Predicting Individuals Mental Health Status in Kenya using Machine Learning Methods

Yara E. Alharahsheh, and Malak A. Abdullah
Computer Science Department
Jordan University of Science and Technology
Irbid, Jordan
yealharahsheh18@cit.just.edu.jo mabdullah@just.edu.jo

2021 12th International Conference on Information and Communication Systems (ICICS) | 978-1-6654-3351-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICICS52457.2021.9464608

***Abstract*—Mental Health diseases affect prominent individuals worldwide. According to WHO, 264 million people globally are affected by one mental health disease, depression. The lack of resources about the disease causes the difficulty of diagnosis and producing an efficient treatment, which eventually increases the number of cases. Depression affects several countries with a lack of knowledge about the disease and lack of resources, such as psychiatrists, psychiatric nurses, mental psychologists. In Kenya, almost 50% of its population suffers from many depression cases. This paper aims to find a robust reliable supervised Machine Learning classifier that gives the best performance evaluation for predicting if an individual is likely suffering from depression or not. The study is based on a data survey made by Busara Center in Kenya. We evaluate different machine learning methods, SVM, Random Forest, Ada Boosting, and Voting-Ensemble models scored the highest f1-score and accuracy with 0.78 and 85%, respectively.**

***Index Terms*—Machine Learning, Support Vector Machine (SVM), Naïve Bayes (NB)), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Ensemble, Gradient Boosting (GB), Ada Boosting (ada), Bagging (BG), XGBosst, Stack, Voting.**

I. Introduction

According to the World Health Organization (WHO), mental health disorders can include ”depression, bipolar disorder, schizophrenia, and other psychoses, dementia, and developmental disorders including autism” [1]. In this paper, we focus on predicting the depression of individuals in Kenya. A definition by WHO, depression can be ”sadness, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, tiredness, and poor concentration”. It also noted that depressed people might also have several physical conditions without any apparent physical cause [1]. It is considered one of the most common mental disorders and one of the leading causes of disability worldwide. WHO estimated 264 million people are affected by depression globally. It is challenging to address such disease in countries with a lack of mental health resources, such as Kenya [2].

Busara Center noted that ”1.3 million Kenyans were estimated to suffer from untreated major depressive disorder every year, yet the mental health treatment in Kenya suffers from a lack of resources” [3]. According to [4], anxiety and depression symptoms are common worldwide, especially in youth. Almost half of the Kenyan population is aged 19 or younger. It is more necessary to have more knowledge correlates of both anxiety and depression in such a country.

According to Dr. Kamau Kanyoro from the University of Nairobi, Kenya has 88 psychiatrists, 427 psychiatrist nurses, and few facilities outside of urban areas [5]. People are unlikely to know about or access them. Sadly, it was caused because most people in Kenya associate mental health and mental illness with negative narratives leading to a soft focus on the importance and benefits of mental health awareness. Mental health awareness is needed in such regions. It helps for having an early interference, which hopefully can result in a fast recovery. Untreated mental illness can lead to increased medical expenses and affect individual performance at school or work. It was also noticed it was the cause of increasing the risk of suicide [5].

As mentioned before, this study focuses on predicting if an individual is likely suffering from depression or not using machine learning methods. It was approved in [6] that analyzing individual behavior using machine learning may help improve diagnostics and have an effective treatment. The data is provided by local clinics and NGOs or community health volunteers. In this study, we have compared various Machine Learning models to predict depression suffering for an individual. The best result was achieved by the Machine Learning model based on the Voting-Ensemble technique. Furthermore, we analyzed data regarding the age of the participants and years of education; affect on.

This paper is organized into five sections. Section 2 presents the related works. Section 3 gives an overview of the dataset, features engineering, and details the experiment setup. While section 4 discusses our results. Finally, section 5 summarizes our conclusion.

II. Related Work

Due to the ease of use of social media, where users can use the different platforms to express their emotions, many studies gathered their dataset for detecting depression and anxiety from social media platforms, such as Twitter. The authors in [7] analyzed tweets to detect depression and categorized its level and severity. In another work, Reece, A.G, et al. [8] developed a computational model to identify depression and Post-Traumatic Stress Disorder within Twitter users. The researchers extracted predictive features, such as linguistic style and context, from tweets. They stated that the supervised learning algorithms successfully discriminated between depressed and healthy individuals based on the content of their tweets. Another study [9] analyzed social media users' feelings and sentiment to detect depression among the users using machine learning methods. The classifiers used are DT, KNN, SVM, and Ensemble classifiers. DT proved to give the best performance with a 73% F-measure using all the features. Also, Fatima, I., et al. [10] focused on exploring how the use of the information available on social media can help to predict postpartum depression among the users. The researchers used multilayer perceptron that outperformed SVM and Logistic Regression in their experiment.

Other studies used both machine and deep learning techniques. In Du, J. et al. [11], they presented this study trying to find methods to help identify suicide-related psychiatric stressors from streaming Twitter's data. The researchers used deep learning and transfer learning strategies, such as CNN, SVM, Extra Trees, Radom Forest, Logistic Regression, and Bi-LSTM models, to find the best techniques for suicide-related tweets. CNN outperformed all the models by recognizing tweets related to suicidal tendencies. Also, in Tyshchenko, Y. [12], they used SVM and CNN to discriminate between ill and healthy individuals. They gathered their dataset from blog posts used to identify depressed peoples and provide the necessary treatment and support.

On the other hand, Al Hanai, T. et al. [13] proposed an audio-text approach to detect depression automatically. They performed interviews between the individual and virtual agent, and through a sequence of questions and answers, the model is supposed to learn without the need to perform based on an explicit topic. LSTM neural network models were used to detect depression based on the answers of the individuals.

Sau, A. et al. [14] developed an appropriate predictive model based on machine learning algorithms to help in diagnosing depressions and anxiety among elderly patients based on different characteristics such as age, literacy, residence, employment status, history of anxiety, and depression and others. Using the WEKA tool, ten machine learning classifiers were used in the study. It included Random tree, J48, minimal sequential optimization, Random subspace, Naïve Bayes (NB), Logistic Regression (LR), Random forest (RF), K Star, Multiple Layer Perceptron (MLP), and Bayesian Network. They also presented another study to compare different machine learning algorithms' performances for predicting anxiety and depressions among seafarers in [15], such as NB, SVM, RF, LR, and Catboost. Also, Priya, A. et al. [**?**] focused on detecting stress, anxiety, and depression using a questionnaire. The R programming language was also used to apply the machine learning algorithms; Decision Tree (DT), RF, NB, SVM, and K-NN. While Walsh, C.G., etc. al. [16] proposed a work to predict the risk of suicide attempts over time using the machine learning method. The RF model was used to predict the risk of suicide attempts.

Previous studies focused on collecting datasets from a wide range of individuals from different regions. We used the Busara Center dataset that collected the dataset via survey for individuals living in Kenya. These individuals are of different ages. The dataset has more than 70 features, including education, information about household composition, economic activity, financial flows, and health. Also, participants were asked to complete a depression screening tool. We believe community surveys can provide more profound opportunities for meaningful participation in a local area. This can improve addressing diseases, such as depression, in countries with a lack of mental health resources, such as Kenya, where 1.3 million Kenyans were estimated to suffer from untreated major depressive disorder every year.

III. Methodology

A. Dataset

The dataset used in this study is provided by the ZINDI competition https://zindi.africa/competitions/busara-mental health-prediction-challenge/data. The dataset is based on a survey study made by Busara Center in western Kenya. The data contains more than 70 features, including information about the participants, such as health, economic activity, financial flows, and household composition. It is good to note here that this data is an epidemiological measure of depression where the organizers believed it represents highly suggestive of depression. Observing the training set, it contains 1,143 instances, where after splitting 800 cases were for the training set and 343 for the validation set, while the testing set has 286 cases.

B. Features Engineering

Random Forest Classifier was used to select the most important features; 51 features remain. Figure 1 shows the most ten important features in the dataset. Figure 2 shows the correlation matrix between several features. We also dropped the least important features; the training dataset has 51 feature for 1143 records and the testing dataset has 51 feature for 286 records.

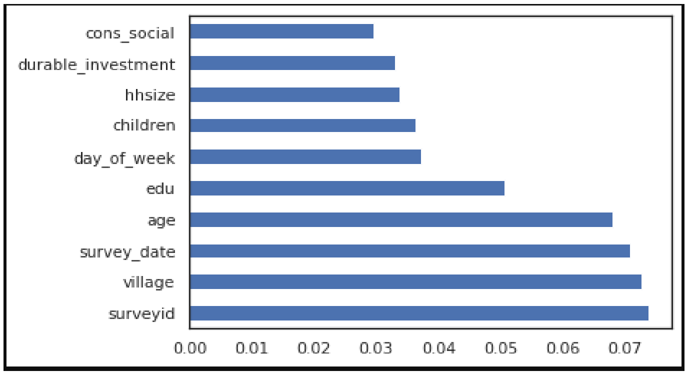

Fig. 1. Most Important Features in Dataset

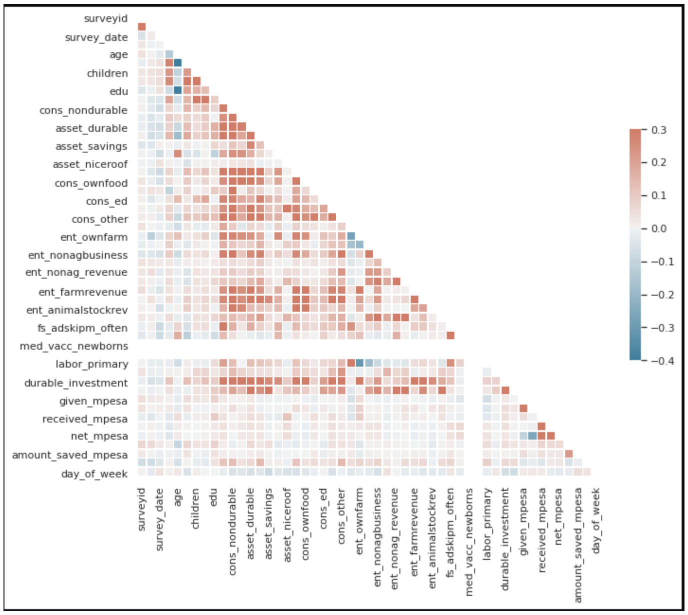

Fig. 2. Correlation Result Between Features.

TABLE I
HYPERPARAMETERS VALUES FOR EACH MODEL

Classifier	**Hyper-parameter**	**Value**
NB	-	default
SVM_1	Kernel C	Liner 5
SVM_2	Kernel C	RDF 5
SVM_3	Kernel Degree	Poly 2
LG	random_state solver max_iter	0 liblinear 1000
DT_1	criterion	Gini
DT_2	criterion max$_depth$ max_leaf_nodes min_samples_leaf min_samples_split splitter	Gini 5 10 2 2 random
DT_3	criterion max$_depth$ max_leaf_nodes min_samples_leaf splitter	entropy 6 10 5 best
RF	criterion max$_features$ max_depth n_estimators	[gini, entropy] [auto, sqrt, log2] [3,5, 10] [100, 120, 150]
GB	n_estimators max_depth max_features random_state ciretion	[10, 20, 50] [3, 5, 10] [auto, sqrt, log2] [10, 20] [friedman_mse, mse, mae]
Ada	n_estimators learning_rate random_state algorithm	[100, 120, 150] [0.1, .001, .0001] [10,20] ['SAMME', 'SAMME.R']
Voting: SVM	Kernel C-parapmeter	Rdf 3
Voting: LG	random_state solver max_iter	0 liblinear 1000
Voting: DT	-	default
Voting: RF	-	default
Voting: Ada	-	default
XGB	seed	3
Stack: GB	n_estimators max_depth random_state	90 3 8
Stack: XGB	seed	3
Stack: RF	random_state n$_estimators$	3 20

C. Experimental setup

This study aims to find the best-supervised Machine Learning (ML) classifier that gives the best accuracy for predicting if an individual is likely suffering from depression or not based on a data survey made by Busara Center in Kenya. In this study, we used various Machine Learning Methods, including the following methods. Table 1 shows each model's hyperparameters values:

- Support Vector Machine (SVM): we run three experiments for SVM with different hyperparameters, kernel, C-parameter, and degree. Naïve Bayes (NB): the hyperparameters were set to default.
- Logistic Regression (LR): the hyperparameters that were tuned were random_state, solver and maz_iter
- Decision Tree (DT):
  1) First experiment with DT, we only tuned criterion-hyperparameter and set it to Gini.
  2) Second experiment hyperparameters were criterion, max_depth, max_leaf_nodes, min_sapmles_leaf, min_samples_split, and splitter.
  3) The final experiment, we tuned the hyperparameters: criterion, max_depth, max_leaf_nodes, min_samples_leaf, and splitter.
- Random Forest (RF): we used the Grid Search to find the better hyperparameter values. The parameter to be tuned were the criterion, max_features, max_depth, and n_estimators.
- Ensemble Methods:
  1) Gradient Boosting (GB): we also used the Grid Search with GB to find the better hyperparameter values. The parameters to be tuned were n_estimators, max_depth, learning_rate, max_features, random_state, and criterion.
  2) Bagging Classifier (BG): two different classifiers were used with BG: NB and DT with n_estimators set to 20 for both classifiers.
  3) Ada Boosting (ADA): we once again used Grid Search to find the better hyperparameters values for n_estimators, learning_rate, random_stat, and algorithm.
  4) Voting-Ensembling: we used five different models; we choose Support Vector Machine, Logistic Regression, Decision Tree, Random Forest, and Gra-

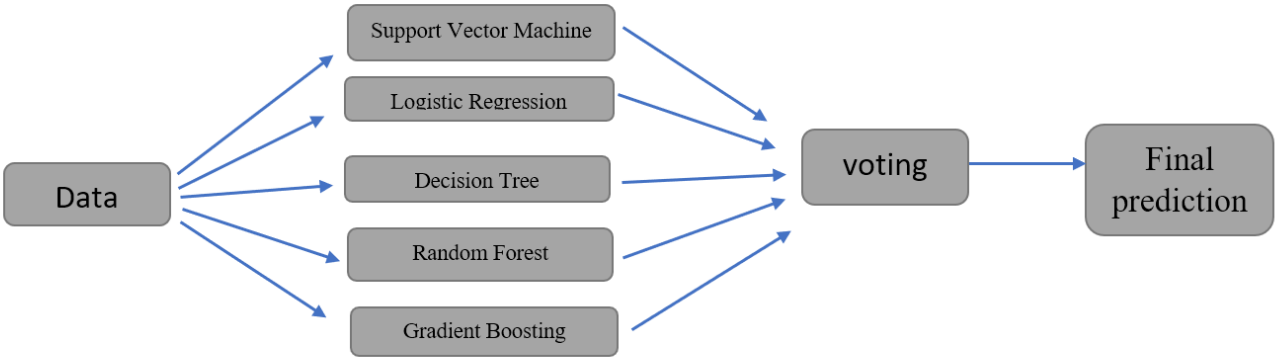

Fig. 3. Voting-Ensamble Model.

dient Boosting. Figure 3 shows the voting-ensemble model.

5) XGB classifier: we only tuned the hyperparameter seed as it was fixed to 3.
6) Stack-Ensemble: three models were used Gradient Boosting, XGBoost, Random Forests.

IV. Results and Discussion

To evaluate the model's performances on the validation dataset, we use F1-score and Accuracy metrics. Table 2 shows the F1-score and accuracy for each model on the validation dataset. SVM with RDF kernel, SNM with Polynomial kernel with degree 2, Random Forest Ada Boosting, and Voting-Ensemble models scored the same f1-score and accuracy with 0.78 and 85%. On the other hand, Naïve Byes had the worst model performance evaluation scores with f1-score 0.26 and 25% accuracy, respectively.

To evaluate the models' performances on the test dataset, the competition score was used. The error metric used will be the percentage of survey respondents to mispredict the target variable. Therefore, the closer to zero, the better the score. Table 2 shows the competition score for each model on the test dataset. The best models were SVM with RDF kernel and Polynomial kernel, Random Forest, Ada Boosting, and Voting-Ensemble model; they scored 0.1958. And the worst models were Naïve Bayes and Bagging-Ensemble model with Naïve Bayes; the score was 0.7587 and 0.7552, respectively. Comparing our best score on the test dataset to the leaderboard, the competitors who got the first three places scored: 0.1713, 01713, and 0.1783, respectively. We believe the difference is due to how the dataset's features have been handled.

Furthermore, we analyzed data in respect of the age of the participants and years of education. Studies show that both age and year of education significantly affect the chance of suffering from depression. In [17] the researchers showed that depression and educational aspirations have a significant association. They stated that individuals who were aspired to have a college degree were less likely to suffer from depression. In contrast, the individuals who only had a high school degree or less, unfortunately, have more chances to be depressed. Their result also showed that the risk of depression increases around age 40 when having low educational expectations in teens years.

TABLE II
Performance Evaluation

Model	F1-Score	Accuracy	Score
SVM_Liner	76.22%	81.04%	0.237762
SVM_RDF	78.29%	85.13%	0.19.5804
SVM$_{ploy}$	78.29%	85.13%	0.195804
NB	26.97%	25.36%	0.758741
Logistic	77.56%	83.67%	0.199301
DT_1	73.46%	74.34%	0.272727
DT_2	74.25%	81.31%	0.206293
DT_3	77.26%	82.21%	0.22028
RF	78.29%	85.13%	0.195804
GB	77.41%	83.38%	0.199301
BG_1	28.65%	26.53%	0.755245
BG_2	73.60%	81.04%	0.195804
Ada	78.29%	85.13%	0.195804
Voting	78.29%	85.13%	0.199301
XGB	74.21%	80.75%	0.671329
Stack	74.01%	79.59%	0.22028

In our experiment, the ages were in the range of 17-91. The average of the ages was 34. Figure 4 shows that most of the participants who are unfortunately suffering from depression are younger than 40 between the participants. For years of education, the longest period was 19, and the shortest was one year. Figure 5 shows the participants with a year of education less than ten years are more, unfortunately, suffering from depression. We think the result proved how crucial mental health awareness is. Diagnosis of mental health earlier can help in improving learning and school performance, also in behavioral and social adjustment as well as decreasing behavioral and emotional problems [18].

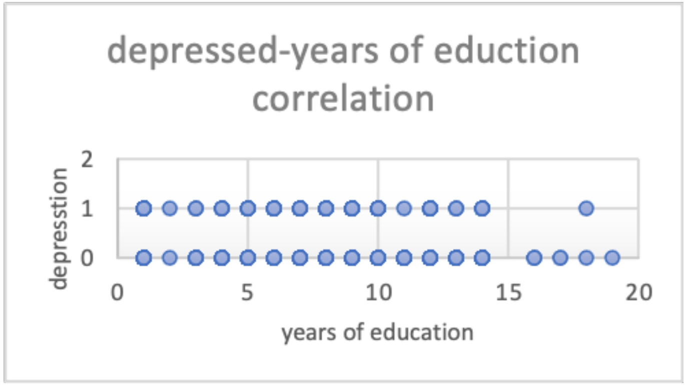

Fig. 4. Corrleation between the year of education for the participants and the chance of them to suffer from depression

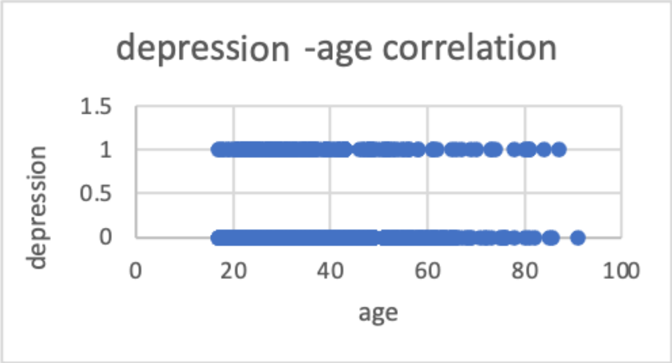

Fig. 5. Correlation between the age of the participants and the probability they suffer from depression

V. Conclusion

According to WHO, depression is considered a severe health condition, especially when someone is suffering long-lasting from it[1]. It can easily affect the affected people's daily lives, at work or school, or their social life. The importance of mental health awareness is that some depression cases, unfortunately, will commit suicide, which is considered the second cause of death between people aged 15 and 29. WHO stated that every year almost 800,000 people die due to committing suicide due to depression. Even though mental health awareness is taking a big part of people's attention, especially at young ages worldwide, people living in low and middle-income countries, such as Kenya, suffer more than those who live in high-income countries. This is due to the lack of knowledge of the disorder, the lack of treatment and resources, such as well-trained healthcare providers. In this study, we focused on predicting individual depression in Kenya, using machine learning methods based on survey data that may help improve diagnostics and have an effective treatment. Our results show that SVM, Random Forest, Ada Boosting, and Voting-Ensemble models scored the highest f1-score and accuracy with 0.78 and 85%, respectively, on the validation dataset.

References

[1] "Mental disorders," https://www.who.int/news-room/fact-sheets/detail/mental-disorders, accessed: 2020-12-22.

[2] T. L. Osborn, K. E. Venturo-Conerly, A. R. Wasil, J. L. Schleider, and J. R. Weisz, "Depression and anxiety symptoms, social support, and demographic factors among kenyan high school students," *Journal of Child and Family Studies*, vol. 29, no. 5, pp. 1432–1443, 2020.

[3] "Busara center for behavioral economics," https://www.busaracenter.org/, accessed: 2020-12-22.

[4] V. Patel and D. J. Stein, "Common mental disorders in sub-saharan africa: The triad of depression, anxiety and somatization." 2015.

[5] "The state of mental health in kenya," https://uonresearch.org/vvc/article/the-state-of-mental-health-in-kenya, accessed: 2020-12-22.

[6] T. Richter, B. Fishbain, A. Markus, G. Richter-Levin, and H. Okon-Singer, "Using machine learning-based analysis for behavioral differentiation between anxiety and depression," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[7] B. Bataineh, R. Duwairi, and M. Abdullah, "Ardep: An arabic lexicon for detecting depression," in *Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence*, 2019, pp. 146–151.

[8] A. G. Reece, A. J. Reagan, K. L. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with twitter data," *Scientific reports*, vol. 7, no. 1, pp. 1–11, 2017.

[9] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, "Depression detection from social network data using machine learning techniques," *Health information science and systems*, vol. 6, no. 1, pp. 1–12, 2018.

[10] I. Fatima, B. U. D. Abbasi, S. Khan, M. Al-Saeed, H. F. Ahmad, and R. Mumtaz, "Prediction of postpartum depression using machine learning techniques from social media text," *Expert Systems*, vol. 36, no. 4, p. e12409, 2019.

[11] J. Du, Y. Zhang, J. Luo, Y. Jia, Q. Wei, C. Tao, and H. Xu, "Extracting psychiatric stressors for suicide from social media using deep learning," *BMC medical informatics and decision making*, vol. 18, no. 2, pp. 77–87, 2018.

[12] Y. Tyshchenko, "Depression and anxiety detection from blog posts data," *Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu, Estonia*, 2018.

[13] T. Al Hanai, M. M. Ghassemi, and J. R. Glass, "Detecting depression with audio/text sequence modeling of interviews." in *Interspeech*, 2018, pp. 1716–1720.

[14] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238–243, 2017.

[15] ——, "Screening of anxiety and depression among the seafarers using machine learning technology," *Informatics in Medicine Unlocked*, vol. 16, p. 100149, 2019.

[16] C. G. Walsh, J. D. Ribeiro, and J. C. Franklin, "Predicting risk of suicide attempts over time through machine learning," *Clinical Psychological Science*, vol. 5, no. 3, pp. 457–469, 2017.

[17] A. K. Cohen, J. Nussbaum, M. L. R. Weintraub, C. R. Nichols, and I. H. Yen, "Peer reviewed: Association of adult depression with educational attainment, aspirations, and expectations," *Preventing Chronic Disease*, vol. 17, 2020.

[18] K. Van Landeghem and C. Hess, "Children's mental health: An overview and key considerations for health system stakeholders," *Children's Mental Health*, 2005.