Risk Prediction of Major Depressive Disorder using Artificial Neural Network

Fatima Hamed
School of Biomedical Engineering and Health Science, Faculty of Engineering
Universiti Teknologi Malaysia
Malaysia

Eko Supriyanto*
School of Biomedical Engineering and Health Science, Faculty of Engineering
IJN-UTM Cardiovascular Engineering Centre
Universiti Teknologi Malaysia
Malaysia
eko@utm.my

Safaa Osman
School of Biomedical Engineering and Health Science, Faculty of Engineering
Universiti Teknologi Malaysia
Malaysia

Tarig Ahmed El Khider Ali
Donegal Mental Health Services
Department of Psychiatry
Letterkenny University Hospital
Republic of Ireland

2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) | 978-1-7281-8406-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISRITI51436.2020.9315463

***Abstract*—Major Depressive Disorder (MDD) is a serious medical condition that can affect many areas of a person's daily life significantly. MDD, caused by a combination of factors, will be debilitating if not detected and managed early. This is why it is the leading cause of disability around the world. If detected early, several treatment and management programs can be done, for example, change of lifestyle. There are models developed to predict the risk of individual suffering MDD but they have low sensitivity and specificity. In this study, a new MDD risk prediction model is developed using a novel equation and Artificial Neural Network (ANN). The model is created using risk factors of MDD that are categorized into three groups, which are psychological, social and biological. Two predictor methods are applied, first, using a conventional equation, then using an ANN tool. From the results, the conventional equation is able to provide the risk estimation for MDD. After comparing, ANN showed the ability to calculate the risk prediction of MDD with 70% test accuracy and found to have a better sensitivity and specificity than the existing models.**

Keywords—risk prediction, artificial neural network, major depressive disorder.

I. Introduction

Major Depressive Disorder (MDD) is one of the most common and significant mental health conditions caused by a number of factors such as a chemical imbalance in the brain caused by adverse life events [1]. MDD is the most prevalent mental disorder causing disabilities globally, according to the World Health Organisation (WHO) [1]. An individual may be diagnosed with the condition by a physician, when they have the symptoms, as listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM), for at least two weeks [2]. People with MDD usually exhibits mood and behaviours changes, affecting their daily activities such as, working, sleeping or eating.

According to the Global Burden of Diseases, Injuries, and Risk Factors Study 2017 (GBD 2017), about 264 million people worldwide have been affected by depressive disorders [3]. It is estimated that 76-85% of people in the developing countries suffer from mental disorder, with 5% of men and 9% of women having depressive disorder [4]. One of the major effects of depression is suicide, the second leading cause of death among 15-29 year old, according to WHO [1].

As a disorder that commonly occurring in adolescence and early adulthood, and with high recurrence and persistence, early prevention and intervention may reduce the number of disabilities and suicide deaths. Effective treatment of MDD thus requires early diagnosis. One of the approaches that can be used is by using a risk predictor, which help to predict the risk of an individual suffering from MDD.

A disease risk predictor is a statistical model that uses the risk factors to create an algorithm that calculates the probability of a person suffering from the disease. In MDD, the risk factors can be categorised as demographic and psychosocial risk factors. The demographic risk factors include gender, age and socio-economic status [5][6][7][8]. Psychosocial risk factors include adverse life experiences, history of traumatic events, job stress, and financial strain [9][10][11]. Risk prediction is obtained from the combination result of key risk factors. Therefore, to accurately measure the MDD risk prediction, algorithms that include the key set of MDD risk factors are required.

This prediction algorithm method is one of the tools that has helped individuals, physicians, and policy makers to make health-related informed decisions. Many algorithms have been developed and used for disease prevention and management, for examples, cardiovascular disease (Framingham Risk Score) [12], and breast cancer risk prediction (Gail Model) [13]. Researchers have used the decision tree forest algorithm method to predict the risk for depression, where the risk factors used include use of alcohol, smoking, life events, health rating, low level of education or mastery, employment status and financial pressure, religious service attendance, age and gender [14]. Weaknesses of this model are the limitation of the decision tree method when applied to disease prevention purpose, and its low sensitivity and specificity. Other method that has been applied for a similar objective is logistic regression using identical predictors [14], which also limited with its low sensitivity and specificity.

In this study, a new MDD risk prediction model is developed. Initially, a conventional algorithm is used, before a machine learning algorithm using Artificial Neural Network (ANN) method is created. The results of both algorithms are then compared.

II. Method

A. MDD Risk Factors Classification

In this study, based on its physiology, the risk factors of MDD can be classified into three categories, psychological, social, and biological. Table I shows the risk factors included in each categories. The psychological risk factors include gender and life events. Financial pressure, employment status, age and education level are listed under social risk factors category. The biological risk factors consist of health rating, body mass index, blood pressure, alcohol intake, smoking, physical activity and diabetes [14].

TABLE I. MAJOR DEPRESSIVE DISORDER (MDD) RISK FACTOR CLASSIFICATION

MDD Risk Factors		
Psychological (PRF)	***Social (SRF)***	***Biological (BSF)***
Gender (G) Life Events (LE)	Age (A) Education Level (EL) Financial Pressure (FP) Employment Status (ES)	Body Mass Index (B) Alcohol (AL) Smoking (S) Diabetes (D) Health Rating (HR) Hypertension (H) Family History (FH) Physical Activity (PA)

Fig. 1 illustrates the relationship between these risk factors that is applied in the algorithm development.

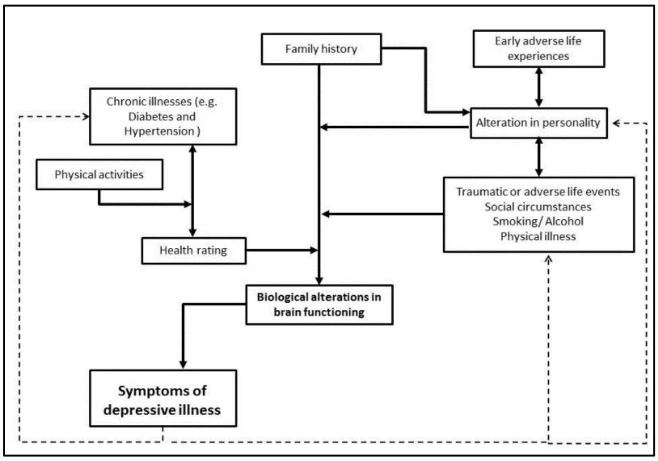

Fig. 1. MDD risk factors relationship.

B. Database from Literature

Fig. 2 shows the effect of stressful life events to MDD risk prediction. These events, which leads to trauma, will cause the person to develop the symptoms of MDD. Another factor in the psychological category, that directly affects the prognosis of depression, is gender. It has been found that more female are diagnosed with depression. In addition, women's symptoms are more readily diagnosed as depression [15][16].

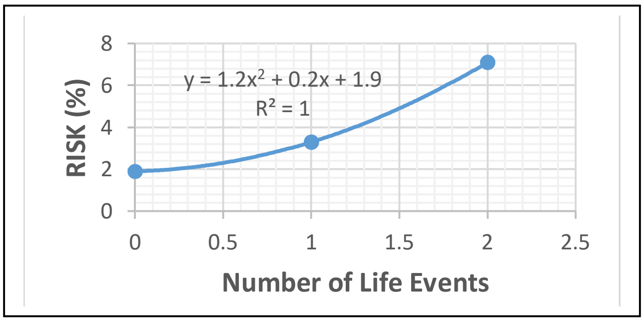

Fig. 2. Relationship between life events and risk of MDD.

The risk factors in the social category are associated to the social pressures which usually will lead to depression. There is a direct relationship between losing a job or financial issues with stress and depression, due to social pressure and loss of social status. These problems may lead to social isolation [17]. Adults are usually at risk of depression because they are more prone to these type of pressures.

Biological risk factors are associated with the individual health status. Statistics showed that people who are accustomed to exercise are less likely to be depressed because the body releases chemicals called endorphins which responsible to the feeling of happiness [18]. People who suffered from chronic diseases like diabetes, and hypertension are more likely to feel anxious due to the fearful thought and feeling of deprivation [19]. As shown in Fig. 3, the risk of depression is increased with diabetes. Similarly, high BMI and obesity can lead to depression due to low self-image and other health problems arise with the conditions.

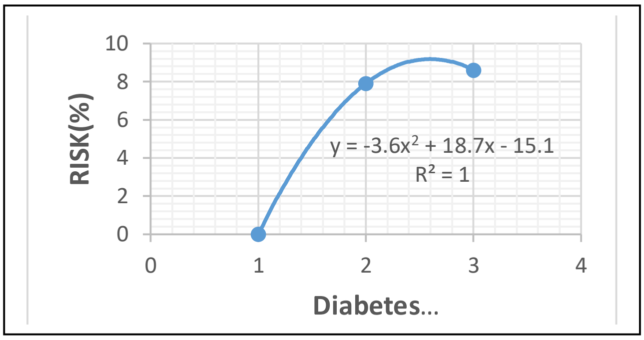

Fig. 3. Relationship between diabetes and risk of MDD; 1: No diabetes, 2: Diabetes Type 1; 3: Diabetes Type 2.

Alcohol and smoking found to be among the key risk factors of depression. Depression can lead to alcohol or nicotine addiction, or vice versa. Recent research suggests that nicotine can damaged important brain pathways responsible for mood regulation [20]. On the other hand, alcohol addiction and depression often occur together. Alcohol can affect the chemistry of the brain, which may increase the risk of depression. The after-effects of alcohol may affect the person's life (relationship, work, memory, etc.) which also can lead to depression.

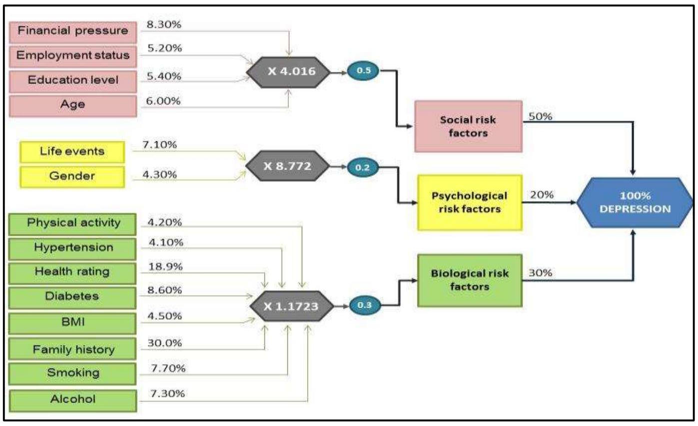

Fig. 4. MDD Conventional Equation Model.

C. Risk Prediction Models

a) Conventional Algorithm: The conventional algorithm of MDD is created based on the relationship between the risk factors and the risk of depression. Each risk factors categories (Table I) will have a total of 100% risk. Fig. 4 shows this model, and the risk percentage of each risk factors as gathered from previous literatures.

The total of the risk percentages will then be normalized to be equal to 100%. Each risk factors has been assigned with the estimated weight from the total risk, based on their effect towards MDD. The final equation is obtained by multiplying the weight of the risk factors in each categories.

MDD risk is calculated using a static risk equation (1) as follow:

$$\boldsymbol{MDD\ Risk\ Static\ Equation} = f(PRF) + f(SRF) + f(BRF) \tag{1}$$

where,

$$f(PRF) = a * ((X_{1G} + X_{2LE}) * N_1)$$

$$f(SRF) = b * ((X_{3A} + X_{4FP} + X_{5EL} + X_{6ES}) * N_2)$$

$$f(BRF) = c * ((X_{7PA} + X_{8H} + X_{9HR} + X_{10D} + X_{11B} + X_{12FH} + X_{13S} + X_{14AL}) * N_3)$$

a, b, and c are the are the weight which are estimated based on the effect of the factors in the risk prediction. N_1, N_2, and N_3 are the normalization numbers.

The conventional equation (2) to calculate the MDD risk developed from the static risk equation is as follow:

$$\boldsymbol{MDD\ Conventional\ Equation} = 0.2\, f(PSF * 8.772) + 0.5\, f(SRF * 4.016) + 0.3\, f(BRF * 1.17234) \tag{2}$$

An example of risk calculation using the conventional equation is shown in Table II. The risks are classified into three levels, according to the percentages. Percentage of 40 to 66 is low-risk, 66 to 85 is medium-risk, and 85 to 100 is high-risk.

b) Machine Learning Algorithm: In order to improve the accuracy and to have a dynamic method that improve with time, ANN method is used in this study. The ANN is trained with the conventional equation in the MATLAB software, using the fed forward pattern recognition method. The data input consists of the MDD features and risk factors, to build a network that able to produce a risk percentage. The ANN in this study is designed with 14 input neurons , 10 hidden layers nurons, and 3 output layers neurons. As shown in Fig. 5, the ANN involved 2 layers, which are the hidden and output layers.

The training function used to update the weights and bias values is based on the scaled conjugate gradient method. The performance cross-entropy method is used for the calculation. The targets are divided into three sets by using random indices.

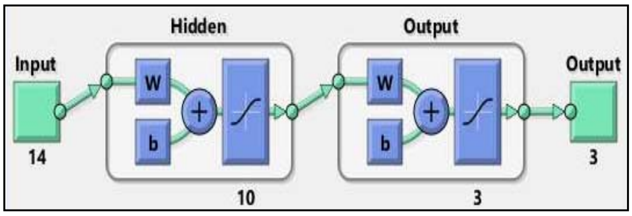

Fig. 5. Artificial Neural Network (ANN).

TABLE II. MDD RISK CALCULATION EXAMPLE USING CONVENTIONAL METHOD.

MDD Risk Calculator			
Risk Factors	***Value***	***Risk Value***	***Notes***
Age	24	6	Age group (20-24), (40-44), (60-64)
Gender	Female	4.3	1 for male; 2 for female
Financial Pressure	Yes	8.3	Have financial pressure?
Employment Status	Unemploy -ed	5.2	What is your employment status?
Life Events	5	7.1	Any adverse life events?
Family History	No	10	Have family history with MDD?
Physical Activity	No	4.2	Any regular moderate physical activity?
Hypertension	Yes	3.9	Diagnosed with hypertension?
Health Rating	Poor	18.9	What is your health rating?
Diabetes	Type 2	8.6	Diagnosed with Diabetes?
Alcohol	Yes	7.3	Do you consume alcohol?
Smoking	Yes	7.7	Are you a smoker?
Education Level	High school	5.2	What is your education level?
BMI (Overweight)	No	4.5	Are you overweight?
Result			
Risk % (Level)	92.49 (High)		

III. RESULTS AND ANALYSIS

A survey has been conducted to collect information from 40 samples to be used in the conventional equation and artificial neural network method. The generated output is classified to three risk levels based on the risk percentage calculated by the conventional equation. The ANN is trained with 20 samples generated from the conventional equation, and then tested with 10 and 40 samples. The results accuracy of ANN tested with 10 samples is 60%. The accuracy increased to 70% when tested with 40 samples. This shows that ANN has the potential to improve the error accuracy with increased number of samples. This is an important characteristic in developing a reliable risk estimation. This also shows the potential to eliminate the drawback and improve the accuracy of the conventional method by increasing the sample size. Table III shows the result of conventional equation for 40 samples. Risk level 1, 2 and 3 are referred to high-, medium- and low-risk, respectively.

ANN model is trained and tested twice, with 20 samples for training, and then with 30 samples using neural pattern recognition app in the MATLAB software. The risk values of features are used for the input dataset, and the level of the risk are used for the target dataset. After dataset training, 10 samples are used for testing to produce the confusion matrix. From the confusion matrix, precision, specificity, and recall (sensitivity) values for each set are calculated, as shown in Table IV.

TABLE III. RESULT OF CONVENTIONAL EQUATION FOR 40 SAMPLES.

No.	Value	Risk Level	No.	Value	Risk Level
1	99.30	1	21	100.00	1
2	85.28	1	22	81.87	2
3	85.64	1	23	83.67	2
4	88.55	1	24	88.55	1
5	92.49	1	25	92.49	1
6	85.84	1	26	86.65	1
7	87.85	1	27	87.85	1
8	66.85	2	28	66.85	2
9	74.48	2	29	61.05	3
10	82.94	2	30	79.91	2
11	67.65	2	31	85.83	1
12	42.78	3	32	84.26	2
13	53.12	3	33	79.21	2
14	72.60	2	34	53.54	3
15	63.83	3	35	77.81	2
16	59.16	3	36	75.48	2
17	64.05	3	37	48.40	3
18	86.04	1	38	91.77	1
19	72.27	2	39	50.11	3
20	78.80	2	40	78.56	2

TABLE IV. PRECISION, SPECIFICITY AND RECALL

No. of Samples	Class	Precision	Specificity	Recall
20	A	0.500	0.750	1.000
	B	1.000	1.000	0.200
	C	0.600	0.710	1.000
30	A	0.670	0.875	1.000
	B	1.000	0.600	0.800
	C	1.000	1.000	0.330

Fig. 6 and 7 shows the output result for ANN training dataset of 30 samples and testing dataset of 10 samples.

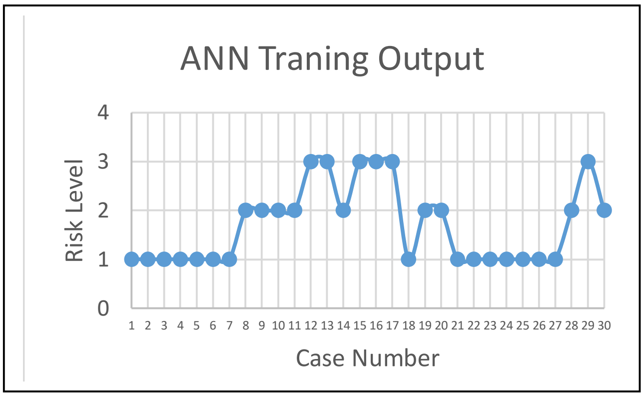

Fig. 6. ANN Output Result for Training Data Set.

The result shows that ANN error accuracy can be improved by increasing the number of samples. Fig. 8 illustrated the accuracy comparison between the two sets.

In comparison, it is found that the ANN method provides a more accurate result than the conventional method. Fig. 9 shows the results of both methods on the test set output.

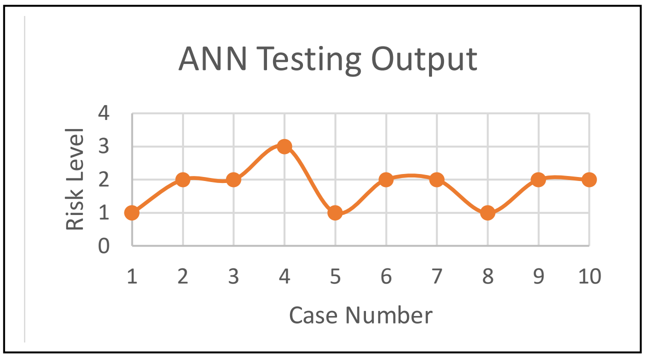

Fig. 7. ANN output result for testing data set.

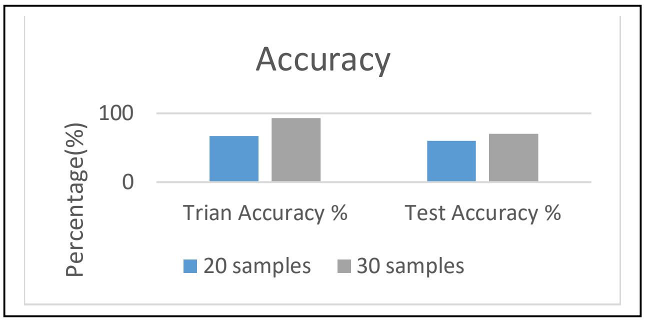

Fig. 8. Accuracy comparison.

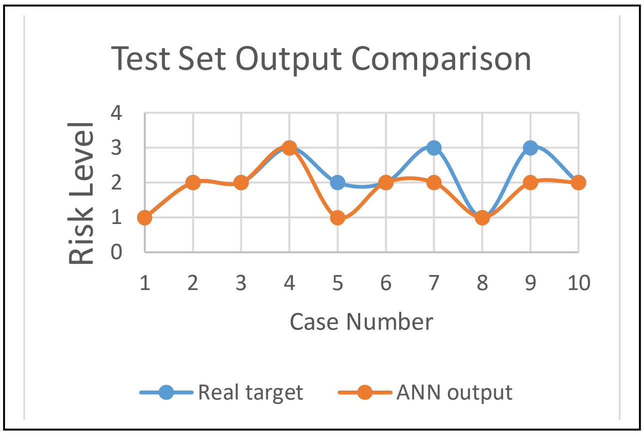

Fig. 9. Test output for both conventional and ANN output.

From the conventional results, a normal and healthy person is classified as low-risk, while an abnormal person is classified as high-risk. The conventional method can be used as an estimation tool but it is not suitable as a clinical risk prediction. The ANN is more suitable in this case, because its prediction algorithm will improved with time and increased number of samples. Therefore, ANN has a better potential to be used as MDD risk prediction method.

IV. CONCLUSION

In this study, risk prediction of Major Depressive Disorder (MDD) is obtained with 70% accuracy, and better sensitivity and specificity compared to the exiting methods. The prediction is calculated using retrospective data on two different methods, the conventional and the ANN methods. The conventional method able to produce the risk percentage, but the ANN method is more suitable and practical for clinical usage. Future studies recommendation are to include more high impact risk factors, use real clinical data, and train with a higher number of samples to ensure the accuracy and reliability of the neural network.

ACKNOWLEDGMENT

The authors are grateful to Universiti Teknologi Malaysia on the support throughout this study and for financial support through High Impact Research Grant (Vot no. 04G29).

REFERENCES

[1] "Depression," 30-Jan-2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression.

[2] American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

[3] James, S.L., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A. and Abdollahpour, I., 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet, 392(10159), pp.1789-1858.

[4] Kessler, R.C., Demler, O., Frank, R.G., Olfson, M., Pincus, H.A., Walters, E.E., Wang, P., Wells, K.B. and Zaslavsky, A.M., 2005. Prevalence and treatment of mental disorders, 1990 to 2003. New England Journal of Medicine, 352(24), pp.2515-2523.

[5] Skapinakis, P., Weich, S., Lewis, G., Singleton, N. and Araya, R., 2006. Socio-economic position and common mental disorders: Longitudinal study in the general population in the UK. The British Journal of Psychiatry, 189(2), pp.109-117.

[6] Weich, S., Sloggett, A. and Lewis, G., 1998. Social roles and gender difference in the prevalence of common mental disorders. The British Journal of Psychiatry, 173(6), pp.489-493.

[7] Weich, S. and Lewis, G., 1998. Material standard of living, social class, and the prevalence of the common mental disorders in Great Britain. Journal of Epidemiology & Community Health, 52(1), pp.8-14.

[8] Wang, J.L., Schmitz, N. and Dewa, C.S., 2010. Socioeconomic status and the risk of major depression: the Canadian National Population Health Survey. Journal of Epidemiology & Community Health, 64(5), pp.447-452.

[9] Wang, J., 2005. Work stress as a risk factor for major depressive episode (s). Psychological Medicine, 35(6), p.865.

[10] Kendler, K.S., Hettema, J.M., Butera, F., Gardner, C.O. and Prescott, C.A., 2003. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. Archives of General Psychiatry, 60(8), pp.789-796.

[11] Wang, J., Patten, S.B., Currie, S., Sareen, J. and Schmitz, N., 2012. A population-based longitudinal study on work environmental factors and the risk of major depressive disorder. American Journal of Epidemiology, 176(1), pp.52-59.

[12] Anderson, K.M., Wilson, P.W., Odell, P.M. and Kannel, W.B., 1991. An updated coronary risk profile. A statement for health professionals. Circulation, 83(1), pp.356-362.

[13] Decarli, A., Calza, S., Masala, G., Specchia, C., Palli, D. and Gail, M.H., 2006. Gail model for prediction of absolute risk of invasive breast cancer: independent evaluation in the Florence–European prospective investigation into Cancer and Nutrition Cohort. Journal of the National Cancer Institute, 98(23), pp.1686-1693.

[14] Batterham, P.J., Christensen, H. and Mackinnon, A.J., 2009. Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach. BMC Psychiatry, 9(1), p.75.

[15] D. F. Levinson and W. E. Nichols, "Major Depression and Genetics," Genetics of Brain Function. [Online]. Available: https://med.stanford.edu/depressiongenetics/mddandgenes.html.

[16] Romans, S.E. and Seeman, M.V. eds., 2006. Women's Mental Health: A Life-Cycle Approach. Lippincott Williams & Wilkins.

[17] D. Semple, R. Smyth "Oxford Handbook of Psychiatry", Oxford University Press (ESC licenced) 2019.

[18] Carless, D. and Douglas, K., 2010. Sport and Physical Activity for Mental Health. Oxford: Wiley-Blackwell.

[19] Mut-Vitcu, G., Mosteoru, S., Timar, R., Gaita, L. and Gaita, D., 2017. Discovering the link between diabetes, depression and cardiovascular disease. Romanian Journal of Cardiology| Vol, 27(3).

[20] Cohen, R.A. and Sweet, L.H. eds., 2010. Brain Imaging in Behavioral Medicine and Clinical Neuroscience. Springer Science & Business Media.