This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

Hyper-Graph Attention Based Federated Learning Methods For Use in Mental Health Detection

Usman Ahmed, Jerry Chun-Wei Lin*, and Gautam Srivastava

***Abstract*—Internet-Delivered Psychological Treatment (IDPT) has become necessary in the medical field. Deep neural networks (DNNs) require large, diverse patient populations to train models that achieve clinician-level performance. However, DNN models trained on limited datasets have poor clinical performance when used in a new location with different data. Thus, increasing the availability of diverse as well as distinct training data is vital. This study proposes a structural hypergraph as well as an emotional lexicon for word representation. An embedding model based on federated learning was developed for mental health symptom detection. The model treats text data as a collection of consecutive words. The model then learns a low-dimensional continuous vector while maintaining contextual linkage. The generated models with attention-based mechanisms as well as federated learning are then tested experimentally. Our strategy is suitable for vocabulary diversification, grammatical word representation, as well as dynamic lexicon analysis. The goal is to create semantic word representations using an attention network model. Later, clinical processes are used to mark the text by embedding it. Experimental results show the encoding of emotional words using the structural hypergraph. The 0.86 ROC was achieved using the bidirectional LSTM architecture with an attention mechanism.**

***Index Terms*—text clustering, NLP, Internet-delivered interventions, word sense identification, adaptive treatments**

I. Introduction

Wearable gadgets as well as innovations on the Internet of Medical Things (IoMT) have made remote patient monitoring possible like never before. Through the process of learning patterns from provided data by devices, machine learning, as well as deep learning algorithms, are significantly helping physicians diagnose patients remotely. Classical machine learning (ML), as well as deep learning (DL) models have the disadvantage of requiring patient data to be transferred from specialized devices, sensors, as well as wearables to centralized servers so that the data can be trained with ML/DL models. Because of the nature of data in healthcare, the techniques mentioned thus far for transferring patient data to centralized servers can pose significant privacy, as well as security issues.

Recent advancement in ML/DL is federated learning, where data is not transmitted to centralized servers. However, the ML model itself is distributed to different nodes (devices) for training data [1]. Parameters of the device models are then transmitted to a centralized model for training the model globally. Federated learning can protect patient data privacy by preventing sensitive information from being exposed to intruders like hackers. COVID-19 is a global health disaster which has threatened the livelihood of millions of people. Advanced ML technologies have been used to build models to predict and diagnose diseases for combating coronavirus. Furthermore, partly because of unstable communication methods, as well as potential attackers, the huge amount of data collected during this time can pose various security and privacy issues. Privacy-preserving federated learning becomes a superior alternative to ensure the security of patient data during the transfer, as well as the training process. Therefore, we have compiled several high quality works on this topic that use state-of-the-art federated learning technologies to protect healthcare data, as well as provide valuable guidance for the current society.

According to WHO [2], *depression* is a severe problem among the most disabling diseases in the world. Depressive disorders affect approximately 264 million people worldwide. Due to a lack of interpersonal interaction, as well as trust, most cases of depression go untreated [3]. Because they do not seek treatment, the leading cause of mortality among individuals between the ages of 15, as well as 29 years is suicide or 76% - 85% within middle-income countries. In addition to mental health issues, early detection is hindered by a lack of resources, inexperienced medical personnel, social stigma, as well as rapid response [2]. People are often humbled by their inability to maintain stability with one's mental state is part due to both shy aspects, as well as nervousness. Anytime that a patient goes through any kind of thorough evaluation of psychological aspects of their current state, the problem persists [4]; as a result, people with depression may decide against future therapy to manage their existing health problems.

When dealing directly with IDPT, or Internet-delivered Psychological Treatment, we can say without a doubt that this level of treatment helps people deal with their own psychological problems with less resources. A tunnel-based approach is rigid as well as not interoperable [5]. The approach is not adaptable due to low adoption, as well as a more significant number of dropouts. User adoption must be considered overtime. With an IDPT system that assesses user behavior, as well as emotional exchanges, adaptation is possible. For example, Mukhiya *et al.* notes that culturally-based mental health issues influence individualized user behavior [6].

U. Ahmed and J. C. W Lin are with the Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, 5063, Bergen, Norway. Email: usman.ahmed@hvl.no, jerrylin@ieee.org. Website: http://ikelab.net. (*Corresponding author: Jerry Chun-Wei Lin)

G. Srivastava is with the Department of Mathematics & Computer Science, Brandon University, Canada as well as the Research Centre for Interneural Computing, China Medical University, Taichung, Taiwan. Email: SRIVASTAVAG@brandonu.ca

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

A. Motivation

This pandemic has exacerbated mental health problems worldwide. According to WHO [1], 93% of countries have seen an increase in mental diseases. The lockdown has exacerbated people's physiological stress, including any fear of disease transmission and fear of the future of humanity [7]. Emotional tensions have grown due to lack of interaction, social isolation, educational insecurity, and irregular work schedules. Anxiety and depression among medical personnel are due to inadequate protective equipment, fear of illness, social isolation, and stressful environments. Overall, the incidence of depression in the pandemic is much higher than expected [8]. Individuals may communicate through online forums and social media platforms, in part because more interactive online systems are being used in many well-known sectors. Since the COVID -19 epidemic, individuals in our "global village" may have become accustomed to Internet communication. Symptoms associated with online screening for depression may help identify individuals at risk for mental illness. Taking medication in a timely manner can help improve overall health [9].

5G/6G and beyond networks are being built and will soon replace 3G/4G in many of the known advanced nations. 6G mobile networks are also emerging. 5G/6G bit speeds will support data-driven applications. IoT applications can connect to data centers with higher capacity and lower latency. This means databases will be dynamically updated. Mobile nodes with sensitive data can be attacked using known attack methods such as denial of service (DoS), replay attacks, eavesdropping, and repudiation. The scientific community has paid little attention to securing data from multiple sources. Many elements can be considered when creating a learning environment.

Researchers use measurement-based methods to overcome dimensionality as well as data scarcity [10]. Initial data representation evaluations include identification as well as classification. Individuals may not be able to apply patterns and learning mechanisms to novel adaptive tasks. Second, the techniques require the structured instances of the database [11]. However, real-world applications are often dynamic and some applications, such as time series, do not allow database reexamination due to sequencing issues. Finally, the problem of extracting patterns from data without sharing them is not adequately addressed. Centralized databases often work together. This structure usually leads to unnecessary overhead, requiring time-consuming approvals due to privacy and ethical concerns. The enterprise values the datasets even if and only if the constraints are resolved; therefore, they prefer not to share them. Mobile or IoT network datasets can be huge, and storing them centrally can be expensive. Therefore, federated learning can solve the above problems by sharing only the model weights across the network, but not the raw data input [1].

B. Contributions

This study presents a method for extracting depression symptoms from texts using NLP, also known as Natural Language Processing, and attention-based learning. An emotion-driven context extraction method and a structural hypergraph propose semantic vectors. Unlabeled text is stripped of essential boundary components before being used in the learning process. The method extends the model training. Repeat the process until the optimal option is found. The pool of unlabeled text is now part of the training set. The study aims to improve the understanding of the learning process by expanding the text material over time. The proposed technique reduces the cost of data annotation while increasing the generalization of the learning system. The semantic vectors of the graph network as well as the synonym expansion contribute to high accuracy without compromising the data annotation.

In addition to attention-based learning, federated learning is used in this study. The proposed model generalizes the low-dimensional embedding for depression data. The attention-based deep learning algorithm reduces the structural hypergraph to a low-dimensional space. The learned embedding is then used to extract and classify symptoms. The federated learning model, which is both intercepted and shared locally, helps improve both global and local performance. By preserving the relationship between entities, embedding helps in constructing low-dimensional continuous vectors. The attention-based method helps in capturing semantic vector associations. We briefly summarize the contributions as follows:

1) To learn a low-dimensional continuous representation, we present an effective attention-based embedding model that is used for handling mental health data.
2) We have shown that federated learning with learned embedding may be able to improve performance even when raw data is not shared.
3) Empirical evaluation of the proposed model is the discussed and compared with the state-of-the-art methods to show that the designed model outperforms the past studies.

II. Related Work

Federated learning can be proposed to avoid data breaches, as well as the building of machine learning models, mostly based on the use of remote datasets [12]. The framework communicates only the model weights with the network nodes and the model trains locally without exchanging real datasets [13], [14]. Previous work has mainly focused on data distribution, imbalanced data, and device optimization capabilities. There are two types of learning models in federated learning: horizontal and vertical [15].

Current research focuses on NLP-based methods for computerized systems. Item Response Theory (IRT)-based Computer Adaptive Tests (CATs) for detecting depression symptoms are investigated by Li *et al.* [16]. Their study achieved a high level of accuracy by using an adaptive questionnaire instead of a static questionnaire. Although the explanatory power as well as interpretation of the model are evident, there is no explanation of why they choose this particular questionnaire or how, for example, changing the response options within a single test affects response behavior. Lerman *et al.* [17] examined Twitter data to see if and only if there were mental

[1] shorturl.at/enwTV

health issues. LIWC, also known as Linguistic Inquiry Word Counts, contrast the experimental as well as control groups, according to McDonnell *et al.* [18]. To evaluate 5-character sequences, two linguistic models are used to examine word probability: (1) a unigram model as well as (2) a character-based 5-gram model. While the other classifiers may be able to discriminate between groups, the trained classifier may be able to discriminate between every group as well as the control group while indicating a positive signal in the language of every group. The associations between the statistical analyzes as well as the classifiers were then examined to see if and only if there were any associations between quantifiable as well as relevant mental health signals on Twitter [18].

Nguyen *et al.* [19] investigates in a neural network that the learned features represent the conditional probability distribution of the input vectors. A variety of designs have been proposed for application-specific domains. A network consists of multiple layers, each and every representing the average of the previous layer's weights. The activation function is used in the last layer, the output layer. Hidden layers, as well as structural approximations, give deep neural networks their predictive power. Optimal architecture selection increases accuracy in tackling complex problems, as well as layer structure as well as hyperparameters are modified on a large scale. The generalizability of learning may be able to be improved by meaningful feature representation. This study aims to find out how to make the user experience more compassionate. In a cinematic approach, users' expressions may be able to then be used to initiate game as well as entertainment events, leading to dynamic application behaviour. To classify emotions, the approach uses the reality factor.

Although deep neural architecture improves scalability and facilitates domain generalization, these models may account for subjective inputs, uncertainty, or human-like reasoning. As shown in [20], fuzzy modeling can help in capturing confidence and human-like reasoning. In this work, fuzzy inference and deep learning are combined to index learned network features as certainty. The task of indexing features is subjective and involves some uncertainty [21]. It requires adaptive learning, which cannot be achieved by stacking multiple modules in sequential order.

Zhuotao *et al.* proposed a layer-based federated learning system with privacy preservation [22]. The model lowers the communication cost by uploading many layers of the model for global averaging and improves privacy protection by using local differential privacy. The method was tested on three datasets in a non-independent as well as identically distributed environment. Jingcheng *et al.* build an efficient privacy-preserving data aggregation federated learning Scheme (EPPDA), an efficient privacy-preserving data aggregation method for FL, based on secret sharing to resist the reverse attack, which may be able to covertly aggregate user-trained models without disclosing the user model [23]. In addition, EPPDA exhibits acceptable fault tolerance in the event of a user disconnection. Even if and only if a significant number of users are disconnected while running the protocol, EPPDA continues to function correctly. The analysis results show that EPPDA may be able to provide locally trained models to the server as a whole without revealing the model of any particular user. In addition, the adversary cannot obtain non-public information about the communication channel. The verification of the efficiency of EPPDA shows that it preserves user privacy and requires less computing and communication resources.

While AI-powered systems can help improve energy distribution between charging stations and charging station providers, the frequent exchange of data between them can lead to security and privacy issues [24]. Federated learning (FL) requires charging stations (CS) to contribute local models rather than full data to address these issues. Wang *et al.* provides a lightweight authenticated FL-based energy demand prediction for electric vehicle infrastructures (EVAs) with the premium penalty mechanism. The architecture might be able to accurately estimate power consumption to resist numerous FL threats to EVIs. Wang *et al.* present a new pairing-free certificate system based on the latest blockchain technology and smart contracts [25]. They then simulated type I and type II attackers to test the security of o the developed system. The solution provides a more reliable security guarantee with lower computational overhead (i.e., reduced by up to 40%) and lower communication overhead (i.e., reduced by 94.7%).

Internet-Delivered Psychological Treatments (IDPT) focus on Internet-based mental health problems. As the epidemic spreads, more Internet technologies are being used to provide evidence-based mental health services. This increase is helping to serve more people with less money. Faster solutions to mental illness are possible through adaptability and flexibility. To classify the material of mentally ill people into different symptoms, Ahmed *et al.* [21] present a fuzzy contrast based model using the attention network for position weighted words. Then, the learned embedding labels the mental data. Then, the attention network expands its lexicons to enable transfer learning. Both similarity and contrast sets are used in the prediction of weighted attention terms. The fuzzy model then uses the sets to categorize the mental health data. Both non-embedding and standard methods are contrasted to illustrate the proposed paradigm. The feature vector achieved a ROC curve of 0.82 with nine symptomatic problems.

Yuan *et al.* used multi-attention to help screen, diagnose, treat, as well as evaluate various cardiovascular and ophthalmic diseases [26]. UNet deep learning model to fully exploit detailed low-level information as well as complementary information stored in multiple layers to properly separate vessels from background. The dense dropout block is designed to preserve maximum vessel information between convolutional layers and avoid overfitting. It is included in the contracting process to automatically evaluate the relevance of each feature channel. Then, an extended way is used to extract multi-level features, which are then used to improve the features in each layer using the attention method.

The model described above uses deep learning to extract text embedding approaches. On the other hand, semi-supervised learning requires the use of labels as well as training the model with output classes. The goal of this project is to develop a tool for extracting depression symptoms from patient-submitted writings in a federated learning environment using natural

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

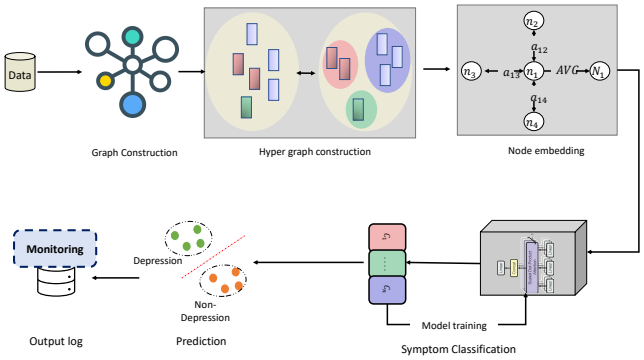


Fig. 1: A workflow architecture of hypergraph.

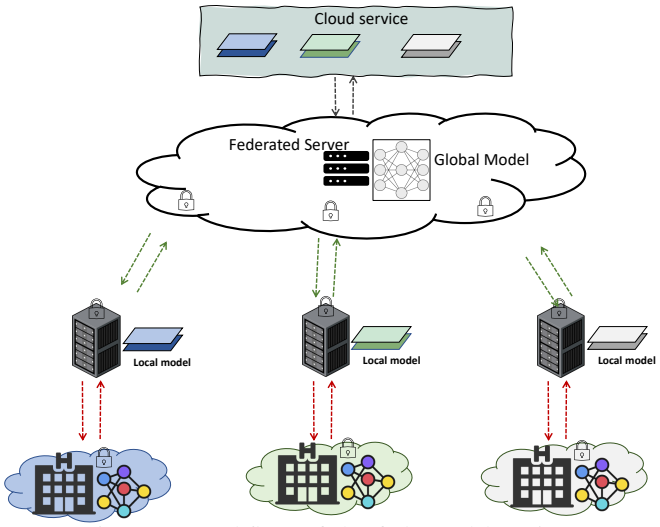


Fig. 2: A workflow of the federated learning.

language processing. The proposed approach uses the graph embedding method to extract, identify, as well as represent the graph attention approach. The patients' mental problems are mainly expressed through their interactions with the system. Recognizing these communication pattern sequences may be able to help clinicians identify the root causes of mental as well as emotional problems. The included interactive online tool (ICT) may be able to help provide contextual information, as well as visualizations for mental health prevention programs.

III. Developed Hyper-graph Attention-based Federated Learning Model

This study interprets a patient authored text as a sequential directed acyclic graph G in which word points denote location as well as time. The edge between every node in this study reflects a set of words that occur sequentially, where the structure is represented by a corpse C as well as the set of lexicons drawn from the set E of the PHQ-9 graph mentioned in Section III-B mentioned PHQ-9 questionnaire $\mathcal{F}(\mathcal{C}) = \{C_1, C_2, \ldots, C_E\}$, word nodes were discovered. A hypergraph represented the nodes in the graph. As shown in Fig. 1, we then consider the hypergraph to process the words according to their semantic meanings. For each edge, it has its own attention network, which is defiend as the edge plan here, and the attention network of the hyperedges are then considered as the node in the designed model. The hyperedgers are then transformed as the hypervertices that is used for sentence embedding by adapting the average measure in the designed model.

In this paper, we present a graph embedding method that may be able to quickly as well as accurately detect depressive symptoms. After embedding, we used cosine similarity to generate symptom scores, as shown in Fig. 1. A structure-aware graph model is used to create the latent space. Comprehensive lexicons are created by combining comprehensive information as well as graph embedding. This study aims to support as well as improve psychiatrists' note-taking using graph attention networks. The frequency of patient messages is used to label every symptom group.

A. Client server embedding method

The proposed framework takes advantage of multi-client data in Fig. 1 as well as Fig. 2. Federated learning is an approach to machine learning in which an algorithm is trained without transmitting data samples across multiple decentralized edge devices or servers that keep local data samples. We split the data across six clients for experimental purposes, with each client receiving an equal share. In addition, each client is provided with non-overlapping data, a local model, and a database. The data is sent to the clients in random order. Because of these distinctions, frameworks often display real-world statistics, such as the various branches of a store in a city connected to a server for supply and demand searches. In this context, the environment is also not independent or evenly distributed. The federated learning approach takes advantage of a server-client context. Client data is managed locally for each model. The local data is required to train the initialization model. We require the client to submit the locally learned weights or their gradient to the server after each iteration *(Algorithm 1-lines 3 to 9)*. The weights are received by the server and aggregated *(Algorithm 1-lines 3 to 9)*. The idea is that the client trains the model using the weights rather than the actual data.

The global model is then updated using the aggregated weights *(Algorithm 1-lines 8 to 9)*. After aggregation, the global weights are used to begin the next cycle of local client models. The client convergence point is reached after a certain number of iterations of federated learning. We used the early termination strategy to experimentally determine the convergence point. During the empirical study, we set the client's early termination value to 10. The client is then able to select the best model for each iteration based on the holdout data. The client tracks the validation loss on the local test set. The client can select either the global aggregated model or the best local iteration model. The federated averaging strategy was chosen because it converges quickly and reduces model overfitting [12]. However, in our empirical study, a larger embedding size is required for optimal performance. The reason for this is the ability of the decoder model to map positional attention in a broader vector space.

The purpose of federated averaging is to reduce the global

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

Algorithm 1 Attention-based federated averaging method.

INPUT: $Client$: number of client for the federeated learning aggregation, R: Number of Federated learning rounds per client, LE: client local training epochs to minimize loss.

OUTPUT: Optimize Gradient

$Weights \leftarrow \phi^{random}$ ▷ Initialize clients weights randomly.
for *all* $r \in R$ **do**
  for *all* $k \in client$ **do**
    $Send \leftarrow Gradient()$; ▷ Graph attention network;
    $Receive \leftarrow \Delta(Gradient)$;
  end for
  $\phi_k^{(r)} \leftarrow \phi^{(r-1)} + \Delta\phi_k^{(r)}$;
  $\phi^{(r)} \leftarrow \frac{1}{\sum_k n_k} \sum_k \left(n_k \cdot \phi_k^{(r)}\right)$; ▷ Aggregate();
end for
Return $Gradient$.

loss function L, which is the result of a weighted combination of K losses from the distributed aggregate function $\{\mathcal{L}_k\}_{k=1}^{K}$. The model could learn the embedding using the phi parameter that minimizes L on local data C_k, where X is a combination of local data sets as well as the embedding representation. Equation 1 represents the loss function.

$$\min_{\phi} \mathcal{L}(X;\phi) \quad \sum_{k=1}^{K} w_k \mathcal{L}_k\left(X_k;\phi\right) \tag{1}$$

B. Psychometric questionnaires (PHQ)

The proposed method collects texts authored by patients using the conventional PHQ-9 questionnaire [27]. The PHQ-9 is a popular method for assessing depressive symptoms, which is able to help in identifying nine unique patterns of behaviour listed in the Diagnostic as well as Statistical Manual of Mental Disorders 5 (DSM-V)[2]. The nine PHQ-9 symptoms are then grouped into disorders such as sleeping, interest, concentration, as well as eating problems[3].

Several researchers have used a method to assess depression. The remedy was then tested using the PHQ-9, which was used to determine the patient's depression score [20]. PHQ-2, which includes two items, as well as PHQ-15, which includes 15 somatic symptoms from the PHQ, are the other methods that can also be used for evaluation. The assessment, as well as diagnosis of mental health according to the ICD10 [28] criteria, is a complicated process. Several elements influence a patient's mental health, including family culture, previous therapies, society, childhood memories, work-life, as well as daily routines. Psychiatrists closely examine the elements that trigger the mental disorder during screening, as well as history collection.

The most common way psychiatrists do this is to list things that happen to their patients, as well as highlight the essential parts. Based on the points the patient has mentioned, psychiatrists suggest some exercises. The second time, psychiatrists use the same questionnaire they always use. It is based on the PHQ-9. The test is used to determine how good the patient's mental health is in the real world. The questionnaires ask about the type of symptoms, the cause, as well as their treatment. A score is determined by adding the frequency of occurrence of every symptom. The score indicates how severe the mental health problems are. For example, nine symptoms may be mild, moderate, or severe. Every symptom is further classified into mild, moderate, or severe conditions. A method called the Clinical Symptom Elicitation Process (CSEP) is used. Following the evaluation of the questionnaire, the psychiatrist creates a rating score. The rating score indicates the severity of the patient's depression. In a typical CSEP procedure, the psychiatrist asks questions about each category and then evaluates the patient's responses to determine the frequency of the category, e.g., score0: never, score1: many days, score2: more than half the days, as well as score3: almost every day. The frequency is based on the patient's responses.

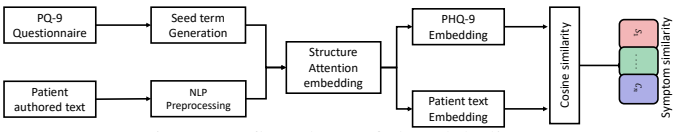


Fig. 3: A flowchart of data labeling.

For the data labeling task, the embedding explained in Section III-C is used, as shown in Fig. 3. The proposed model takes two texts, *PHQ-9 questionnaire* as well as *patient authorized text* as well as converts them into two vectors, both of which are latent sentiment representations. For the semantic expansion of *PHQ-9 questionnaire*, we used a WordNet-based seed term generation [20]. Then, the similarity of the vector *patient authorized* with all extended latent representations of symptoms $(S1 - S9)$ is calculated. The output of the model is a multi-class prediction based on the latent similarity of nine discrete symptoms.

C. Structural-aware Hypergraph

We have interpreted a patient's writing as a sequential directed acyclic graph G in which word points denote location and time. The edge between every node in this study reflects a series of words occurring sequentially, with the structure represented by the word nodes. A hypergraph represented the nodes in the graph. As shown in the designed framework, this paper uses a hypergraph to organize the words based on their semantic information. The edges have their attention network, the edge level attention, as well as the hyperedges have their attention network, which is called the node. The hyper-edges are organized into hyper-vertices for sentence embedding using the averaging approach. In the designed model, $G = (\mathcal{V}, \mathcal{E})$ stands for the set of nodes connected with two or more nodes, where $\mathcal{V} = \{v_1, \ldots, v_n\}$ stands for the set of nodes connected with two or more nodes. The hypergraph G has a structure that may be able to be described as an incidence matrix with the following entries:

[2]https://www.psychiatry.org/psychiatrists/practice/dsm
[3]https://www.uspreventiveservicestaskforce.org/Home/GetFileByID/218

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

$$\mathbf{I}_{xy} = \begin{cases} 1, & \text{if } v_x \in e_y \\ 0, & \text{if } v_x \notin e_y \end{cases} \tag{2}$$

Each as well as every node in G is a d-dimensional attribute vector. Therefore, a node in the graph is defined as a vector $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^{\mathrm{T}} \in \mathbb{D}^{n \times d}$. We may be able to represent the graph $G = (\mathbf{I}, \mathbf{X})$, which is used to state the hypergraph in the designed model. Note that only one-hot vector is then utilized by applying the structural word model; this strategy can thus preserve the sequential order of the authorized words of the patient.

Algorithm 2 Client level graph attention embedding

INPUT: C represents corpse, n represents number of words, w represents embedding size, m represents numbers of instances per batch, and K is the sampling data.

OUTPUT: Symptoms label c_i of $t_i \in Questionare$.

1: **while** $K \leq \{1, 2, ; \frac{n}{m}\}$ **do**
2: Select training samples from T;
3: **foreach** $synonym = 1, T\epsilon\ w$ **do**
4: $Emotionallexicon \leftarrow word2vec_{dep}(vocabulary, corpus, w, window = 2)$;
5: **end foreach**
6: $Struture_{graph} \leftarrow Structure_{hypergraph}()$;
7: $Attention_{PositionalEncoder} \leftarrow (Emotionallexicon,\ Struture_{graph})$;
8: Apply pooling strategy by embeddings;
9: Determine similarity by utilizing `cosine` matrix;
10: Update λ by the GD algorithm; ▷ GD is gradient descent
11: **end while**
12: **Return** $Gradientdescent$.

D. Structural graph embedding

The nodes are represented by latent learning representations after word-level modeling. We used the ordered degree sequence to collect the nodes $S \subset V$. The jumps between k locations in G are denoted by $T_k(x)$ nodes. For example, $T_1(x)$ denotes the collection of neighbors of vertex x when its distance is set to 1. $T_k(x)$ represents the extension of the node structure at a distance of k. Comparing the ordered degree sequences of the two vertices x as well as y (two vertices in the vertex network). $f_k(x, y)$ denotes the structural distance between x as well as y. Their neighborhoods are denoted as k (all words within a k radius). The function is defined as follows in Eq. 3.

$$\begin{gathered} f_k(x,y) = f_{k-1}(x,y) + g\left(s\left(T_k(x)\right), s\left(T_k(y)\right)\right) \\ k \geq 0 \text{ as well as } \left|T_k(x)\right|, \left|T_k(y)\right| > 0 \end{gathered} \tag{3}$$

Only if and only if both x as well as y have an edge at distance k is Eq. 3 defined. The function $f_k(x, y)$ may be able to be used to calculate the degree sequences of the nodes that are at the same distance from x as well as y using structure growth at a distance, as well as k may be able to aid in the computation of degree sequences of nodes that are at the same distance from x, as well as y. Using Dynamic Time Warping to calculate the distance between two ordered degree sequences (DTW). This strategy helps the extraction process of usable distance that discard the variants of lengths regarding the processed sequences; the sequence structure belongs to the loose compression [20]. In addition, DTW model is useful to choose the optimal alignment of the processed x and y sequences, in which both of them are considered as the growth sequences. For every and each term in the sequences, it is possible to define a distance function for the calculation, e.g., $d(x, y)$. DTW then aligns the sequences and calculates the minimum value of the sum of the distances among the matched terms [29]. This distance function can be discovered in Eq. 4. The reasons for this calculation is that the trajectory is then indicated by using the sequence degrees of a nodes connected with its neighbors.

$$d(a,b) = \frac{\max(x,y)}{\min(x,y)} - 1 \tag{4}$$

There is no distance between two identical nodes with ordered sequences ($x = y$ as well as $d(x, y) = 0$. The multilayer weighted network encodes the nodes as word sequences to generate contexts. The diameter of $k*$ hops is associated with the structure node $G = (V, L)$. We define the multilayer network by looking at the k hop neighbors of the node. The weights are assigned to the nodes using the function given above. The edge weight of a layer is defined by Eq. 5. In this approach, the data are labeled in the manner shown in Fig. 3.

$$w_k(x,y) = e^{-f_k(x,y)} \tag{5}$$

We have n_k^* vertices as well as at most $k^*\,(n2) + 2n\,(k^* - 1)$. Vertices are then used for weighted edges in the designed model. Contextual information about word order is generated using a multi-layer network. No labeling information was required to evaluate structural similarity based on the nodes. We explored the multi-layer network using a biased random walk, creating random selections as well as weighted sequences. Based on the probability ($q > 0$), the random walk decides whether to advance through the layers or stay in the current layer. Eq. 6 gives the probability that a node u connected to a node v in layer k will stay in the current layer.

$$p_k(x,y) = \frac{e^{-f_k(x,y)}}{Z_k(x)}, \tag{6}$$

in which $Z_k(x)$ represents the normalization factor for vertex x in layer k. It may be able to thus be defined by Eq. 7.

$$Z_k(x) = \sum_{\substack{v \in V \\ v \neq x}} e^{-f_k(x,y)} \tag{7}$$

Every step is explained in detail in the Algorithm 2. Given a phrase with node information, the gradient values are used to generate the training embeddings f_w as well as λ (Algorithm 2, input). We then extract lexicons using m samples (Algorithm 2, lines 2-6). The word2vec model with window

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

size = 2 is used to train the emotion lexicon. The emotion lexicon helps to convert nodes into vector representations (Algorithm 2, lines 3-5), which are then combined for the hypergraph as well as the attention network (Algorithm 2, lines 6-7). The word structure (Algorithm 2, line 7) is inserted into the text. Then, as described in Section III-D (Algorithm 2, lines 8-10), we create the pooling algorithm for the attention network [21]. Then, a small batch with average embedding is selected (Algorithm 2, line 10), as well as the similarity is computed to obtain the labels of the sentences (Algorithm 2, line 9). The gradient approach is then updated (Algorithm 2, line 10).

TABLE I: Experimental setup.

Device	Intel Core i7-9700K	RTX 2070
Architecture	Coffee lake	Turing (TU106)
Base clock	3.6 Ghz	1410 Mhz
Boost clock	4.6	1710 Mhz
Total cores	6	2304
Memory	16GB	8GB
Memory bandwidth	32GB/s	448GB/s
Deep learning library	Sklearn	Tensorflow
Language — library	Python 3.8	CUDA 8.0 — CUDNN

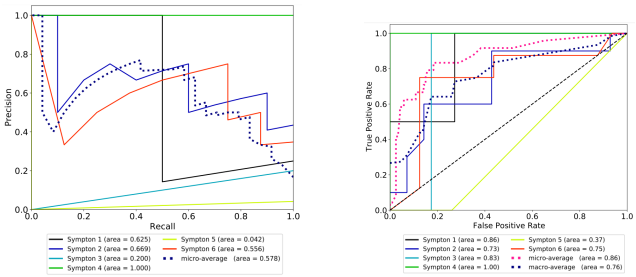
IV. Experimental Result as well as Analysis

In this research, we used the Table I processing configurations. In this paper, two feature extraction approaches are used. The first is based on an emotional lexicon, while the second is based on a structure-aware graph model. For vectorization, both models used a 300-dimensional glove vector. The embedding was used in the lexicon of nine symptoms to convert the text into node vectors. The structure embedding model then uses the hypergraph to extract word-based node patterns. The text is then labeled using the trained embedding depending on the question.

We used a dataset from several Internet forums as well as websites to classify the nine specific symptoms in this study [6]. The learned knowledge is extended with an entropy-based method that ensures that unusual events do not damage the proposed system. According to [6], the labeling is based on a nine-point PHQ-9 scale, where 0 indicates no depression, 1 indicates mild depression, 2 indicates moderate depression, as well as 3 indicates severe depression. For every symptom, we convert the labeling to a binary class, where 0 represents no symptoms, as well as 1 represents the presence of symptoms.

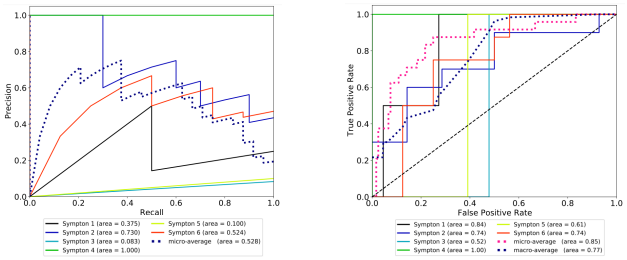
A. Model designation

A feed-forward neural networks is thus considered as a baseline for the result evaluation. The moderate strategy was used to maintain consistency in the length of comments. The developed approach is with $(30, 20, 10)$ hidden layers, as well as a ReLU activation function for the parameter settings. The main idea of the designed is to have outstanding performance in identifying 9 different symptoms by using the hypergraph and the emotional lexicon that are developed in the designed model. The final layer is a 9-link `sigmoid` function. The cross-entropy function is then considered as a loss function in the developed model.



(a) Precision Recall Curve (b) ROC curve

Fig. 4: Baseline model classification result.



(a) Precision Recall Curve (b) ROC curve

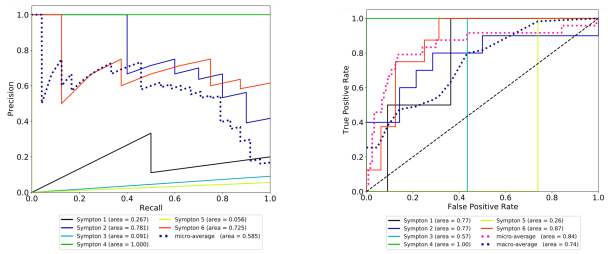
Fig. 5: LSTM model classification result.

To improve performance on sequential tasks, we used a recurrent neural network with gated units. The long-term memory (LSTM) architecture stores a complete record of the embedding sequence. Then, the process assigns one parameter to the forward roll and another to the backward roll. As a result, position encoders can operate in two modes: Input and Output. In addition to the LSTM units, we have used the attention position layer [20]. For regularization as well as overfitting reasons, we set the dropout ratio for the hidden LSTM layers to 0.5. To use the word importance [30], the sequencing of a hypergraph, as well as the emotional lexicon, are kept separate. The attention network helps extract structural information to select the essential word in the attention layer.

The baseline model has a ROC of 0.86. As shown in Fig. 4, the proposed feature extractor may be able to contribute to high accuracy. However, considering the precision-recall curve for every class, the model may be able to not detect clear patterns between classes. The model achieves a precision-recall curve of 0.81. At certain thresholds, the model curve fluctuates. Therefore, additional development is needed to achieve even better performance. The depression data is a sequential prediction task, where word sequences are ordered in order of relevance. Consequently, a design that prioritizes sequence as well as information storage may be able to achieve high performance.

The LSTM network outperforms the baseline model but suffers from the same fluctuation issue, as seen in Fig. 5. This shows that the cross-class learning of the model is not optimal, improving from 0.81 to 0.85. Due to the complexity of the LSTM cells, the gradient disappears as it goes from one cell to the next. Hyper-tuning, as well as extended run durations in the architecture, may be able to mitigate this problem. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics



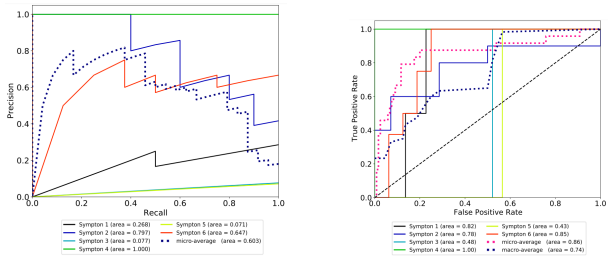
(a) Precision Recall Curve (b) ROC curve

Fig. 6: BILSTM model classification result.

network must be run for a more extended period to achieve human-level performance. The LSTM prefers minimal weight initialization as well as follows the baseline model's pattern.

The LSTM network is able to achieve a relatively high level of performance when directly compared to our baseline models as presented. Moreover, LSTM also exhibits a level of similar fluctuations as a problem given clearly in Fig. 5. From this, we can decipher that our cross-class learning methodologies in our model may not be considered as optimal, because it improves only from a 0.81 level to a 0.85 level. The well-known complex nature of LSTM's cells may have issues within the actual gradient being lost when it moves from 1 cell to a different one. In our architecture, the levels of hypertuning, as well as the run times being longer, may be able to assist in the reduction of this problem. The network itself may be able to run longer for the achievement of performance as close to known human levels as possible. The LSTM may actually prefer an initialization with a group of small weights, as well as follows the baseline model's pattern.

The bidirectional model performed well, as you may be able to see in Fig. 6. Based on the input, as well as output, the model worked in two ways. Two different RNN models demonstrate the model's ability to learn sequence problems. The testing set has the fewest errors, as well as the curve shows the top corner. This shows the minimum number of false positives, as well as false negatives. The forward, as well as backward motion of the BILSTM model helps maintain long-lasting dependencies that allow for long-term memory.



(a) Precision Recall Curve (b) ROC curve

Fig. 7: BILSTM with positional attention layer model classification result.

Fig. 7 shows positional attention with a bidirectional LSTM. The model was used to compute the position vector to understand the correlated high-quality words. The model makes the fewest errors. The Precision-Recall curve is 0.60, and the Receiver Operating Characteristic (ROC) value is 0.86. This shows the high percentage of true positives. The model includes the most important words that contribute most to the categorization. The network reduces the cost of vector computation by focusing on the most important words. A directional coder could teach the attention network to discover meaningful words. Individual mental health data are complicated. In addition, a larger vocabulary and grammatical combinations can help strengthen the attention network.

Table II, provides strategies for detecting depression associated with mental illness using texts [31], [32], [20], [33]. These approaches have put the strategy through its paces on several datasets (eRisk 2017 / 2018, Twitter, as well as Reddit). However, semi-supervised learning remains a viable strategy to improve the model. This study shows how node-level and edge-level hypergraphs can help in training techniques, and aims to extend the trainable situations through a semantic extension approach. The proposed model aims to reduce the data annotation overhead. Thus, the method helps to generalize the system. The semantic vectors are classified using the hypergraph information obtained from the context. The resulting word embedding uses the semantic information to select a subset of the unlabeled text. This technique finds instances using the unlabeled text generated during the hypergraph learning phase. The approach inserts the newly obtained training points into the initialization of the model.

B. *Federated learning analysis*

We evaluated the federated learning method based on the convergence and privacy factor. Due to the limited wireless resources in communication networks, only a subset of users may be active in uploading their local model parameters to the center in any single learning phase. However, due to the variability of training data among different users, the center would like to incorporate the local models of all users FL to create the best global FL model. Therefore, the scheduling of the upload by the users is crucial and affects the performance of FL as well as the convergence time. In contrast to the word embedding, the proposed model can improve performance by 0.2 to 1, as shown in Table II. However, it had an impact on the performance of the small dataset. This is because the proposed model needs to be adjusted to account for the different characteristics of the dataset. We do not need to use the embedding to fine-tune the categorization layer. The ultimate goal is to create a federated learning environment for medical data. However, the model can be improved by fine-tuning the hyperparameters. The federated environment allows learning to be embedded without sharing accurate data. The results show that learning medical data and applying it to classification can be improved with some hyper-tuning. The attention network may be able to signifimay be able to derive total relational lexicons. Machine learning has transformed data processing for large-scale applications in recent decades. At the same time, increasing privacy concerns for popular applications have led to a rethinking of traditional data processing methods. In particular, traditional machine learning uses centralized data training, where both the data is collected and the entire training

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2022.3172269, IEEE Journal of Biomedical and Health Informatics

TABLE II: Critical analysis of the methods.

Paper	Data-set	Method	Machine learning	Unsupervised	Adaptive learning	Performance
[31]	Online Forum	Feature based	Naive Bayes, Maximum Entropy, as well as Decision Tree	No	No	54.5
[32]	Erisk 17/18 — Anx 18/19	Bag-of-Words / Word Embeddings	One class SVM as well as KNN	No	No	62, 56, 77, 6.8
[20]	Amazon Mechanical Turk	Synonyms	Attention network	Yes	No	88.8
[33]	Erisk 17/18	TF / Embedding	Machine / Deep learning	No	No	66 / 61
Proposed	Amazon Mechanical Turk	Word Embeddings / Federated learning	Hyper-graph attention learning	Semi-Supervised	Yes	86

process is performed on a single server. Despite the great convergence, this training poses various privacy risks to the participants when they share their data with the centralized cloud server. In this regard, federated learning has surpassed distributed data training in importance. In particular, federated learning allows users to jointly train local models on local data without sharing sensitive information with the central cloud server.

Machine learning has transformed data processing for large-scale applications in recent decades. At the same time, increasing privacy concerns for popular applications have led to a rethinking of traditional methods for data processing. Ahmed *et al.* present a semantic vector based strategy for synonym expansion using a deep learning model [20]. Semantic vectors are grouped based on semantic information from the context in which they occur. Based on the semantic information, the resulting similarity metrics help in selecting a subset of unlabeled text. The proposed technique isolates the unlabeled text and includes it in the next cycle of the active learning process. The approach uses the new training points to update the model training. The cycle is repeated until the optimal solution is found. At this point, all unlabeled text is converted into the training set. In particular, centralized data training is used, where both the data is collected and the entire training process is performed on a single server. Despite the great convergence, this training poses several privacy risks to the participants when they share their data with the centralized cloud server. We proposed that the federated learning method surpasses distributed data training in importance. In particular, federated learning allows users to jointly train local models on local data without sharing sensitive information with the central cloud server.

Traditional methods such as encryption may be able to be used to ensure the security of the entire FL algorithm, as may be able to more recent developments such as secure multi-party computation as well as physical layer security, which may be able to provide security in situations (such as massively deployed IoT) where more conventional methods may be able tonot be used. The proposed model needs to be run over longer epochs to produce robust results. This helps to increase the weights per instance. However, it should also be noted that the associated set of features loses performance as the training time increases. When treated as a set of lexicons, all approaches perform well on the classification task. When combined with the attention network, the stacked model at the node and edge level is able to extract critical work that contains both emotional meaning and connections to the source class. The aggregation layers help to obtain the vector representation of the node and informative edge words. The technique analyzes inspirational words to form phrases, which are then concatenated to identify symptoms in the text or conversation. When some words are misleading and others are important, the important words are given more weight than their neighbors. The results show that adding a larger vocabulary as well as grammatical changes can improve the direct performance of our model. Our proposed novel techniques are able to directly learn the important sequences and, based on this, add attention-based known weights to any node- and edge-level vector representation. Our given representations are then directly combined with words (others) to directly form our proposed sentence vectors. This direct aggregation vector can then contain all the semantic meaning as well as the pattern information needed for instance class classification. Our proposed system then achieves a direct weighted F1 score of 0.86 as well as synonym expansion, which improves the overall training accuracy.

Moreover, we can say that the developed lexicon for emotions can help to reduce both overfitting and generalization. Our hierarchical attention model based on vector representation support with our node and edge based hypergraph. From this, we can infer that node attention directly contributes to an emotional event being the triggering event, while edge attention, in contrast, can directly contribute to the context of the analyzed text. The tool we created can then be used for any online adaptive intervention used during an online virtual session on mental illness. Any psychiatrist can provide the tools needed for reference notes.

V. Conclusion

Both Natural Language Processing (NLP) and Deep Learning (DL) have been applied to clinical text analysis in recent years. Symptoms are retrieved from text data written by patients. In addition, strategies for mental health concepts are not presented. We used structural data to solve a sequencing problem. The structural hypergraph, in which nodes explore the structure of their neighbors, was used, which is effective in the LSTM model based on attention. We also embedded the model in an emotional lexicon to generate a high-level node sequence. The structural embedding then supports the weighted attention network. The strategy based on federated mean learning can completely reduce the global loss of joint weights. The experimental results show that federated learning has practical advantages over classical supervised learning approaches. The model may be able to perform better at the local level without the need to share and disseminate the raw data throughout the network. This approach has the potential for wide application. The experimental results clearly show that the model performs better in direct comparison with other models, with an F-value of 0.86. To improve the selection for language lexicons, we will perform automatic embedding selection before training in the near future. This study will be extended to include additional factors such as the

patient's geographic, cultural, religious, and social background to provide more information about the associated conditions. In addition, a similarity learning network will be used to investigate the scalability of the model. This will help to develop a more robust model for data accuracy, reliability, and integrity.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *The International Conference on Artificial Intelligence and Statistics*, A. Singh and X. J. Zhu, Eds., vol. 54, 2017, pp. 1273–1282.

[2] S. L. James, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J. Abdela, A. Abdelalim *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017," *The Lancet*, vol. 392, no. 10159, pp. 1789–1858, 2018.

[3] M. G. Mazza, R. D. Lorenzo, C. Conte, S. Poletti, B. Vai, I. Bollettini, E. M. T. Melloni, R. Furlan, F. Ciceri, P. Rovere-Querini, and F. Benedetti, "Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors," *Brain, Behavior, and Immunity*, vol. 89, pp. 594–600, 2020.

[4] S. K. Mukhiya, J. D. Wake, Y. Inal, K. I. Pun, and Y. Lamo, "Adaptive elements in internet-delivered psychological treatment systems: Systematic review," *Journal of Medical Internet Research*, vol. 22, no. 11, p. e21066, 2020.

[5] S. K. Mukhiya, J. D. Wake, Y. Inal, and Y. Lamo, "Adaptive systems for internet-delivered psychological treatments," *IEEE Access*, vol. 8, pp. 112 220–112 236, 2020.

[6] S. K. Mukhiya, U. Ahmed, F. Rabbi, K. I. Pun, and Y. Lamo, "Adaptation of IDPT system based on patient-authored text data using NLP," in *IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 2020, pp. 226–232.

[7] E. A. Troyer, J. N. Kohn, and S. Hong, "Are we facing a crashing wave of neuropsychiatric sequelae of COVID-19? neuropsychiatric symptoms and potential immunologic mechanisms," *Brain, Behavior, and Immunity*, vol. 87, pp. 34–39, 2020.

[8] C. Karmen, R. C. Hsiung, and T. Wetter, "Screening internet forum participants for depression symptoms by assembling and enhancing multiple NLP methods," *Computer Methods and Programs in Biomedicine*, vol. 120, no. 1, pp. 27–36, 2015.

[9] A. Neuraz, I. Lerner, W. Digan, N. Paris, R. Tsopra, A. Rogier, D. Baudoin, K. B. Cohen, A. Burgun, N. Garcelon *et al.*, "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*, vol. 22, no. 8, p. e20773, 2020.

[10] H. Cheng, X. Yan, J. Han, and C. Hsu, "Discriminative frequent pattern analysis for effective classification," in *IEEE International Conference on Data Engineering*, R. Chirkova, A. Dogac, M. T. Özsu, and T. K. Sellis, Eds., 2007, pp. 716–725.

[11] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. L. Rubin, and J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 945–954, 2018.

[12] J. Konecný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *CoRR*, vol. abs/1610.02527, 2016.

[13] D. Połap, G. Srivastava, A. Jolfaei, and R. M. Parizi, "Blockchain technology and neural networks for the internet of medical things," in *IEEE Conference on Computer Communications Workshops*. IEEE, 2020, pp. 508–513.

[14] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," *Future Generation Computer Systems*, 2020.

[15] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine llearning: concept and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 12:1–12:19, 2019.

[16] T. M. H. Li, M. Chau, P. W. C. Wong, and P. S. F. Yip, "A hybrid system for online detection of emotional distress," in *Intelligence and Security Informatics*. Springer, 2012, pp. 73–80.

[17] E. Chen, K. Lerman, and E. Ferrara, "Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data set," *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. e19273, 2020.

[18] M. McDonnell, J. E. Owen, and E. O. Bantum, "Identification of emotional expression with cancer survivors: Validation of linguistic inquiry and word count," *JMIR Formative Research*, vol. 4, no. 10, p. e18246, 2020.

[19] G. Nguyen, S. Dlugolinsky, M. Bobák, V. D. Tran, Á. L. García, I. Heredia, P. Malík, and L. Hluchý, "Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, 2019.

[20] U. Ahmed, S. K. Mukhiya, G. Srivastava, Y. Lamo, and J. C.-W. Lin, "Attention-based deep entropy active learning using lexical algorithm for mental health treatment," *Frontiers in Psychology*, vol. 12, p. 471, 2021.

[21] U. Ahmed, J. C.-W. Lin*, and G. Srivastava, "Fuzzy contrast set based deep attention network for lexical analysis and mental health treatment," *Transactions on Asian and Low-Resource Language Information Processing*, 2022.

[22] Z. Lian, W. Wang, H. Huang, and C. Su, "Layer-based communication-efficient federated learning with privacy preservation," *IEICE Transactions on Information and Systems*, vol. 105, no. 2, pp. 256–263, 2022.

[23] J. Song, W. Wang, T. R. Gadekallu, J. Cao, and Y. Liu, "Eppda: An efficient privacy-preserving data aggregation federated learning scheme," *IEEE Transactions on Network Science and Engineering*, 2022.

[24] W. Wang, M. H. Fida, Z. Lian, Z. Yin, Q.-V. Pham, T. R. Gadekallu, K. Dev, and C. Su, "Secure-enhanced federated learning for ai-empowered electric vehicle energy prediction," *IEEE Consumer Electronics Magazine*, 2021.

[25] W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, and C. Su, "Blockchain-based reliable and efficient certificateless signature for iiot devices," *IEEE transactions on industrial informatics*, 2021.

[26] Y. Yuan, L. Zhang, L. Wang, and H. Huang, "Multi-level attention network for retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[27] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: Validity of a brief depression severity measure," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[28] W. H. Organization *et al.*, *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*. World Health Organization, 1993, vol. 2.

[29] T. Han, S. Niu, and P. Wang, "Multimodal-adaptive hierarchical network for multimedia sequential recommendation," *Pattern Recognition Letters*, vol. 152, pp. 10–17, 2021.

[30] Y. Huang, J. Chen, S. Zheng, Y. Xue, and X. Hu, "Hierarchical multi-attention networks for document classification," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 6, pp. 1639–1647, 2021.

[31] L. Xu, R. Jin, F. Huang, Y. Zhou, Z. Li, and M. Zhang, "Development of computerized adaptive testing for emotion regulation," *Frontiers in Psychology*, vol. 11, p. 3340, 2020.

[32] J. Aguilera, D. I. H. Farías, R. M. Ortega-Mendoza, and M. Montes-y Gómez, "Depression and anorexia detection in social media as a one-class classification problem," *Applied Intelligence*, pp. 1–16, 2021.

[33] R. M. Ortega-Mendoza, D. I. Hernández-Farías, M. Montes-y Gómez, and L. Villaseñor-Pineda, "Revealing traces of depression through personal statements analysis in social media," *Artificial Intelligence in Medicine*, vol. 123, p. 102202, 2022.