# Learning Graph Representation for Predicting Student Mental Wellbeing in Robot Assisted Journal Writing Context

Name Surname
*dept.*
*organization*
City, Country
email address

Name Surname
*dept.*
*organization*
City, Country
email address

Name Surname
*dept.*
*organization*
City, Country
email address

Name Surname
*dept.*
*organization*
City, Country
email address

***Abstract*—Conversational agents are increasingly introduced in mental health and well-being research. One promising application is in assessing students' mental well-being while facilitating self-disclosure activities such as administering questionnaires or assisting in journaling. In these applications, the data collected is mostly multimodal and strategies for fully harnessing the potential of this data is an open challenge. Moreover, most of the machine learning (ML) research within healthcare has been carried out using conventional, baseline models which fail to capitalise on the inherent graphical characteristics of health data. Learning from graph representations encompasses techniques such as feature propagation and aggregation methods to combine feature information and graph structure in order to learn better data representations. Therefore, in this paper we propose a novel graph representation learning strategy to predict student mental health status from multimodal data in a robot assisted journal writing context. Twenty undergraduate / graduate students interacted with the Pepper robot in a dyadic interaction setting over four weeks (one session per week). Graph representation learning was applied on their questionnaire data (mental wellbeing state, mood changes, perceptions toward Pepper, and motivational state) and physiological data (electro-cardiogram). Comparative experimental results demonstrate that graph representation learning based classification enables mental wellbeing assessment with an accuracy of 92.5% which is superior to standard ML methods that do not take into account graphical structures.**

***Index Terms*—social robot, self-disclosure, journal writing, student mental health, graph representation learning**

## I. Introduction

The significance of university students' mental health and their academic performance has prompted a need for comprehensive institutional strategies to support mental health across universities [1]–[3]. However, accurately measuring and monitoring student well-being remains a challenge despite significant investments in support services [1], [4]. There are disparities among institutions in how they assess well-being [5], and the higher education sector is behind in directly measuring well-being [6]. Several surveys, including The WHO World Mental Health (WMH) International College Student (WMH-ICS) project [7], as well as others [8], [9], employ traditional paper-and-pen or online self-reporting techniques to evaluate the mental well-being of college students. While these surveys have been successful in understanding the prevalence of mental health issues in college students, they depend greatly on the assumption of the accuracy of self-reported responses [10]. For example, stigma and discrimination may present substantial obstacles to seeking care and openly reporting symptoms [11], potentially impeding students' access to treatment and leaving mental health issues unaddressed.

Journal writing has frequently been employed in mental health as an intervention method. However, within therapy, the journal serves as a distinctive tool for continuous evaluation throughout the intervention process [12]. By utilizing a journal, the individual is provided with chances to articulate personal thoughts, emotions, and concepts in a setting that might feel more secure than direct self-disclosure [12]. Additionally, such disclosures are important to provide contextual information to mental health assessment questionnaires which is required for a complete diagnosis and can be currently accessed mostly by clinical judgement and training [13]. On the other hand, among all the possible advantages of journal writing as an assessment tool, it is suggested that the guidance provided in planning, writing, and discussing the journal is important to benefit from it in full extent [12]. However, shortage of mental health professionals in the universities create barriers in providing such support for students in their mental wellbeing.

Conversational agents are increasingly introduced in mental health and well-being research. One promising application is in assessing students' mental well-being while facilitating self-disclosure activities such as administering questionnaires or assisting in journaling [14], [16], [17]. These agents can

X. XXXXX is supported by XXXXXXXX XXXXXXXXXXXXXXXXXXXXXXXX, XXXXXXX XX XXXXXXXXXXX XXX XXXXXXXX, XXXXXXXXXXXXXX XXX XXXXXXXXXX XXXX, XXXXXXX X XXXX XXXX XXXXXXXXXXX XXXXXX XXXXX XXXXXXXXXXX XXX XX XXXX XXXXXXX XXX XXXXXXXXXXX XXXXXX XXXXX XXXXXXXXXXX XXXXXXXXXX XXXXXXXXXX XXXXXXXXXXXX. X. XXXXX is supported by XXX XXXXXXXXX XXXXX XXXXX XXX. XXXXXXXXXXXXX XXXXXXX.

alleviate stress and fear associated with potential judgement, thereby reducing emotional barriers to disclosure [15], [73], and enhancing accessibility to therapeutic services [18], [19], [21]. However, in applications involving these agents, the gathered data is predominantly multimodal, presenting a challenge in fully exploiting its potential. Additionally, much of the machine learning (ML) research in healthcare has relied on traditional, conventional models that overlook the intrinsic graphical aspects of health data [62]. Therefore, in this paper we utilise graph representation learning to predict student mental health status using multimodal data in the context of robot-assisted journal writing. Our contributions are two-fold. We show for the first time in the literature that it is possible to assess student mental wellbeing in journal writing context from robot administered questionnaires and physiological reactions. We propose a graph representation learning methodology for mental wellbeing prediction and demonstrate that it is more effective than conventional machine learning techniques that do not leverage graphical structures.

## II. Literature Review

### A. Conversational Agents in healthcare and mental wellbeing

[14] did a study in which university students engaged in daily journaling with a chatbot prompting them to record moods and experiences for three weeks, effectively motivating disclosure. Another study [20], proposed an intelligent chatbot analysing students' chat to determine mental states. Additionally, researchers [18], [21] developed a robotic coach delivering positive psychology interventions in on-campus dormitories to enhance psychological well-being. Moreover, research [11] demonstrated that individuals tend to disclose more about depressed thoughts when interviewed by a virtual agent compared to face-to-face interviews. [22] showed that especially on sensitive topics such as drugs and gender, people disclose more to an embodied virtual agent (completing a chatbot-based survey via written language responses) than to a human. Another study [24] explored using computational methods to assess children's well-being during interactions with robots, showing promise in analysing audio responses to robot-assisted procedures. Thus, conversational agents can help make university students express their thoughts and feelings without fear of being stigmatised.

### B. Machine Learning for Mental Wellbeing

With the advance of machine learning and deep learning, many computational models have been used to learn representations from different modalities such as questionnaires, video/audio and physiological data for recognizing mental health disorders [26]–[32], [59], [61]. Recent applications include the use of ML algorithms for improvement of virtual medical assistants who provide support to individuals with mental health problems [68], [69]. One of the primary focuses of this body of research is the application of ML to diagnosis tasks with some of the most commonly assessed conditions being schizophrenia, depression, bipolar disorder, anxiety and loneliness [63], [71]. However, the overall quality of existing mental health prediction models is generally poor. Moreover, most of the machine learning research within healthcare has been carried out using standard, baseline models such as Random Forest and Support Vector Machines which fail to capitalise on the graphical characteristics of health data [62].

### C. Why Graph Representation for Wellbeing Assessment?

Graph can well and explicitly model the relationships among: (i) different modalities; and (ii) status at different time stamps, for multi-modal time-series data analysis. It allows the target components/objects to be represented by graph nodes and thus the relationship between each pair of components (e.g., modalities or features of different time stamps) can be explicitly modelled via graph edges during the propagation of the employed Graph Neural Networks (GNNs). Compared to standard machine learning models (e.g. CNNs), they can effectively capture comprehensive relationships within a single modality as well as relationships across different modalities [33]. Such comprehensive relationships between each pair of nodes can be well and explicitly described by the deep-learned multi-dimensional edge features [34]. Because of such advantages, graph-based machine learning systems have been increasingly utilized in mental health prediction tasks [35], [36], [64], [65]. For instance, in [35], GCN and GCN-based diff-pool are used on fNIRS data for depression cognition. They focus on GNN's advantage to model dependency among graph nodes and use this to extract the connection between brain function regions and channels. [37] introduced a novel node-based antagonistic learning approach for the task of autism recognition. [39] created a semi-supervised Graph Instance Transformer(SS-GIT) for preliminary mental illness detection with mobile sensing data. Later, [40] proposed a BrainNN framework embedding multi-modal connectivities between brain regions for GNN to diagnose mental illness. [41] built a GNN model to accurately learn features from EEG signals for Major Depressive Disorder (MDD) detection and introduced a self-attention graph pooling process. Recently, a Multi-modal fusion framework MS2-GNN [42] explored the modal-shared and modal specific information on the EEG and audio data to detect MDD, showing a promising performance. However, none of these models were developed to predict student mental health status from physiological and subjective measures in robot assisted journal writing context. Accordingly this study proposes a novel graphical ML approach in this context and using multi-dimensional edge feature to describe the relationship between each pair of nodes while existing approaches generally used one value to represent each edge.

## III. Methods

### A. Participants

Our dataset consists of data from 4 sessions which are spread over four weeks (1 session per week). In total 20 students (7 female, 11 male, 1 non-binary, 1 transgender male), 19 to 43 years old (M= 23.35 ± 5.38) (9: undergraduate students, 11: graduate students) interacted with a Pepper robot

[43]. We recruited all students through flyers, posters displayed in corridors, the university website, and newsletters distributed on campus. Each student who participated in the study provided written informed consent before joining. All participants completed four sessions and received a 30X incentive for their involvement.

### B. Protocol

The study protocol received approval from the University Ethics Committee. Prior to the first session, students completed online questionnaires, including demographic information (DemographicQ) and self-assessment using The Warwick-Edinburgh Mental Wellbeing Scales (WEMWBS [33]). Additionally, they were introduced to Pepper through a brief online video. The study was conducted in a designated room within the department, where each student engaged in dyadic interaction with Pepper under the supervision of the experimenter, who monitored the interaction from a hidden control room. Each student was requested to seat on a chair positioned 1.5 meters away from the Pepper, which was placed directly in front of them. Tasks and activities were presented on a laptop in front of each student, operated using PsychoPy (v2021.2.3). As presented in Fig. 1, students completed a total of 4 sessions (2 with questionnaire activity (Q), 2 with questionnaire activity (Q) + journal writing (J)) across 4 weeks. For Q and Q+J, Pepper administered the following questionnaires: the Positive and Negative Affect Schedule (PANAS) [45] and the Short Depression-Happiness Scale (SDHS) [46]. Additionally, in Q+J, Pepper administered the journal writing activity adapted from [47]. During the sessions, Pepper verbally presented statements from PANAS (e.g., "you have felt interested during the past week") and SDHS (e.g., "you have felt dissatisfied with your life"), while students verbally expressed their responses. For PANAS, response options included "Very slightly or not at all", "A little", "Moderately", "Quite a bit", and "Extremely", while for SDHS, options were "Never", "Rarely", "Sometimes", and "Often". These response options were displayed on the laptop screen, serving as visual aids and eliminating the need for students to memorize them. Each session (refer to Fig. 1) began with students becoming familiar with the tasks. As the students were being familiarised with the tasks, neurotechnological device Superhero Collar [48] for collecting physiological data (electrocardiogram (ECG)) were placed on participants. Subsequently, the experimenter exited the room to supervise the interaction from the control room. Operating via a Wizard of Oz (WoZ) setup, the same experimenter controlled Pepper from the control room. The activities assisted by Pepper (J and Q) were pre-scripted and integrated into the WoZ setup. Before and after Pepper-assisted activities during each session, students completed the Brief Mood Introspection Scale (BMIS) on the laptop, which computes sub-scores for Pleasant Unpleasant, Arousal-Calm, Positive-Tired, and Negative-Relaxed Mood [49]. At the end of each session, students were prompted on the laptop to assess their perceptions of Pepper using the Robot Social Attributes Scale (RoSAS) [50]. Additionally, participants were asked to complete a Self-concordant motivation questionnaire (SCM) on the laptop to gauge their motivation to continue engaging in the Pepper-assisted activities [51]. To counterbalance the potential order effect of the activities, the assignment of the order of Q (week 1 & week 2) and Q+J (week 1 & week 2) was randomized. Within the second and final sessions, participants were requested to fill in the same self-assessment questionnaire (WEMWBS) completed at the beginning of the study.

## IV. Data Analysis & Prediction

### A. Mental Wellbeing Labelling from WEMWBS

We evaluated the student's mental wellbeing state after each Q and Q+J block using the self-report results of the WEMWBS questionnaire, which provides an outcome measure of the mental wellbeing of the population and has already been widely used in National surveys in X and in other countries [52]. Students completing the scale are required to choose their experience of each statement (e.g ''I've been feeling optimistic about the future") over the past two weeks using a 5-point Likert scale (ranging from "none of the time" to "all of the time"). The overall score for the WEMWBS is calculated by totalling the scores from each item (1 to 5 respectively) giving a minimum score of 14 and maximum score of 70. A higher WEMWBS score therefore indicates a higher level of mental well-being state [53]. WEMWBS has been validated against established measures of depression, and it's feasible to propose scores that indicate possible and probable clinical conditions on this scale [54]. Accordingly, we set the score of $^{<}$ 41 as indicative of probable clinical depression, a score of 41-44 as indicative of possible/mild depression, and score of 44 $^{>}$ as indicative of no depression state. For each Q and Q+J block, we then classified all the data collected from students into "probable clinical depression", or "possible/mild depression", ''no depression". Data from one Q block and one from Q+J block were excluded from the analysis due to missing WEMWBS measures. This resulted in the following grouping, for the Q block: 12 participants belonged to the "no depression" category, 6 belonged to the "possible/mild depression" category, 1 belonged to the "probable clinical depression" category; for the Q+J block: 12 participants belonged to the "no depression" category, 3 belonged to the "possible/mild depression" category, 4 belonged to the "probable clinical depression" category. Due to the dataset's limited size and unbalanced distribution, we chose to constrain our task to binary classification in the following sections bringing "possible/mild depression" and "probable clinical depression" into the same category as "depression".

### B. Predicting Mental Wellbeing from PANAS and SDHS

In the process of development and validation of WEMWBS, [53] showed that scores from PANAS [45] and SDHS [46] correlate with the scores from WEMWBS with the data from the student sample collected in an interval of one week. The PANAS scale assesses emotional well-being with 20 items, divided into Positive (PA) and negative (NA) affect dimensions. Scores may range from 10 – 50 and a very low

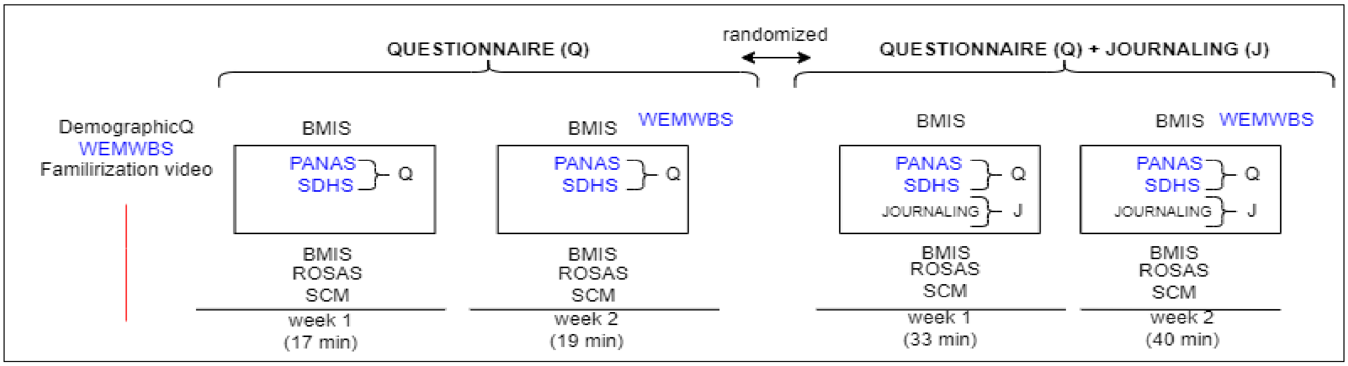


Fig. 1. Diagram of the experimental design. WEMWBS, The Warwick-Edinburgh Mental Wellbeing Scales; BMIS, Brief Mood Introspection Scale; PANAS, Positive and Negative Affect Schedule; SDHS, The Short Depression-Happiness Scale; RoSAS, Robot Social Attributes Scale; SCM, Self-concordant motivation.

level of PA might be a clinical concern and a high score on the NA might be an indicator of psychological distress. The SDHS is a six-item scale assesses well-being as a continuum between the two states of depression and happiness. Scores may range anywhere from 6 – 24 and a score $<$ 10 might be an indication of mild but clinically relevant depression. However, it is important to note that each of these measures can explain the wellbeing state at a certain level and uniquely as it can be also seen from the degree of correlations (ranges from medium to high) between these questionnaires [53]. Accordingly, we could see a possible inconsistency between these questionnaires when we also checked overall scores from our PANAS and SDHS findings (see Fig. 2, 3, 4). Regarding PANAS, when we checked all the scores from PA and NA, none of the participants, neither in week1 nor week2 reported a very low level of PA which could be a clinical concern. Similarly, a high score of NA which might be an indicator of psychological stress was not reported. Regarding SDHS, none of the participants, neither in week1 nor week2 reported score $<$ 10 which might be an indicative of mild but clinically relevant depression. However, on the contrary, some of the WEMWBS scores reveal that some of the participants might already belong to the "possible/mild depression" or "probable clinical depression" category. As a result it is worth investigating further the unique properties of PANAS and SDHS questionnaires using computational approaches especially using graph representation learning in terms of predicting WEMWBS scores including also other questionnaires and modalities such as physiological reactions.

### C. Graph Representation Learning

As detailed in section I.C, graph representation has a great potential for learning to predict mental wellbeing from structured data such as questionnaires. Therefore, we propose the construction of graph representation to encode the PANAS questionnaire and compare it with baseline models such as Support Vector Machines (SVMs) and Multilayer Perception (MLP) using classification accuracy and Unweighted Average Recall (UAR) as in [77]. Secondly, we propose the construction of graph representation to encode all subjective and

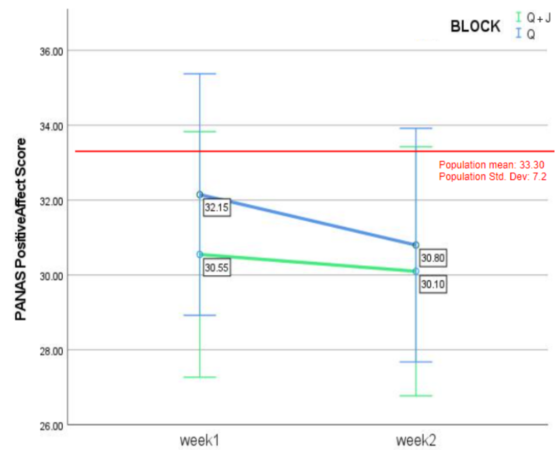


Fig. 2. Mean PA scores, in Q & Q+J, in week1 & week2. Scores may range from 10–50. Although the sample means for each week is lower than population means, the distributions do not reveal any data points with a very low PA which might be a clinical concern.

physiological measures separately and combined, using raw and processed versions of physiological data.

*a) Questionnaire Data:* As a first step, considering the possible relationship between WEMWBS and other subjective measures, we designed a node model which takes each week data as a node, taking into account the temporal relationship between nodes (week 1 to week 4). While doing so the node model changes from a graph result to node result with the WEMWBS labels at the end of each Q and Q+J block which results in a total of 20 graph representations. Accordingly the classification task was formulated as a 4-class problem as follows – class1: [no_depression, no_depression] (n=11), class2: [no_depression, depression] (n=2), class3: [depression, no_depression] (n=2), class4: [depression, depression] (n=5). For example, if a student's WEMWBS label after Q block was no_depresssion but after Q+J block was depression, this student's data was labelled as class2. If one of the labels were missing (after Q or Q+J), the missing label was replaced with the label from the available block. To validate the effectiveness of our strategy, we also trained and evaluated baseline models (SVM and MLP). For these models we used the data from Q

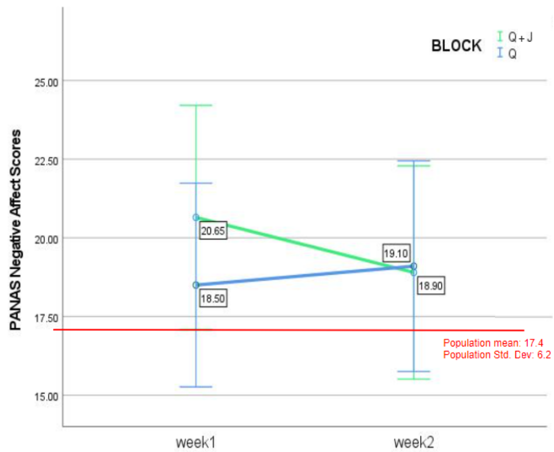


Fig. 3. Mean NA scores, in Q & Q+J, in week1 & week2. Scores may range from 10–50. Although the sample means for each week is higher than population means, the distributions do not reveal any data points with a high NA which might be an indicator of psychological distress.

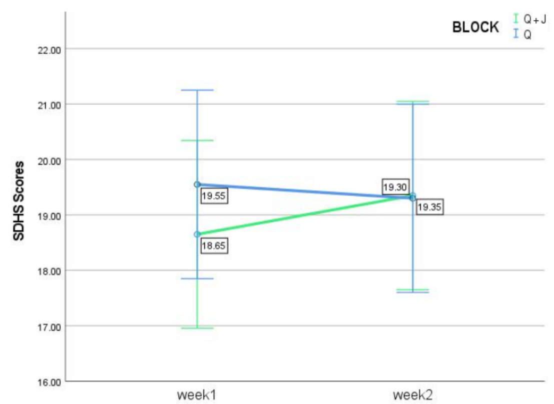


Fig. 4. Mean SDHS scores, in Q & Q+J, in week1 & week2. Scores may range from 6–24. The distributions do not reveal any data point with a score lower than 10 which might be an indicative of clinically relevant depression.

and Q+J block separately and formulated the classification task as a 2-class problem with class1: [depression] (n=26), class2: [no_depression] (n=52). We first trained and evaluated the above-mentioned approaches with responses from the PANAS questionnaire (see Fig. 5). The response from each PANAS item was used as input to the classifiers instead of the overall summed questionnaire score. For the graph model, we utilised the Adam optimizer with a learning rate of 1e-3, trained the model for 1000 epochs and split the data into 80% training and 20% test sets, with shuffling performed before each epoch for robustness and run it 10 times. As the dataset is not large, we utilised 10-fold cross validation to ensure the models were not over-fitting, similarly to the approach in [34]. Comparative results are presented in Table I.

We followed the same model formulation for other questionnaires, separately and in combination, as shown in Table II (see models from (1) to (10)). In our analysis, we defined difBMIS score as $difBMIS = postBMIS - preBMIS$. For each questionnaire which is contributing to the formation of node

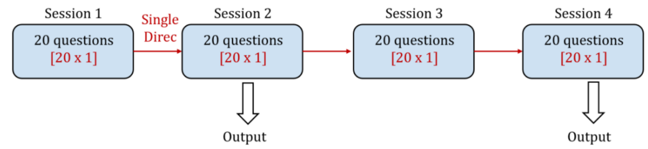


Fig. 5. Forming the node model based on responses from PANAS.

models mentioned previously, we took the response from each questionnaire item instead of having an overall questionnaire score as described earlier. In the process of combining the data from each questionnaire, we used a padding approach to ensure that questionnaire items are fitted to the same length to fill in the graph neural network. For example, PANAS has 20 questions, while SDHS has 6 questions, so padding is applied (see Fig. 6) prior to filling in the graph neural network.

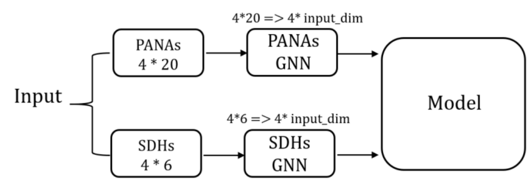


Fig. 6. PANAS & SDHS padding using GNN (hyper parameter input_dim).

*b) Psychophysiology Data:* As greater emotional expressivity can be associated with greater physiological responses [25], we incorporated physiological measures in the prediction task. We used the ECG data recorded while participants were answering the PANAS questionnaire. We also investigated the effect of using raw and processed versions of ECG signal in the prediction task, as there are mixed findings in the literature regarding using raw or filtered physiological signals [55]–[57]. For the filtered ECG signal we used the band-pass filters (0.05-45Hz) with sample entropy to verify the quality and a band-pass filter for external Power-Line Interference (PLI) as specified in [57]. As we found that raw ECG signal does not have those common 50-60 hz and low-frequency noises, we did not apply further filtering. Moreover, as the measured ECG signals were different in length for each session (e.g., answering duration of the PANAS questionnaire differed between each participant and in each session), before incorporating the ECG signal in node models as a part of a variable-length input task, we used attention weights and LSTM to pad the signals into the same length as in [58]. As in the previous classification experiment, we first validated the effectiveness of our strategy on ECG_raw by comparing it with the baseline models (SVM and MLP) (see Table I). Seeing the promising results from this comparison study, we continued applying our node model formulation to other measures as shown in Table II (see models from (11) to (17)). For each node model, we ran the analysis 10 times using the same approach mentioned in the previous section.

TABLE I

COMPARISON OF BASELINE MODELS (SVM AND MLP) AND GNN MODEL FOR PREDICTING STUDENT MENTAL HEALTH STATUS FROM PANAS AND ECG_RAW. REPORTING IS WITH METRICS SUCH AS ACCURACY (ACC) AND UNWEIGHTED AVERAGE RECALL (UAR) (PRESENTED IN %). VALUES IN BOLD DENOTE THE HIGHEST ACCURACY FOR THE SPECIFIC MODEL

| | **Fold1** | | **Fold2** | | **Fold3** | | **Fold4** | | **Fold5** | | **Fold6** | | **Fold7** | | **Fold8** | | **Fold9** | | **Fold10** | | **Avg** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **PANAS** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** |
| (1) MLP | **100** | **100** | 43.8 | 87.5 | **66.7** | **75** | 78.6 | 75 | **75** | 87.5 | 50 | 50 | **75** | **75** | **100** | **100** | **83.3** | 87.5 | 50 | **75** | 72.2 | 81.3 |
| (2) SVM | **75** | 87.5 | 37.5 | 62.5 | **66.7** | **75** | 50 | 75 | 66.7 | 75 | 66.7 | 37.5 | 37.5 | **75** | 73.5 | 75 | **83.3** | 87.5 | 50 | 62.5 | 60.7 | 71.3 |
| (3) GNN | 75 | **100** | 75 | **100** | 50 | **75** | **100** | **100** | **75** | **100** | **100** | **100** | **75** | **75** | 50 | **100** | 75 | **100** | **75** | **75** | **75** | **92.5** |
| **ECG_raw** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** | **UAR** | **Acc** |
| (1) MLP | 33.3 | 50 | 58.5 | 62.5 | **100** | **100** | **71.4** | **50** | 62.5 | 62.5 | **62.5** | **62.5** | 50 | 50% | 66.7 | 62.5 | **80** | **75** | **50** | 87.5 | 63.5 | 66.3 |
| (2) SVM | 33.3 | 50 | 41.7 | 62.5 | 42.9 | 75 | **71.4** | 50 | 50 | 50 | 37.5 | 37.5 | 50 | 50 | 66.7 | 62.5 | 70 | 50 | 42.9 | 75 | 50.6 | 56.3 |
| (3) GNN | **75** | **75** | **75** | **75** | **100** | **100** | 50 | **50** | **75** | **75** | 50 | 50 | **75** | **75** | **75** | **100** | 75 | **75** | **50** | **100** | **70** | **77.5** |

TABLE II

GNN MODELS FOR PREDICTING STUDENT MENTAL HEALTH STATUS FROM PHYSIOLOGICAL AND SUBJECTIVE MEASURES. REPORTING IS WITH THE METRIC ACCURACY (ACC) (PRESENTED IN %). VALUES IN BOLD DENOTE THE HIGHEST ACCURACY FOR THE SPECIFIC MODEL.

| **Models with Questionnaire data** | **Fold1** | **Fold2** | **Fold3** | **Fold4** | **Fold5** | **Fold6** | **Fold7** | **Fold8** | **Fold9** | **Fold10** | **Avg Acc** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) PANAS | **100** | **100** | 75 | **100** | **100** | **100** | **75** | **100** | **100** | **75** | **92.5** |
| (2) SDHS | 50 | 50 | 25 | 25 | 25 | 25 | 25 | 50 | 25 | 50 | 35 |
| (3) ROSAS | 75 | 75 | **100** | 50 | 50 | 75 | 50 | 50 | 50 | **75** | 65 |
| (4) SCM | 25 | 50 | 25 | 25 | 25 | 25 | 50 | 25 | 25 | 25 | 30 |
| (5) difBMIS | 75 | 75 | 50 | 50 | 50 | 50 | **75** | 50 | 50 | 50 | 57.5 |
| (6) PANAS + SDHS | 50 | 50 | 12.5 | 62.5 | 37.5 | 62.5 | 25 | 25 | 37.5 | 50 | 41.3 |
| (7) PANAS + ROSAS | 50 | 50 | 50 | 62.5 | 25 | 37.5 | 50 | 25 | 25 | 50 | 42.5 |
| (8) PANAS + SCM | 62.5 | 50 | 37.5 | 50 | 37.5 | 50 | 37.5 | 25 | 37.5 | 62.5 | 45 |
| (9) PANAS + difBMIS | 37.5 | 25 | 37.5 | 50 | 50 | 62.5 | 50 | 37.5 | 37.5 | 50 | 43.8 |
| (10) All questionnaires | 35 | 40 | 40 | 55 | 55 | 55 | 45 | 35 | 40 | 35 | 43.5 |
| **Models with Physiology data** | **Fold1** | **Fold2** | **Fold3** | **Fold4** | **Fold5** | **Fold6** | **Fold7** | **Fold8** | **Fold9** | **Fold10** | **Avg Acc** |
| (11) ECG_raw | 75 | **75** | **100** | 50 | **75** | 50 | 75 | **100** | 75 | **100** | 77.5 |
| (12) PANAS + ECG_raw | **100** | **75** | **100** | **100** | **75** | **100** | **100** | **100** | **100** | 75 | **92.5** |
| (13) SDHS + ECG_raw | 50 | 25 | 50 | 75 | 50 | 25 | 50 | 50 | 50 | 50 | 47.5 |
| (14) PANAS + SDHS + ECG_raw | 50 | 62.5 | 37.5 | 50 | 37.5 | 50 | 37.5 | 25 | 37.5 | 50 | 43.8 |
| (15) All questionnaires + ECG_raw | 45 | 40 | 40 | 45 | 50 | 45 | 55 | 35 | 40 | 30 | 42.5 |
| (16) PANAS + ECG_processed | 50 | **75** | 75 | 75 | 50 | 75 | 75 | 50 | 50 | 75 | 65 |
| (17) PANAS+SDHS+ECG_processed | - | - | - | - | - | - | - | - | - | - | underfitting |

## V. RESULTS

In this work, we trained different models for predicting student mental health status from physiological and subjective measures in robot assisted journal writing context over two consecutive sessions. The accuracy obtained by each model is provided in Table I and Table II for each run. As seen from Table I, our proposed approach to learn a graph representation from PANAS questionnaire compared to baseline models performs better. Likewise, our approach from ECG_raw compared to baseline models performs better (see Table I). As the results from GNN perform best, for the subsequent analyses we employed GNN. Accordingly, when we applied our approach using other questionnaires and physiological measures, as seen from Table II, the model (1) which is developed just using PANAS and the model (12) PANAS + ECG_raw data provide the highest accuracy compared to the models which incorporate other questionnaire data. In other words, different questionnaires cannot ensure the improvement of performance in the prediction task. In fact, the result after combining different questionnaires lead to the decrease of performance. Regarding incorporating raw and processed physiological measures into the model development process, we see that using ECG data in a raw format results in better accuracy compared to processed format.

## VI. DISCUSSION

Results suggest that it is possible to assess the mental wellbeing states of students from the robot administered questionnaires and their physiological reactions in journal writing context. These results reflect those of [73] who also found that robotised evaluation from questionnaire responses can be a promising tool to potentially evaluate mental wellbeing related concerns but their focus was in children. Experimental results show that the proposed graph representation learning approach is superior to standard ML methods that do not take advantage of graphical structures. Although these results are aligned with the advantage of using graph representation learning [33], it is the first time that such models with multi-dimensional edge feature were developed to predict student mental health status from physiological and subjective measures in robot assisted journal writing context.

### A. Why SDHS is not as informative as PANAS?

When we analysed the questionnaires, and each questionnaire's contribution to the prediction task, we see that only PANAS significantly predicted the WEMWBS scores (see Table 2). This was partially aligned with our expectation as PANAS is a measure that is shown to correlate with WEMWBS [25]. On the other hand, although SDHS is a measure that is shown to correlate with WEMWBS too, we found that SDHS was not contributing to prediction tasks as expected when it was evaluated alone or combined with other measures. This inconsistency may be related to the design of the experiment. Students had to respond to SDHS always after PANAS and if the questions in SDHS are considered repetitive to PANAS, answers to SDHS might have been answered in a fast fashion just to complete the task and might not provide detailed information as PANAS. Asking questions about the same topic one after the other can lead to reflex answers given without thought [74]. It is also important to note that in the study where SDHS was a measure that is shown to correlate with WEMWBS, the SDHS and WEMWBS were both recorded in an interval of one week.

On the contrary, in our experimental design we have SDHS measured in an interval of two weeks and we collect SDHS once per each week including week1 and week2 SDHS scores together in our model formulation. It might be possible that the SDHS score in week1 not contributing to the prediction task as the score in week2 and might have an impeding effect on the model performance. It is suggested that self-reporting should take place as close as possible to the experience or event which is studied [75]. Not seeing other questionnaires (BMIS, SCM, ROSAS) contributing to the prediction task as much as PANAS, might be related to the fact that these questionnaires are not usually being used for wellbeing assessment although they might reflect momentarily the improvement in wellbeing state (BMIS) and evaluation of the possible benefit of an intervention to improve wellbeing (SCM, ROSAS). What is curious about the results from ROSAS (18-item) and BMIS (16-item) is that although they do not contribute to the prediction task significantly, the performance of these questionnaires are relatively high compared to questionnaires with few number of items (SCM (4-item) and SDSH (6-item)). This finding might be related to the fact that GNNs are not able to distinguish simple graph structures [76].

### B. Is graph good for ECG representation?

Regarding analysing ECG alone and with other questionnaires, and its contribution to the prediction task, first of all our results indicate the possible contribution of physiological reactions from questionnaire responses (See Table 2 ECG_raw model (11)) in the evaluation of mental wellbeing state. This finding is consistent with that of [25] which pointed out that greater emotional expressivity can be associated with greater physiological responses. Moreover, fused models including ECG_raw performed better than models including ECG_processed in this prediction task, as it is possible to expect that raw data containing more information than filtered data. Surprisingly, when ECG_raw fused with the best performing model PANAS (See Table II model (12)), we do not see an improvement in the model performance. A possible explanation for this might be related to the fact that GNNs are not robust to noise and missing information in the data [76]. Although we found that our raw ECG signal does not have those common 50-60 hz and low-frequency noises, it might still include different types of noises that could affect the performance of GNNs in the fusion task. Moreover, answering duration of the PANAS questionnaire differed between each participant and in each session naturally affecting the size of the ECG recordings for each case. This might be another reason for the information presented within some of the ECG signals not being rich enough to contribute to the prediction task accordingly. Several studies compare manually extracted short-term Heart Rate Variability measurements to investigate their reliability to reflect the reactions during cognitively demanding tasks and it has been shown that the 30-second window is mostly presented as the smallest time frame considered to have reliable measurements [76].

## VII. Limitations & Future Work

While our study findings hold potential, it is important to highlight several constraints. First, although SDHS is a measure that is shown to correlate with WEMWBS, we found that SDHS was not contributing to prediction tasks as expected when it was evaluated alone or combined with other measures. To investigate this further, administration of PANAS and SDHS should also be done in a randomised manner to control for possible carryover effects. Second, in addition to self-reports and ECG, analysis from other measures such as journal entries (content analysis), Electrodermal Activity (EDA) and/or visual cues can also be considered for a more holistic understanding of mental health status. Finally, our dataset was imbalanced. It was imbalanced in two ways: 1) Data from one Q block and one from Q+J block were excluded from the analysis due to missing WEMWBS measures. 2) Most of the data belonged to the ‘’no depression” category. We aimed to address this limitation by reporting the Unweighted Average Recall (UAR) which can be reported for all experimental results as needed.

## VIII. Conclusions

This work contributes an approach to learn graph representations from multimodal data for mental health status prediction in students in robot assisted journal writing context. We discuss limitations and offer suggestions for improving our approach. The findings indicate the feasibility of evaluating students’ mental wellbeing states through their robot-assisted tasks in journal writing contexts. Our methodology proves more effective than conventional machine learning techniques that do not leverage graphical structures. The gathered evidence can be used to design new virtual agent interactions that assist journal writing for assessing and improving student mental well-being in classroom and counselling settings.

## Ethical Impact Statement

The study protocol underwent review and approval by the University Ethics Committee. All participant data utilized in our experimental trials adheres to the protocols outlined by relevant X and X laws and ethical standards, as guided by university resources. Each participant’s ID-related information and other data types (such as psychophysiological and behavioral data) are linked to a unique alphanumeric ID. To safeguard participant privacy, the alphanumeric ID file associated with participants' IDs is securely stored separately from other documentation on a university-controlled, password-protected disk. Access to this information is restricted to the researchers involved in the project, overseen by the principal investigator.

## REFERENCES


[1] Okanagan Charter. (2015, June 22-25). An international charter for health promoting universities and colleges[Paper presentation]. 2015 International Conference on Health Promoting Universities and Colleges/VII International CongressKelowna, BC, Canada. https://open.library.ubc.ca/cIRcle/collections/53926/items/1.0132754

[2] Van Agteren, J., Woodyatt, L., Iasiello, M., Rayner, J., & Kyrios, M. (2019). Make it measurable: Assessing psychological distress, wellbeing and resilience at scale in higher education. Student Success, 10(3), 1-13.

[3] Orygen. (2017). Under the radar: The mental health of Australian university students. https://www.orygen.org.au/Policy/Policy-Reports/Under-the-radar

[4] Carter, M. A., Pagliano, P., Francis, A., & Thorne, M. (2017). Australian university students and mental health: Viewpoints from the literature. International Journal of Innovation, Creativity and Change, 3, 1-25.https://www.ijicc.net/images/vol3iss3/1CarterPaglianoFrancisThorne.pdf

[5] Travia, R. M., Larcus, J. G., Thibodeau, K. R., Hutchinson, C. R., Wall, A., & Brocato, N. MEASURING WELL-BEING IN A COLLEGE CAMPUS SETTING.

[6] Hewitt, R. (2019). Measuring Well-Being in Higher Education. HEPI Policy Note 13. Higher Education Policy Institute.

[7] Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., ... & Kessler, R. C. (2018). WHO world mental health surveys international college student project: prevalence and distribution of mental disorders. Journal of abnormal psychology, 127(7), 623.

[8] Campbell, F., Blank, L., Cantrell, A., Baxter, S., Blackmore, C., Dixon, J., & Goyder, E. (2022). Factors that influence mental health of university and college students in the UK: a systematic review. BMC public health, 22(1), 1-22.

[9] Paton, L. W., Tiffin, P. A., Barkham, M., Bewick, B. M., Broglia, E., Edwards, L., ... & Heron, P. (2023). Mental Health Trajectories in University Students Across the Covid-19 Pandemic: Findings from the Student Wellbeing At Northern England Universities Prospective Cohort Study. Frontiers in public health, 11, 1188690.

[10] Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response.

[11] Lucas, G. M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J., & Morency, L. P. (2017). Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. Frontiers in Robotics and AI, 4, 51.

[12] Haertl, K. (2008). Journaling as an assessment tool in mental health. Assessments in occupational therapy mental health: An integrative approach (second edition). USA: SLACK Incorporated, 61-79.

[13] Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). Psychological Science in the Public Interest, 18(2), 72-145.

[14] Kawasaki, M., Yamashita, N., Lee, Y., & Nohara, K. (2020). Assessing users' mental status from their journaling behavior through chatbots. Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents. https://doi.org/10.1145/3383652.3423870

[15] DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., ... & Morency, L. P. (2014, May). SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems (pp. 1061-1068).

[16] Callejas, Z., & Griol, D. (2021). Conversational agents for mental health and wellbeing. Dialog systems: a perspective from language, logic and computation, 219-244.

[17] Denecke, K., Vaaheesan, S., & Arulnathan, A. (2020). A mental health chatbot for regulating emotions (SERMO)-concept and usability test. IEEE Transactions on Emerging Topics in Computing, 9(3), 1170-1182.

[18] Jeong, S., Aymerich-Franch, L., Arias, K., Alghowinem, S., Lapedriza, A., Picard, R., ... & Breazeal, C. (2023). Deploying a robotic positive psychology coach to improve college students' psychological well-being. User Modeling and User-Adapted Interaction, 33(2), 571-615.

[19] Laban, G., Ben-Zion, Z., & Cross, E. S. (2022). Social robots for supporting Post-Traumatic Stress Disorder diagnosis and treatment. Frontiers in psychiatry, 12, 752874.

[20] Patel, F., Thakore, R., Nandwani, I., & Bharti, S. K. (2019, December). Combating depression in students using an intelligent chatBot: a cognitive behavioral therapy. In 2019 IEEE 16th India Council International Conference (INDICON) (pp. 1-4). IEEE

[21] Jeong, S., Alghowinem, S., Aymerich-Franch, L., Arias, K., Lapedriza, A., Picard, R., ... & Breazeal, C. (2020, August). A robotic positive psychology coach to improve college students' wellbeing. In 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 187-194). IEEE.

[22] Bhakta, R., Savin-Baden, M., & Tombs, G. (2014, June). Sharing secrets with robots?. In EdMedia+ Innovate Learning (pp. 2295-2301). Association for the Advancement of Computing in Education (AACE).

[23] Abbasi, N. I., Spitale, M., Anderson, J., Ford, T., Jones, P. B., & Gunes, H. (2022, August). Can robots help in the evaluation of mental wellbeing in children? an empirical study. In 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (pp. 1459-1466). IEEE.

[24] Mendolia, M., & Kleck, R. E. (1993). Effects of talking about a stressful event on arousal: Does what we talk about make a difference?. Journal of personality and social psychology, 64(2), 283.

[25] Butler, E. A., Lee, T. L., & Gross, J. J. (2009). Does expressing your emotions raise or lower your blood pressure? The answer depends on cultural context. Journal of Cross-Cultural Psychology, 40(3), 510-517.

[26] Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: an overview. Current psychiatry reports, 21, 1-18.

[27] Joshi, M. L., & Kanoongo, N. (2022). Depression detection using emotional artificial intelligence and machine learning: a closer review. Materials Today: Proceedings, 58, 217-226.

[28] Mohamad Al Jazaery and Guodong Guo. Video-based depression level analysis by encoding deep spatiotemporal features. IEEE Transactions on Affective Computing, 12(1):262–268, 2018.

[29] Xingchen Ma, Hongyu Yang, Qiang Chen, Di Huang, and Yunhong Wang. Depaudionet: An efficient deep model for audio based depression classification. In Proceedings of the 6th international workshop on audio/visual emotion challenge, pages 35–42, 2016

[30] Nicholas Cummins, Jyoti Joshi, Abhinav Dhall, Vidhyasaharan Sethu, Roland Goecke, and Julien Epps. Diagnosis of depression by behavioural signals: a multimodal approach. In Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge, pages 11–20, 2013.

[31] Le Yang, Dongmei Jiang, Xiaohan Xia, Ercheng Pei, Meshia C´edric Oveneke, and Hichem Sahli. Multimodal measurement of depression using deep learning models. In Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge, pages 53–59, 2017.

[32] Tao Gui, Liang Zhu, Qi Zhang, Minlong Peng, Xu Zhou, Keyu Ding, and Zhigang Chen. Cooperative multimodal approach to depression detection in twitter. In Proceedings of the AAAI conference on artificial intelligence, volume 33, pages 110–117, 2019.

[33] Guimin Dong, Mingyue Tang, Lihua Cai, Laura E. Barnes, and Mehdi Boukhechba. Semi-supervised graph instance transformer for mental health inference. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 1221–1228, 2021.

[34] Siyang Song, Yuxin Song, Cheng Luo, Zhiyuan Song, Selim Kuzucu, Xi Jia, Zhijiang Guo, Weicheng Xie, Linlin Shen, and Hatice Gunes. Gratis: Deep learning graph representation with task-specific topology and multi-dimensional edge features. arXiv preprint arXiv:2211.12482, 2022.

[35] Cheng Luo, Siyang Song, Weicheng Xie, Linlin Shen, and Hatice Gunes. Learning multi-dimensional edge feature-based au relation graph for facial action unit recognition. arXiv preprint arXiv:2205.01782, 2022.

[36] Qiao Yu, Rui Wang, Jia Liu, Long Hu, Min Chen, and Zhongchun Liu. Gnnbased depression recognition using spatio-temporal information: A fnirs study. IEEE Journal of Biomedical and Health Informatics, 26(10):4925– 4935, 2022

[37] Yuzhong Chen, Jiadong Yan, Mingxin Jiang, Tuo Zhang, Zhongbo Zhao, Weihua Zhao, Jian Zheng, Dezhong Yao, Rong Zhang, Keith M. Kendrick, and Xi Jiang. Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification. IEEE Transactions on Neural Networks and Learning Systems, pages 1–12, 2022.

[38] Xinzhou Xu, Jun Deng, Nicholas Cummins, Zixing Zhang, Li Zhao, and Bj¨orn W. Schuller. Exploring zero-shot emotion recognition in speech using semantic-embedding prototypes. IEEE Transactions on Multimedia, 24:2752–2765, 2022

[39] Guimin Dong, Mingyue Tang, Lihua Cai, Laura E. Barnes, and Mehdi Boukhechba. Semi-supervised graph instance transformer for mental health inference. In 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 1221–1228, 2021

[40] Yanqiao Zhu, Hejie Cui, Lifang He, Lichao Sun, and Carl Yang. Joint embedding of structural and functional brain networks with graph neural networks for mental illness diagnosis. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), pages 272–276, 2022.

[41] Tao Chen, Yanrong Guo, Shijie Hao, and Richang Hong. Exploring selfattention graph pooling with eeg-based topological structure and soft label for depression detection. IEEE Transactions on Affective Computing, 13(4):2106–2118, 2022

[42] Tao Chen, Richang Hong, Yanrong Guo, Shijie Hao, and Bin Hu. Ms²-gnn: Exploring gnn-based multimodal fusion network for depression detection. IEEE Transactions on Cybernetics, 53(12):7749–7759, 2023.

[43] Aldebaran. 2023. Pepper the humanoid and programmable robot. http://www.aldebaran.com/en/pepper

[44] Sayis, B., & Gunes, H. (2024, March). Technology-assisted Journal Writing for Improving Student Mental Wellbeing: Humanoid Robot vs. Voice Assistant. In Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (pp. 945-949).

[45] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology 54, 6 (1988), 1063.

[46] Stephen Joseph, P Alex Linley, Jake Harwood, Christopher Alan Lewis, and Patrick McCollam. 2004. Rapid assessment of well-being: The short depressionhappiness scale (SDHS). Psychology and psychotherapy: Theory, research and practice 77, 4 (2004), 463–478.

[47] Yvo MC Meevissen, Madelon L Peters, and Hugo JEM Alberts. 2011. Become more optimistic by imagining a best possible self: Effects of a two week intervention. Journal of behavior therapy and experimental psychiatry 42, 3 (2011), 371–378.

[48] B. Sayis, R. Ramirez, and N. Pares, 'Mixed reality or LEGO game play? Fostering social interaction in children with Autism', Virtual Reality, vol. 26, no. 2, pp. 771–787, 2022.

[49] Mayer, J. D., & Gaschke, Y. N. (1988). The Brief Mood Introspection Scale (BMIS).

[50] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. 2017. The robotic social attributes scale (RoSAS) development and validation. In Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction. 254–262.

[51] Sheldon, K. M., & Lyubomirsky, S. (2006). How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves. The journal of positive psychology, 1(2), 73-82.

[52] Taggart, F., & Stewart-Brown, S. (2019). A Review of questionnaires designed to measure mental wellbeing. Retrieved on, 25.

[53] Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health and Quality of life Outcomes, 5(1), 1-13.

[54] Collect, score, analyse and interpret WEMWBS. (2024). Welcome to the University of Warwick. https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/howto/

[55] He, L., Niu, M., Tiwari, P., Marttinen, P., Su, R., Jiang, J., ... & Dang, W. (2022). Deep learning for depression recognition with audiovisual cues: A review. Information Fusion, 80, 56-86.

[56] Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. CNS neuroscience & therapeutics, 24(11), 1037-1052.

[57] Wasimuddin, M., Elleithy, K., Abuzneid, A. S., Faezipour, M., & Abuzaghleh, O. (2020). Stages-based ECG signal analysis from traditional signal processing to machine learning approaches: A survey. IEEE Access, 8, 177782-177803.

[58] Oh, S. L., Ng, E. Y., San Tan, R., & Acharya, U. R. (2018). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in biology and medicine, 102, 278-287.

[59] Chung, J., & Teo, J. (2022). Mental health prediction using machine learning: taxonomy, applications, and challenges. Applied Computational Intelligence and Soft Computing, 2022, 1-19.

[60] Song, Siyang, Shashank Jaiswal, Linlin Shen, and Michel Valstar. "Spectral representation of behaviour primitives for depression analysis." IEEE Transactions on Affective Computing 13, no. 2 (2020): 829-844.

[61] Song, Siyang, Yiming Luo, Tugba Tumer, Michel Valstar, and Hatice Gunes. "Loss Relaxation Strategy for Noisy Facial Video-based Automatic Depression Recognition." ACM Transactions on Computing for Healthcare (2024).

[62] Crowley, R. (2022). Graph Representation Learning for Child Mental Health Prediction [Master dissertation, University of Cambridge]. University of Cambridge repository. https://www.mlmi.eng.cam.ac.uk/files/2021-2022_dissertations

[63] Jivnani, L., Goodman, F., Rottenberg, J., & Canavan, S. (2023, September). Predicting Loneliness from Subject Self-Report. In 2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII) (pp. 1-8). IEEE.

[64] Xu, Jiaqi, Siyang Song, Keerthy Kusumam, Hatice Gunes, and Michel Valstar. "Two-stage temporal modelling framework for video-based depression recognition using graph representation." arXiv preprint arXiv:2111.15266 (2021).

[65] Abbasi, Nida Itrat, Siyang Song, and Hatice Gunes. "Statistical, spectral and graph representations for video-based facial expression recognition in children." In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1725-1729. IEEE, 2022.

[66] Lin, E., Lin, C.-H., and Lane, H.-Y. (2020). Precision psychiatry applications with pharmacogenomics: Artificial intelligence and machine learning approaches. International journal of molecular sciences, 21(3):969.

[67] Le Glaz, A., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., Devylder, J., Walter, M., Berrouiguet, S., et al. (2021). Machine learning and natural language processing in mental health: Systematic review. Journal of Medical Internet Research, 23(5):e15708

[68] Balan, O., Moldoveanu, A., and Leordeanu, M. (2021). A machine learning approach to ˘ automatic phobia therapy with virtual reality. In Modern Approaches to Augmentation of Brain Function, pages 607–636. Springer.

[69] Yang, P.-J. and Fu, W.-T. (2016). Mindbot: a social-based medical virtual assistant. In 2016 IEEE International Conference on Healthcare Informatics (ICHI), pages 319–319. IEEE

[70] Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior, 37:94– 100

[71] Chung, J. and Teo, J. (2022). Mental health prediction using machine learning: Taxonomy, applications, and challenges. Applied Computational Intelligence and Soft Computing, 2022.

[72] Moriarty, A. S., Meader, N., Snell, K. I., Riley, R. D., Paton, L. W., Chew-Graham, C. A., Gilbody, S., Churchill, R., Phillips, R. S., Ali, S., et al. (2021). Prognostic models for predicting relapse or recurrence of major depressive disorder in adults. Cochrane Database of Systematic Reviews, (5).

[73] Abbasi, N. I., Spitale, M., Anderson, J., Ford, T., Jones, P. B., & Gunes, H. (2022, August). Can robots help in the evaluation of mental wellbeing in children? an empirical study. In 2022 31st IEEE international conference on robot and human interactive communication (RO-MAN) (pp. 1459-1466). IEEE.

[74] Johnson, C. (2023, October 20). Response bias: Definition, 6 types, examples & more (Updated). Nextiva Blog. https://www.nextiva.com/blog/response-bias.html

[75] Can we trust self-reported data? (2022, February 24). Research on the go. https://www.researchonthego.eu/post/can-we-trust-self-reported-data

[76] W&B. (2024, March 19). Weights & biases. https://wandb.ai/syllogismos/machine-learning-with-graphs/reports/18-Limitations-of-Graph-Neural-Networks–VmlldzozODUxMzQ

[77] Nessiem, M. A., Mohamed, M. M., Coppock, H., Gaskell, A., & Schuller, B. W. (2021, June). Detecting COVID-19 from breathing and coughing sounds using deep neural networks. In 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS) (pp. 183-188). IEEE.