the area of explainability as well as interpretability for the accuracy performance of machine learning models. By combining machine learning as well as fuzzy modeling approaches, researchers have begun to address the challenge of making artificial intelligence models explainable to humans. The goal of Artificial Intelligence systems is to maximize accuracy while providing users with interpretable conclusions as well as results. With various unstructured data available, Artificial Intelligence systems can be optimized for better performance. However, fuzzy based systems retain their fundamental nature of comprehensible intelligibility, as they can enhance the performance of Artificial Intelligence models as well as help construct explainable artificial intelligence for humans.

According to the World Health Organization (WHO), depression is one of the most debilitating diseases. [17]. To date, approximately 264 million cases of depression have been recorded, with nearly as large a proportion likely to go undiagnosed or untreated [17]. Depression is often misdiagnosed as well as neglected. In addition, depression is a leading cause of suicide. According to WHO, nearly 800,000 people die each year as a result of suicide [17]. People between the ages of 15 and 29 committing suicide is the second leading cause of death. Between 76 and 85 percent of people in low- and middle-income countries do not receive treatment for their illnesses. Key difficulties include lack of resources, inadequate training of health workers, as well as flawed assessments [17]. People tend to be ashamed as well as fear having an examination of their psychological pain [33].Also, many people are reluctant to admit that they are depressed or seek therapy for it. The overburdened health care system is facing pandemic pressures and economic issues. This burden can be reduced by having an adaptive system that is explainable to treat patients more efficiently. This will reduce waiting times for treatment and provide intervention at a faster rate as well as reduced cost. Internet delivered Psychological Treatment (IDPT) can reduce physical distress for a large population using reduces set of resource [32].

1.1 Motivation

The COVID 19 pandemic in 2 years has also increased mental health issues globally. According to the report from WHO[1], 93% of countries worldwide have experienced an increase in mental health issues. The physiological stress factor for people has increased due to lockdown, which has also included fears of illness as well as uncertainty of the future of humankind [43]. Social isolation, lack of interactive activities, educational uncertainty, as well as irregular work hours have resulted in a much higher emotional stress rate. Health workers have also been suffering from anxiety as well as depressive symptoms due to fear of illness, lack of protective equipment, social disconnections, as well as working in a high stress environment. Overall, Depression during the pandemic is significantly more than anyone could have expected [19]. As the world moves toward online systems in many different areas, numerous Internet forums as well as social media platforms enable individuals to interact with each other [25]. Members of our global society have become accustomed to interacting online due to the pandemic [30]. Online depression detection can help to determine

[1] shorturl.at/enwTV

the high risk people with the issue of mental health. The timely medication can then help to improve general well being [35].

The aim of this work is to extract depressive symptoms from writings written by patients. To increase accuracy, our technique uses a fuzzy categorization model. The proposed framework enables the identification as well as visualization of depressive symptoms by using a deep attention based method. In a clinical psychiatrist's practice, the patient will communicate or use an online method to express their psychological problems (specifically in the current pandemic era). We then categorize emotional text elements into depression related symptoms to extract depressive components. By using fuzzy rules, our fuzzy categorization model helps psychiatrists as well as other medical professionals in interpretation. The framework acts as a decision support system as it incorporates entropy based labeling of data, attention based Deep Learning embedding, as well as an interpretable fuzzy model. Our online interactive tool (ICT) helps provide context as well as visualization of detected depressive symptoms related to mental health. We evaluated different architecture design for the attention mechanism as well as combined them with fuzzy rules to increase explanation based on the attention word frequency. This study directly addresses the interpretability of a deep learning model by using learned embeddings as input features for a fuzzy classifier in the deep learning scheme.

2 RELATED WORK

Many works as well as studies have been developed recently making use of NLP based methodologies in computer aided systems. We provide a brief summary of the related literature in this section

Li *et al.* [24] implement a Item Response Theory based Computer Adaptive Tests (CAT), that is able to help in the detection of depression. The authors system is able to achieve a very high accuracy making use of adaptive questionnaires instead of static questionnaires. That being said, interpretation as well as explainability of the model are still unknown, in that we do not know why the model selects certain aspects of the questionnaire or how varying the response options can influence the respondents behaviour.

Recently, a great deal of research as well as studies have been conducted using natural language processing (NLP) techniques in digital systems. We provide a brief summary of work that has been studied in the area.

Herbert *et al.* [12] analyzed short texts for linguistic text features to detect depression distress. Their designed model is then performed using a binary classification task using ML as well as NLP. All of the texts that are short are given the classification of "distressed" or "non-distressed". 4 text classes are created in a fine grained level which are happiness, response, low distress, as well as finally high distress. The dataset that was used is made up of 200 posts that are from public forums that deal with depression as well as anxiety. Decision Tree, Naïve Bayes, as well as Maximum Entropy were conducted as well as applied on the given dataset, respectively. From the experimental results, their model can achieve 54.5% accuracy when classifying four ways of distress level.

Chen *et al.* [4] examine health data for mental health from Twitter to aid in the diagnoses of mental illnesses such as SAD, bipolar disorder, as well as depressed. LIWC or Linguistic Inquiry Word Count is made use to compare the experimental group to the control group [4, 28]. Two language models are used, for example, (1) a unigram model to analyze the probability of words, and (2) a character based 5 gram model to examine 5 character sequences. Compared to these classifiers, the trained classifier separates each group from the control group as well as exhibits a valuable signal in the language of each group. The associations between the results of the statistical analyzes as well as the classifiers were then examined to determine the relationships between measurable as well as meaningful signals of mental health on Twitter [28].

Patel *et al.* [38] study health diagnosis on Twitter for diabetes. Motivated by the fact that social networks have as well as are a common venue to share experiences as well as information about health, the authors make use of the Twitter API to discover trends in Twitter threads to aid in the diagnosis of diabetes. The authors learned that patterns exists that can aid in the diagnosis of other diseases as well based on the discussion of health issues on Twitter.

Neuman *et al.* [34], crawled depression metaphorical relations from websites. Their model utilized a dependency tree to increase the prediction quality. A depression lexicon is constructed through the use of metaphorical phrases and then expanded through the use of synonymous first and second degree terms. The depression lexicon was developed to evaluate the degree of depression in various texts. A statistical based method lacks generalization as a model trained on one dataset might not perform better for different distributions.

Li *et al.* [24] show CAT based on Item Response Theory (IRT) that can be used to diagnose depressive symptoms. By using an adaptive questionnaire instead of a static questionnaire, the technique achieves a high degree of accuracy. However, the explanatory power as well as interpretation of the model remain unexplained, i.e., why the model selects this questionnaire or how changing response options within a single test affects response behavior.

Nguyen *et al.* [36] investigate a distinct representation of the feature space in a neural network. The learned features represent the conditional probability distribution of input vectors. The number of architecture is proposed for domain specific applications. One network is a multi layer perception network; each hidden layer averages the previous layer as well as weights. The function for activation is used on the final stage or output layer. The gradient approach is used to update the weights by using the function for loss. In their supervised learning method, the neural network requires to reduce the loss. The non linear optimization problem uses the weight as well as bias values to achieve it. This method is based on the gradient descent technique. The gradient based techniques start with random initialization for each input vector. It then runs for several epochs or iterations for the set of the batch instances. The function for loss computes the loss by using the gradient as well as prediction from the neural network. After that, weights are updated, as well as loss is reduced. This process continuously performs until it reaches the convergence point. The prediction power of the deep neural networks is coming from the hidden layers as well as their structure alignments. The selection of correct architecture improves accuracy, as well as majorly tuning is performed on the layer structuring as well as hyperparameters to solve complex problems. The learned higher feature representation helps to achieve generalization as well as increase predictive power.

In modern neural network research, networks with low computational complexity as well as high predictive capacity are preferred. As can be seen in [42], neural networks are classified according to the number of hidden layers, the type of layers, their shape, as well as the connections between them. Wainberg *et al.* [45] have introduced the z technique for extracting higher dimensional features from tabular data. Their Convolutional Neural Network (CNN) learns how to incorporate features from image pixels. The network benefits from translation invariant pixels. In [9], a recurrent neural network (RNN) architecture was developed as well as applied to NLP applications with sequential data. The RNN model is an encoder decoder framework in which the encoder takes the input sequence as well as decodes it into a fixed length vector. Cho *et al.* proposed a model that uses different gates to process the input data depending on the function for loss. Alignment of RNN encoder as well as decoder models is an area of research in the field of RNN. The value of the neighbor feature may have an impact on the sequence. Therefore, in another form of RNN, the creation of a new network called an attention mechanism is proposed. By assigning weights to specific inputs, it focuses attention on the input vector. Based on this selection, it determines the position of the relevant information. The decoder then uses the position with the appropriate weights as well as corrected vectors for the higher feature representation. The RNN makes predictions using the learned weights [6, 26]. There are several variants of the attention architecture, including a

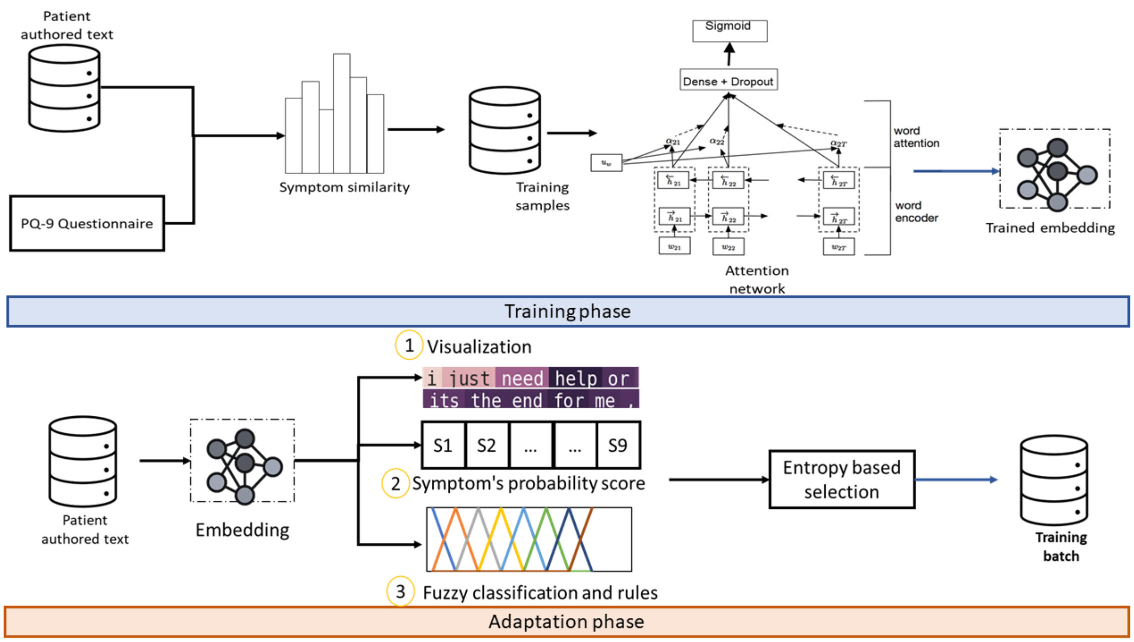

Fig. 1. Attention-based domain adaptation is used to create a flow of training and adaptation. The visualization, symptoms' probability score, fuzzy classification and triggered rules are suggested to the psychiatrists that will be provided to patient for certain symptoms.

soft, a hard, as well as a global design for the attention mechanism. In soft attention, the model [3] is initiated first, followed by the average of the hidden ones, as well as finally the context vector is constructed.

Xu *et al.* constructed the context vector at hard attention by sampling the hidden states [47]. Luong *et al.* argue for both local as well as global attention [27]. Their paradigm of global attention is a synthesis of soft as well as hard attention. The model takes advantage of batch based attention, which contributes to rapid convergence. The local attention model takes advantage of prediction based attention. Both local as well as global attention can be efficiently computed. However, it requires domain knowledge as well as diverse data instances for tuning the architecture, not a trivial task as well as inefficient.

While deep neural architecture can achieve better scalability that helps achieve domain generalization, these models cannot capture data subjectively, uncertainty, as well as human like reasoning [23]. As shown in [8], fuzzy modeling can help capture trust as well as human like reasoning. The purpose of this research is to show how fuzzy inference can be used in conjunction with Deep Learning to index network properties as a measure of trust. The task of indexing properties is subjective as well as introduces an element of ambiguity. It requires adaptive learning that cannot be achieved by sequentially stacking different modules.

3 METHODOLOGY

This article proposes an embedding strategy to detect depressive symptoms quickly as well as effectively. As shown in Fig. 1, symptom scores are calculated using cosine similarity. Lexicons are created using extended knowledge as well as word embedding. An example of the authored text from a patient (an anonymous user) is mentioned below.

Hello, long time depression sufferer here. I'm just having a hard day today as well as figuring out why I am even fighting this battle. No one cares about me etc. Why keep fighting this fight? I think I should just give up as well as let the Depression win so I can pass on as well as be at peace in heaven.

In mental health, classification according to the ICD10 [37] is complicated. The dynamic nature of emotional symptoms has different degrees as well as natures concerning patients. As a result, treatment on specific disease processes requires different methodologies to handle complex problems, i.e., Depression as well as anxiety. The assessment process includes mental health issues that were obtained by listening to the patient's queries. Psychiatrists as well as other medical professionals extract valuable elements to assess the clinical features of patients. For diagnosis, the psychiatrist uses a conventional technique that involves questionnaire based analysis, such as the PHQ 9[2]. The survey helps determine diagnostic accuracy as well as reliability in relation to the clinical circumstances of the person experiencing symptoms. The questionnaire includes a section on the types of symptoms as well as their frequency of occurrence. The total score is then used to diagnose mental disorders according to a predetermined cut off point. The results of the questionnaire help classify the behavior as mild, moderate, or severe. This technique is called Clinical Symptom Elicitation Process (CSEP) [37].

The aim of this research is to improve the explainability of the process by using fuzzy categorization based on a deep attention network. Active learning is used to improve adaptability. Each symptom group is labeled based on the frequency of patient messages.

3.1 Psychometric questionnaires (PQ)

The proposed technique uses the standard PHQ 9 questionnaire to collect patient written texts [22]. The PHQ 9 is a widely used method for assessing symptoms of depression. The PHQ 9 approach helps identify nine unique behavioral patterns included in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM V)[3]. These nine symptoms are then classified as sleep, interest, concentration, as well as eating problems, as shown in Table 1 as well as a sample document[4]. The psychiatrist generates the assessment score following the questionnaire assessment. The rating score reflects the patient's level of depression. The psychiatrist asks the questions on each category as well as checks the patient's response to assign the frequency of the class as follows:

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

3.2 Seed term generation

In this work, seed term can be extracted directly from PHQ 9 questionnaires. Any lexicon can be built making use of symptoms of depression given as common resources for psychiatrists, as given in Table 1. In this work, we made used of Wordnet [29] for the extraction of first order antonyms, hypernyms, as well as hyponyms. From Wordnet, each and every word from a given category may possess different synsets that can be used fro the expression of unique concepts. Synsets can be categorized both into semantic as well as lexiconc based relations. In may other works the use of various lists for symptoms of depression in a variety of systems for classification is presented [31]. The major systems for

[2]https://www.mdcalc.com/phq-9-patient-health-questionnaire-9
[3]https://www.psychiatry.org/psychiatrists/practice/dsm
[4]https://www.uspreventiveservicestaskforce.org/Home/GetFileByID/218

classification given for Depression for example like DSM V[5] as well as the International Statistical Classification of Diseases as well as Related Health Problems 10 (ICD 10) [37] are well known as well as used widely to asses depression scales that have since been merged [31].

The seed term in this study was derived from the PHQ 9 questionnaire. The vocabulary was created using symptoms of depression from widely available resources for psychiatrists, as shown in Table 1. We extracted the first order of hypernyms, hyponyms, as well as antonyms in this study using Wordnet [29]. Each category in Wordnet has its own collection of synsets that are used to convey specific terms. The synsets are classified according to their semantic as well as lexical relationships. Other research has used a variety of different lists of symptoms of depression classified according to different systems for classification [31]. Major systems for classification for depression such as the DSM V[6] as well as the International Classification of Diseases and Related Health Problems 10 (ICD 10) [37] are widely used depression scales that have been merged to create a fine basic list of symptoms [31].

3.3 Preprocessing

Preprocessing is a crucial step in the process of text processing. Any text written by a patient is subjected to a certain method, as follows:

(1) Each text is parsed and formatted according to the UTF 8 encoding convention. This helps to ensure consistency.
(2) Convert each word to its lowercase equivalent.
(3) Remove tabs or spaces between words
(4) Eliminate unique characters that have no meaning. (#, +, -, *, =, HTTP, HTTPS).
(5) Convert text based terms to complete words, e.g., *can't* by cannot and so on.

3.4 Word embedding using emotional lexicon

Knowledge based emotion embedding (EKB) makes use of word sense lexicon as well as a variety of learned contextual embeddings. Here, we determined word embedding for each and every word in patient written texts using 300 dimensions using a model for pre training for the global vector for word representation (GLOVE) [39]. We then trained our model for hyper tuning as well as domain adaptation using the embedding. Since the glove embedding was trained on data from the general structural language, it needs to be extended. To incorporate word sense, the transfer learning approach is used. The reason for this is that most of the embedding is done using open source data, in particular *(Wikipedia texts)* as well as sentiment knowledge, i.e. *(Twitter data)*, is trained. The terms *sad* as well as *happy* both refer to *feelings*. However, these expressions convey a different mental state. Therefore, it is necessary to extend the embedding using word sense. In this study, we extracted target terms from a part of speech, namely (*noun*, *verb*, *adverb*, as well as *adjective*). We used the corpus T, which consists of the text collection $T = \{d_1, d_2, \ldots, d_n\}$, as well as the *WordNet* used to extract synonyms, antonyms, hypernyms, as well as physical meaning for each as well as every extracted part of speech. This increased the number of emotional words in the provided word list $W = \{w_1, w_2, \ldots, w_K\}$. In this way, we develop the emotional lexicon. Then, the vocabulary is created from the set W used to train the model. The resulting embedding is the learned vector X, i.e. $X = \{x_1, x_2, \ldots, x_m\} \in \mathbb{R}^{m\times\delta}$, where δ is the vector dimension. It is possible to calculate sentence embedding by averaging the word vectors in the patient author's text. The trained model is used to convert the patient's information into a vector, as well as nine symptoms, using the text author's information. The

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

Table 1. PHQ 9 questionnaire and seed terms of each and every symptoms.

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

cosine similarity approach is used for the two matching embeddings. We have a similarity value of $0 - 1$ for each and every one of the nine symptoms.

The vector X represents the text written by the patient, while the vector Y represents the symptom lexicon. Additionally, V is used to construct semantically aware vectors from textual features. The similarity between two embeddings indicates that the published text is strongly associated with certain symptoms, as shown in Fig. 1.

3.5 Dataset

We used a dataset from various online forums as well as online sites to classify the nine specific symptoms in this study [31]. In our approach, an entropy based strategy extends the learned knowledge as well as ensures that rare events do not affect the performance of the proposed system. The annotation is based on the PHQ 9 rating scale, where 0 indicates no depression, 1 indicates mild depression, 2 indicates moderate depression, as well as 3 indicates severe depression, respectively [31]. We convert the annotation to a binary class for each and every symptom, where 0 represents the absence of symptoms as well as 1 represents their presence. The data collection process is described in detail in Table 2.

Table 2. The statistical summary of the training and testing set.

Type	Statistics
Corpus size (# of posts collected)	15044
# of sentences	133,524
Avg. sentences per post	8.87
Avg. words per post	232
set for training size (# of posts)	14,944
Testing set size (# of posts)	100

3.6 Deep learning model

In the study, we used a feed-forward neural network as a baseline for comparison. The tokens are extracted from the text by using the glove embedding approach. To approximate the length of the comments, the average method is used. The model is composed of hidden layers $(30, 20, 10)$ that are activated using the ReLU function for activation [3]. Our objective is to be able to classify nine unique symptoms using several labels. The last layer has the nine unit sigmoid function. The function for loss is the cross entropy function.

RNN with GRU was employed in the tests, as well as the findings indicated that LSTM cells as the RNN architecture worked well for the sequential task. A last time step of the hidden state can be delivered into the output layer using the LSTM unidirectional architecture. Additionally, we employed a bidirectional LSTM architecture that reads input token lists sequentially as well as set a single parameter for forward unrolled LSTM. As a result, each and every token's location has two input states that combine to generate the output state, increasing the attention layer as specified in the following equation as well as the designed model proposed by Huang *et al.* [16]. The dropout ratio is adjusted at $0:5$ to prevent overfitting as well as to regulate the LSTM layer.

$$\boldsymbol{i}_t = \sigma\left(\boldsymbol{x}_t\boldsymbol{W}^{(i)} + \boldsymbol{h}_{t-1}\boldsymbol{U}^{(i)}\right) \tag{1}$$

$$\boldsymbol{f}_t = \sigma\left(\boldsymbol{x}_t\boldsymbol{W}^{(f)} + \boldsymbol{h}_{t-1}\boldsymbol{U}^{(f)}\right) \tag{2}$$

$$\boldsymbol{o}_t = \sigma\left(\boldsymbol{x}_t\boldsymbol{W}^{(o)} + \boldsymbol{h}_{t-1}\boldsymbol{U}^{(o)}\right) \tag{3}$$

$$\tilde{\boldsymbol{c}}_t = \tanh\left(\boldsymbol{x}_t\boldsymbol{W}^{(o)} + \boldsymbol{h}_{t-1}\boldsymbol{U}^{(o)}\right) \tag{4}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \circ \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \circ \tilde{\boldsymbol{c}}_t \tag{5}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \circ \tanh\left(\boldsymbol{c}_t\right) \tag{6}$$

In addition, the developed framework had a mechanism for using the term "important" in the text [16]. We replaced the LSTM layer with the attention approach. This structure helps in extracting descriptive terms for classification problems. The dropout layer is provided with the attention output vector as input. The formal representation of the network is explained below. Training large networks using supervised learning traditionally requires a large dataset of labeled data. The greatest need as well as dependency of the application is on the labeled data. The active learning model is the process of generating the most informative data set for training a supervised model. Active learning is used in situations where the amount of data is too large to label manually. In this study, we used similarity based features to intelligently label a small amount of data, as well as then trained the entire dataset using the entropy based approach to instance selection. To expand the small number of instances as well as select the data distribution, the entropy process of instance selection is used [14]. This process helps to expand the knowledge gradually over time.

Nevertheless, the framework as designed is adapted from attention methods that can be used in the utilization of the importance of words in text[16]. An attention method was added instead of using a generic LSTM layer. This change/add on may assit in the extraction of words that are informative in classification tasks. We feed the output vector from the attention method in as input into the dropout layer. We give a more formal definition as well as description next. In a more traditional was supervised learning needs a large labelled dataset for training in any large networks. The primary dependency of the application is the labelled data. The relevant data then is generated as well as is the primary responsibility of the active learning model. We only keep the data that has the highest predictive nature for training the model. When the data from the applications is too big to manually label, we can use an active learning model. In this work, a similarity based feature process is used to label data in a smart fashion. Afterwards, entropy based selection is used for training on the entire dataset. We adopt an entropy based selection methodology for the expansion of the lower number of instances as well as to choose the data distribution. The entire process assists in expanding knowledge as time progresses.

$$\boldsymbol{v}_t = \tanh\left(\boldsymbol{h}_t\boldsymbol{W}_a + \boldsymbol{b}_a\right) \tag{7}$$

$$s_t = \boldsymbol{v}_t\boldsymbol{u}_a^\top \tag{8}$$

$$\alpha_t = \frac{\exp\left(s_t\right)}{\sum_{t=1}^{T}\exp\left(s_t\right)} \tag{9}$$

$$\tilde{\boldsymbol{h}} = \sum_{t=1}^{T}\alpha_t\boldsymbol{h}_t \tag{10}$$

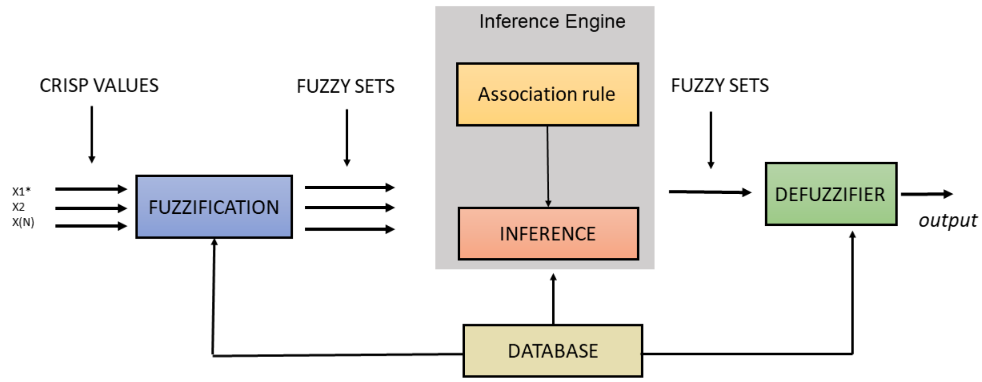

Fig. 2. A workflow of the designed fuzzy model (mamdani framework).

3.7 Fuzzy modeling

The Classification Based on Associations (CBA) approach is used in this research to develop rules for the classification problem [11]. First, the objectives of the fuzzy controller in Fig. 2, which is the output of the produced text of the was trained based on user interaction with the online system. Based on the CBA, a fuzzy set is selected. The Mamdani framework is used for the text classification part based on the Deep attention method output [41]. The rule based classification helps to address the reasoning behind the prediction of the attention network as well as word explainability. The physical intuition is used for the fuzzy objective to construct the fuzzy rules based on fuzzy inputs. This part required in depth domain knowledge.

3.7.1 Fuzzifier model. In fuzzification, the input is frequency of the attention words used for the prediction. This word frequency is used as input. A membership function determines the crisp input as well as generates the fuzzy output value in terms of degree. each and every word frequency is mapped to order paired document or degree od goodness [41].

$$\text{Triangle } (x; a, b, c) = \begin{cases} 0, & x \leq a \\ \frac{X-a,}{b-a} & a \leq x \leq b. \\ \frac{c-x,}{c-b} & b \leq x \leq c. \\ 0, & c \leq x. \end{cases} \tag{11}$$

We used a triangular fuzzifier that takes word frequency as input as well as converts it into a triangular fuzzy one. In this paper, we used a triangular membership function. Using *min* as well as *max* operations, we have an alternative expression for Eqs. 11 as well as 12.

$$\text{Triangle } (\mathrm{x}; \mathrm{a}, \mathrm{b}, \mathrm{c}) = \max\left(\min\left(\frac{\mathrm{x}-\mathrm{a}, \mathrm{c}-\mathrm{x}}{\mathrm{b}-\mathrm{a}}\mathrm{c}-\mathrm{b}\right), 0\right) \tag{12}$$

3.7.2 Fuzzifier parameters. The inference engine helps to convert the fuzzy values into rules. Then the output is passed to the defuzzifier, which converts the fuzzy input to a fuzzy output. To develop rules, we used the CBA (Classification Based on Associations) technique [11] to classify the fuzzy rules. It first finds the frequent itemsets, which helps to construct support as well as confidence. Then, rules are sort as well as organize to form a classifier.

3.7.3 Defuzzifier model. In this paper, we used the centre average de fuzzifier technique [41]. The centre average de fuzzification formula is mentioned in equation 12, where y^{-i} denotes the centre value of region μ_0^i small, medium, large etc. as well as M is the total number of rules in the rule base. Based on the word frequency association, the output is mapped to nine symptoms mentioned in Table 1.

3.7.4 Fuzzy training. Given the word frequency based on the attention method, the fuzzification is done, fuzzy rule classification is made based on the CBA, as well as De fuzzifier helps to convert fuzzy set as well as fuzzy rules. The prediction is based on the computed membership value as well as compared with previous computer membership values, with the threshold value of $+0.001$ to -0.001. Then, the centroid average method helps to compute output results. While the rule is being changed, the result is entered into the database.

3.7.5 Fuzzy prediction. It may be explained using the following steps:

(1) Define fuzzy set based on triangle membership function of word frequency extracted from the attention method.
(2) Generate fuzzy rules by using association rule mining.
(3) Compute as well as assign a degree of membership to each as well as every rule.
(4) Compute the membership value for the set for training as well as make a prediction based on the threshold value.

3.8 Performance metrics

In the last few months, a collection of many studies have been able to show that there is a great deal of interest in the use of ROC curves to measure the performance of ML based models. In fact, some speculate it can be considered one of the main methods for measuring performance for ML methodologies [1, 2]. Here, we make use of false negative rate (FPR), true positive rate (TPR). We also make use of the standard metrics such as accuracy, recall, precision, as well as F-measure. TPR is given as TPR = $\frac{TP}{TP+FN}$. Furthermore, FPR can be given as FPR = $\frac{FP}{FP+TN}$.

4 EXPERIMENTAL ANALYSIS AND DISCUSSION OF RESULTS

The material written by the patient himself serves as the basis for the emotional lexicon. Text representation is then trained on a separate architecture. We used the Glove based vectorizer for transfer learning. In addition, the embedding is trained using average vectorization as well as tagged vectors. To further categorize the data instances, a similarity based technique is used between the questionnaire as well as the text written by the patient. For each and every architecture, we employed the Adam optimizer as well as the learning rate was limited to 0.0005 [20]. We varied the cell type as well as hidden size for each and every architecture. Along with the directed LSTM layer, an attention approach was used to increase the performance of the model. The concept of early stopping was used to store the model incrementally. To avoid and eliminate difficulties with the gradient, a gradient pruning approach was used [5]. As seen in Table 3, the attention method with fuzzy classification rules achieves a 0.89 F measure. We evaluated the architecture designed by considering the attention focus, aggregation of the focus as well as attention score as well as using the activation output with as well as without fuzzy classification model. Whenever a focus item exists in the attention based model, then it tends to perform better. However, in the absence of the focus word, attention networks miss classified the label. Whereas the fuzzy model, successfully able to classify in both cases.

Using different k fold values, we tested the embedding architecture, i.e., two, three, five, as well as ten folds. To check the detection performance under different epochs, we set the training/test set percentages as 0.50, 0.66, 0.80 and 0.90. The number of epochs is set to 50. Overall, all models can achieve a high F1 measure. However, the fuzzy classifier with

Table 3. The F-measure of testing set.

Architectures	Test set
Baseline	0.85
Bidirectional LSTM	0.87
Bidirectional_LSTM_Attention	**0.88**
Bidirectional_LSTM_Attention with Fuzzy classification	**0.89**

rule based configuration performed better despite the heavy unbalanced dataset, where symptoms two are dominant in the dataset. The detection results of all LSTM architecture looks similar to the small percentage of increase.

The model in Table 4 achieves a high recall of 0.87, but has the lowest accuracy of all models. The design is inefficient for the data at hand. The depression data depends on the order of the words, which the base network did not maintain. In other words, an architecture that prefers sequences as well as stores the crucial word information is needed.

Table 4. K-fold approaches with variations on baseline model are used to calculate Precision, Recall and F-measure

	Precision	Recall	F-measure
2-Fold	0.78	0.81	0.79
3-Fold	0.84	0.81	0.82
5-Fold	0.84	0.85	0.84
10-Fold	0.84	0.87	**0.85**

4.1 Deep Attention model Analysis

In Tables 5 and 6, bidirectional LSTM is still able to achieve the similar performance of 0.87 as well as 0.88, respectively. The highest values are marked bold that represent the effective architecture concerning the number of folds. However, bidirectional LSTM with the attention model achieved 0.89 precision. The reason is that the model works in two directions: forward from the past to the future as well as backward from the future to the past. The two hidden state models maintain both future as well as past information. The two separate RNNs run in parallel, allowing backward as well as forward connections between the networks. The error is lowest in the training as well as evolution sets. The precision recall curve in the upper right corner shows high recall as well as accuracy, illustrating the low incidence of false positives as well as negatives. The BILSTM model accounts for any hidden state that was dependent on the previous state. This poses a major problem as the network is forced to wait for data. Long term dependencies affect performance because it is difficult to retain knowledge over time.

Table 5. K-fold approaches with variations on Bidirectional LSTM model are used to calculate Precision, Recall and F-measure

	Precision	Recall	F-measure
2-Fold	0.78	0.77	0.77
3-Fold	0.83	0.81	0.82
5-Fold	0.83	0.87	0.85
10-Fold	0.88	0.87	**0.87**

4.2 Fuzzy rule classification analysis

The fuzzy classifier in Table 7 achieved high recognition as well as accuracy of 0.89. The fuzzy model also generated rules with extensible knowledge using the attention based word list. With an increase in the cross validation score, the fuzzy classifier performed better; this improved performance resulted in a high true positive rate (TP). The results demonstrate the presence of important terms that contribute significantly to as well as support the classification of depressive symptoms. In addition, the network contributes to cost reduction by focusing on specific phrases. This is because the model is able to recognize the target word of the task as well as has learned the meaning of the topic in both directions. Thus, we can display the optimal model as well as evaluate its predictive power in randomized experiments.

Table 6. *K*-fold approaches with variations on Bidirectional LSTM using a fuzzy classification attention model are used to calculate Precision, Recall and F-measure

	Precision	Recall	F-measure
2-Fold	0.81	0.81	0.81
3-Fold	0.86	0.84	0.85
5-Fold	0.89	0.87	**0.88**
10-Fold	0.81	0.87	0.84

Table 7. *K*-fold approaches with variations on Bidirectional LSTM fuzzy classification model are used to calculate Precision, Recall and F-measure

	Precision	Recall	F-measure
2-Fold	0.81	0.83	0.82
3-Fold	0.88	0.87	0.87
5-Fold	0.88	0.9	**0.89**
10-Fold	0.89	0.9	**0.89**

4.3 Visualization analysis and usage

Fig. The visualization illustrates the attention based probability as well as triggered word rules of the sentence. The visualization allows the psychiatrist to see the categorization reasons for the diagnosis choice as the words , *peace*, *heaven*, as well as *depression*. Weights as highlighted give 2 symptoms, namely $S3$ as well as $S2$ (hopeless/feeling down). The fuzzy approach can successfully detect as well as highlight focused words which may help psychiatrists in the recording as well as the diagnosis of patients. The fuzzy model may be able to detects successfully triggering rules as well as highlights on focused words that may help psychiatrists make strong diagnoses. The visualization helps to get the context behind the symptoms extraction, attention model with support of fuzzy rule make the system more explainable as well as less complex. As seen from Fig. 3, dark highlighted words are trigging the decision. The proposed model helps to understand the reasoning behind the prediction. The prediction model's intention helps to support its decision making, as well as if psychiatrist see any irregularity, it can be seen from the word attention as well as fuzzy rules.

4.4 Discussion

Prediction of future values is entirely satisfactory, even when data is highly non linear. The accuracy of prediction can be arbitrarily fixed to an acceptable limit by increasing membership functions over the data domain. With each as well

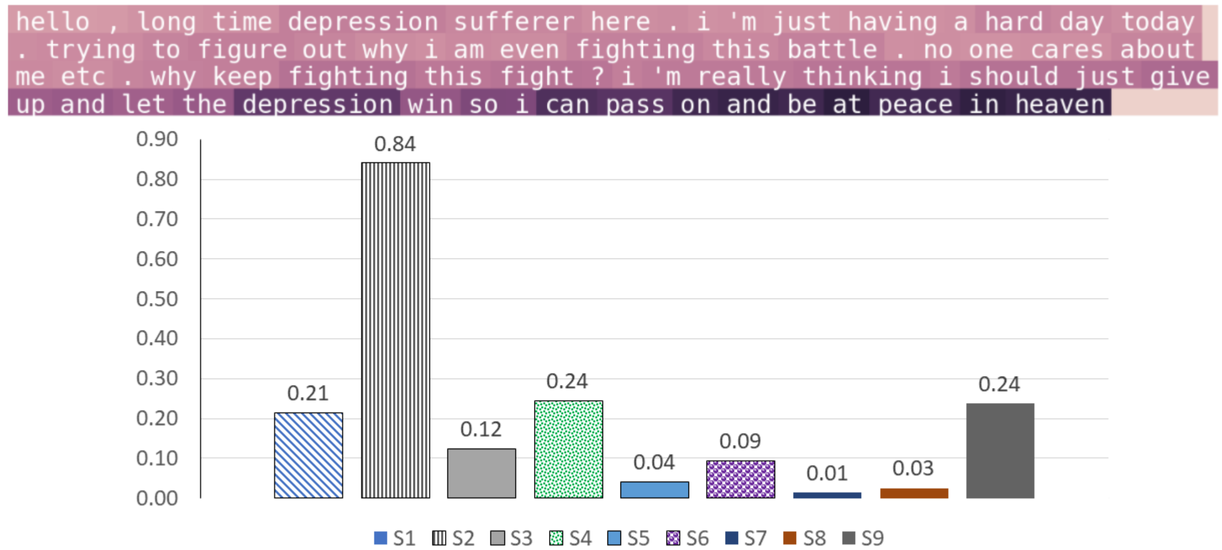

Fig. 3. An example patient authored text, symptom probability, fuzzy rule visualization of symptoms of depression extracted by our approach.

as every prediction, a new rule is added to the rule base. Thus, the 'quality of prediction keeps on increasing. Because of their similarity to human control logic, fuzzy controllers have proven to be an excellent choice for a variety of control systems. They can be incorporated into anything from a small handmade item to a large automated processing control system. Due to their superior performance, high reliability as well as durability, fuzzy based controllers have become very competitive in today's market. The current research does not need a sophisticated thought process. It was simply beautiful as well as easy to apply. The proposed application demonstrates a straightforward method to arrive at a final result based on confusing, inaccurate, noisy or missing input data. The fuzzy logic controller method for problem solving simulates how a human might make a decision more quickly. Because the system does not require perfect, noise free inputs, it is resilient. Despite a wide range of input variations, the output control is a smooth control function. The controller consists of user defined rules that control the target system, as well as it can be easily updated to significantly improve the system's performance. By developing appropriate control rules, new symptoms can be easily incorporated into the system. The fuzzy controller is not bound to a small number of inputs or outputs. It allows low cost as well as imprecise extraction of symptoms, which makes the whole system cost effective as well as simple.

5 CONCLUSION AND FUTURE DIRECTIONS

This research explores concepts of embedding by expanding emotional lexicons. The fuzzy model then uses the embedding to draw conclusions as well as classify the symptoms. The proposed approach creates the emotional embedding for the fuzzy inference system using bidirectional LSTM gates. Our research shows that fuzzy logic components are able to handle ambiguity as well as correctly interpret the operation of the embedding. For similarity based labeled data, the fuzzy based model achieved an F value of 0.89. Furthermore, the model able to expand its knowledge by using entropy based selections. Fuzzy logic rules are gathered from classification based on the association rules. The method as proposed given an adaption in the IDPT system by autonomously expanding its knowledge of the symptoms. The intervention as adapted provides feedback personal in nature on the recommended exercises. For the future, to increase the detection rate, we will use the dynamic architecture search as well as automatic rule generation methods. We suggest that the fuzzy recurrent attention model performed as well as interpretation analysis efficiently. The enhanced version should incorporate the visuals feature of the patients to help in the predictions tasks.

REFERENCES

[1] Usman Ahmed, Muhammad Aleem, Yasir Noman Khalid, Muhammad Arshad Islam, and Muhammad Azhar Iqbal. 2019. RALB-HC: A resource-aware load balancer for heterogeneous cluster. *Concurrency and Computation: Practice and Experience* (2019), e5606.

[2] Usman Ahmed, Jerry Chun Wei Lin, Gautam Srivastava, and Muhammad Aleem. 2020. A load balance multi-scheduling model for OpenCL kernel tasks in an integrated cluster. *Soft Computing* (2020), 1–14.

[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.). 1–10.

[4] 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.

[5] 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).

[6] 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.). 1724–1734.

[7] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. 2016. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine learning for healthcare conference*. 301–318.

[8] Palash Dutta and Soumendra Goala. 2018. Fuzzy decision making in medical diagnosis using an advanced distance measure on intuitionistic fuzzy sets. *The Open Cybernetics & Systemics Journal* 12, 1 (2018).

[9] Hassan Ismail Fawaz. 2020. Deep learning for time series classification. *CoRR* abs/2010.00567 (2020).

[10] Robin C Feldman, Ehrik Aldana, and Kara Stein. 2019. Artificial intelligence in the health care space: how we can trust what we cannot know. *Stan. L. & Pol'y Rev.* 30 (2019), 399.

[11] Jiří Filip and Tomáš Kliegr. 2018. *Classification based on Associations (CBA)-a performance analysis*. Technical Report. EasyChair.

[12] 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.

[13] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* 2, 2 (2017).

[14] Alex Holub, Pietro Perona, and Michael C Burl. 2008. Entropy-based active learning for object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1–8.

[15] Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. 2019. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9, 4 (2019), e1312.

[16] Yingren Huang, Jiaojiao Chen, Shaomin Zheng, Yun Xue, and Xiaohui Hu. 2021. Hierarchical multi-attention networks for document classification. *Int. J. Mach. Learn. Cybern.* 12, 6 (2021), 1639–1647.

[17] 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.

[18] Kipp W Johnson, Jessica Torres Soto, Benjamin S Glicksberg, Khader Shameer, Riccardo Miotto, Mohsin Ali, Euan Ashley, and Joel T Dudley. 2018. Artificial intelligence in cardiology. *Journal of the American College of Cardiology* 71, 23 (2018), 2668–2679.

[19] 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.

[20] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[21] Chayakrit Krittanawong, HongJu Zhang, Zhen Wang, Mehmet Aydar, and Takeshi Kitai. 2017. Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology* 69, 21 (2017), 2657–2664.

[22] 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.

[23] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature* 521, 7553 (2015), 436–444.

[24] 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.

[25] 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.

[26] 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.). 289–297.

[27] 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.). 1412–1421.

[28] 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.

[29] 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.

[30] 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.

[31] 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, 226–232.

[32] 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. https://doi.org/10.1109/ACCESS.2020.3002793

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

[38] Krunal Dhiraj Patel, Kazi Zainab, Andrew Heppner, Gautam Srivastava, and Vijay Mago. 2020. Using Twitter for diabetes community analysis. *Network Modeling Analysis in Health Informatics and Bioinformatics* 9 (2020), 1–16.

[39] 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*. 1532–1543.

[40] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. 2018. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine* 1, 1 (2018), 1–10.

[41] Alejandro Santiago, Bernabé Dorronsoro, Antonio J Nebro, Juan J Durillo, Oscar Castillo, and Héctor J Fraire. 2019. A novel multi-objective evolutionary algorithm with fuzzy logic based adaptive selection of operators: FAME. *Information Sciences* 471 (2019), 233–251.

[42] 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.

[43] 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.

[44] Alfredo Vellido. 2019. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications* (2019), 1–15.

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

[46] Fei Wang, Rainu Kaushal, and Dhruv Khullar. 2020. Should health care demand interpretable artificial intelligence or accept "black box" medicine?

[47] 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.). 2048–2057.