A Neural Network Algorithm-Based Prediction Model for Depression Among Elderly Individuals with Diabetes Mellitus

Xiaojuan Wang[1†], Shaofeng Wang[1†], Wencong Xu[1], Xiaoqin Zhao[1], Zhang Zhao[1], Jing Li[1]*

[1]School of Health and Medical Technology, Neusoft Institute Guangdong 528225, China

Keywords:

Geriatric Diabetes;
Depression;
Random Forest;
BP Neural Network;
CHARLS

ABSTRACT

Objective: To address the "syndemic" of depression among elderly diabetic patients in China by developing a non-invasive machine learning predictive model for large-scale community screening.

Methods: Data from 4,827 participants (aged 60–85) were extracted from the 2020 CHARLS database. Random Forest (RF), a tree-based machine learning technique, was used to select important variables. Then, a Backpropagation Neural Network (BPNN), which can capture complex, non-linear relationships, was built. Model performance was evaluated using Accuracy (correct prediction rate) and AUC (area under the curve, a measure of discrimination ability).

Results: The prevalence of depressive symptoms was 33.7%. RF identified age, physical pain distribution, and sleep duration as the top predictors. Although post-noon nap duration was not significant in a simple statistical analysis, it emerged as important in a neural network, which detects hidden patterns. The model achieved an accuracy of 0.67 (percentage of correct predictions) and AUC of 0.59 (ability to distinguish cases) on the test set, indicating strong negative predictive value for risk triaging.

Conclusion: Physical pain, sleep patterns, and social engagement are decisive warning signs for depression in elderly diabetics. This non-invasive model is an essential, scalable solution for early identification and intervention within primary healthcare systems.

CONTACT:

[1]Author: wangxiaojuan@nuit.edu.cn; wangshaofeng@nuit.edu.cn

[*]Corresponding Author: xuwencong@nuit.edu.cn

Note: Authors designated with "†" are recognised as co-first authors.

DOI: 10.64549/jaai-ii.v1i1.52

This work is licensed under the CC BY 4.0 HTTPS://CREATIVECOMMONS.ORG/LICENSES/BY/4.0/

1.Introduction

China's ageing population has led to a marked rise in diabetes among the elderly. The comorbidity of diabetes with other diseases presents a significant challenge for global diabetes health interventions and management. Mental health issues associated with diabetes, particularly depression, are frequently overlooked due to a predominant clinical focus on physical health. Within global public health research, the co-occurrence of diabetes and depression has been identified as a "syndemic." Recent epidemiological studies indicate that the prevalence of depressive symptoms among elderly diabetic patients in China ranges from 35% to 38.6%, a rate substantially higher than that observed in non-diabetic elderly populations. This prevalence is especially pronounced among women, individuals with lower education levels, and those with chronic complications [1].

The comorbidity of diabetes and depression represents a complex, bidirectional, and mutually reinforcing relationship. Laboratory studies indicate that prolonged hyperglycemia and associated metabolic disorders constitute the endocrine basis for depressive states in individuals with diabetes by inducing systemic low-grade inflammation and activating the hypothalamus-pituitary-adrenal (HPA) axis [2, 3].

Despite the high predictive value of biomarkers in individual experimental settings, large-scale community screening for routine risk assessment remains challenging due to high testing costs, invasive sampling procedures, and reliance on primary healthcare infrastructure. Existing studies indicate that depressive episodes in elderly patients with chronic diseases are associated with factors such as age, lifestyle, social support, and the accumulation of chronic conditions [4, 5, 6]. Consequently, intelligent prediction tools utilising non-invasive and easily accessible features offer greater applicability for large-scale population screening. Traditional clinical prediction models primarily employ statistical methods such as logistic regression (LR). While these models provide strong interpretability, they assume linear relationships between variables. In elderly populations, however, complex multidimensional nonlinear interactions often exist between lifestyle variables and metabolic indicators, which limit the predictive accuracy of traditional models [7].

Machine learning approaches are therefore more effective in constructing depression prediction models for elderly diabetic patients, as they can accommodate complex interactions. Recent studies have applied single algorithms, including random forests and neural networks, to develop depression prediction models for elderly patients [8, 9]. Although random forests (RF) demonstrate strong performance in feature selection and classification tasks involving chronic disease data, their reliance on a stepwise decision space limits their ability to capture smooth, nonlinear interactions among high-dimensional variables, such as lifestyle and disease status [10]. Neural network-based analyses, while advantageous in global function approximation, are susceptible to local optima and often lack interpretability when applied to structured data from lifestyle questionnaires [11]. Integrating the robust feature extraction capabilities of random forests with the deep mapping strengths of neural networks can therefore improve the accuracy of complex biopsychosocial model construction [12].

The fifth round of follow-up questionnaires from the 2020 China Health and Retirement Longitudinal Study (CHARLS) database was analysed. Data on lifestyle, social support networks, and physiological functions of elderly individuals were integrated. Random forests were applied for feature engineering optimisation to remove redundant variables, and a backpropagation neural network was used to construct a nonlinear predictive model. This research paradigm aims to develop a low-cost, efficient digital early warning tool, offering empirical support for the early identification and intervention of depressive disorders in elderly diabetic patients.

2. Data Sources and Research Methods

2.1 Data Sources

The fifth round (2020) follow-up questionnaire data from the China Health and Retirement Longitudinal Study (CHARLS) served as the primary data source. Depressive symptoms were measured using the CES-D score as the dependent variable. The CES-D comprises 10 items, each categorised by symptom frequency as "rarely or never," "not very often," "sometimes or half the time," and "most of the time," corresponding to values of 0, 1, 2, and 3, respectively. Items 5 and 8 were reverse-scored, resulting in a total possible score ranging from 0 to 30. A score of 10 or higher was classified as indicative of depressive symptoms (assigned a value of 1), while a score between 0 and 9 was considered within the normal range (assigned a value of 0).

Sixteen independent variables related to depressive symptoms were included and categorised into the following five types:

(1) Demographic variables: gender and age.

(2) Health status variables: Number of chronic diseases (hypertension, diabetes, dyslipidemia, etc., 15 types), number of pain locations (head, shoulder, arm, etc., 15 types), whether or not a hip fracture has been experienced, whether or not a traffic accident has been experienced, whether or not outpatient visits have been made in the past month, whether or not hospitalization has been made in the past year, etc., 6 types.

(3) Health behaviour variables. Frequency of alcohol consumption, sleep duration, duration of afternoon nap, regular exercise. Regular exercise was scored as follows: 0, 1, 2, and 3 points were assigned for no regular exercise, light exercise, moderate exercise, and high-intensity exercise, respectively. The highest applicable score was recorded for each respondent. Social activity scores were determined by the frequency of participation in activities such as visiting friends and family, playing mahjong, chess, or cards, providing assistance to non-cohabiting relatives or neighbours, engaging in physical activities, participating in community organisations, volunteering, attending educational courses, and other social activities. Frequency was scored as 3 points for "almost daily," 2 points for "almost weekly," and 1 point for "not often." The total social activity score was calculated by summing these values. were then added together to obtain the respondents' social activity score.

(4) Family status variables: marital status and pension payment frequency. The initial CHARLS 2020 dataset contained 19,367 respondents. A systematic data-cleaning process was used to screen for valid samples. The specific steps were as follows: First, the research subjects were limited to the elderly aged 60-85; second, samples with no response records on the depression scale and missing values in the dependent variable were removed; finally, based on clinical rationality and data distribution characteristics, outliers were strictly judged and excluded. Specific exclusion criteria included: (1) sleep time ≤ 4 hours; (2) nap duration ≥ 4 hours; (3) by fitting a curve to a scatter plot of pain location and depression score, outliers significantly deviating from the overall trend were identified and excluded. After the multi-stage screening described above, 4,827 valid samples were included for subsequent analysis.*2.2 Data Collection (Times New Roman 12pt, Italics)*

Body texts: The research method explains the implementation methods employed in the study. The method is described clearly and in detail.

2.2 Feature Screening

The importance of predictive features related to depression levels was quantified and ranked using the random forest ensemble learning algorithm. The 4,827 samples (1,627 positive and 3,200 negative) were randomly divided into training and test sets at a 1:1 ratio. First, oversampling was used to balance the training set. Second, grid search cross-validation was used to systematically optimise the key hyperparameters of the random forest to ensure the stability of the feature importance ranking. Finally, the relative importance differences between features were visualised using bar charts.

2.3 Model Construction and Performance Evaluation

For the BP neural network model, the dataset was split into training and test sets at a 7:3 ratio, and the selected important features were standardised using Z-scores. A multilayer perceptron (MLP) classifier was constructed with sklearn.neural_network.MLPClassifier. The training data were used to fit the MLP classifier. Model performance was evaluated by plotting the receiver operating characteristic (ROC) curve and calculating the area under the curve (AUC).

3. Results

3.1 General Information

A total of 4,827 samples were included. Based on depression scale scores, participants were categorised into a depressive symptoms group (n=1,627) and a non-depressive symptoms group (n=3,200). The sample comprised 68.13% females (3,288 cases) and 31.87% males (1,539 cases). The age distribution was as follows: 60–69 years (59.98%, 2,901 cases), 70–79 years (33.62%, 1,623 cases), and 80–85 years (6.40%, 303 cases). Statistically significant differences between the depressive and non-depressive groups were observed for gender, age group, marital status, frequency of alcohol consumption, history of hip fracture, number of chronic diseases, sleep duration, number of pain sites, internet use, and regular exercise ($P<0.05$). No statistically significant differences were found for history of traffic accidents, nap duration, or social frequency ($P>0.05$). (Table 1).

Table 1 Comparison of characteristics between depressed and non-depressed elderly patients with diabetes (N=4827).

Characteristic	Category	Depression (n=1627)	Non-depression (n=3200)	$\chi2$	P-value
Gender	Female	961 (59.1%)	2327 (72.7%)	91.95	< 0.001
	Male	666 (40.9%)	873 (27.3%)		
Age Group (years)	60–69	919 (56.5%)	1982 (61.9%)	13.38	0.001

Characteristic	Category	Depression (n=1627)	Non-depression (n=3200)	χ2	P-value
	70–79	596 (36.6%)	1027 (32.1%)		
	80–85	112 (6.9%)	191 (6.0%)		
Marital Status	Married	1052 (64.7%)	2476 (77.4%)	88.03	< 0.001
	Living alone	575 (35.3%)	724 (22.6%)		
Alcohol Consumption	Never	1181 (72.6%)	2004 (62.6%)	48.73	< 0.001
	≤ 3 times/month	80 (4.9%)	243 (7.6%)		
	≥ 1 time/week	366 (22.5%)	953 (29.8%)		
Chronic Conditions	0	924 (56.8%)	2072 (64.8%)	57.33	< 0.001
	1	412 (25.3%)	794 (24.8%)		
	≥ 2	291 (17.9%)	334 (10.4%)		
Sleep Duration	< 6 hours	645 (39.6%)	798 (24.9%)	116.64	< 0.001
	6–8 hours	837 (51.4%)	2120 (66.2%)		
	> 8 hours	145 (8.9%)	282 (8.8%)		
Pain Locations	0	504 (32.8%)	1799 (56.2%)	258.32	< 0.001
	1	251 (16.3%)	448 (14.0%)		
	2–3	365 (23.8%)	515 (16.1%)		
	≥ 4	417 (27.1%)	437 (13.7%)		
Internet Use	No	1316 (80.9%)	2203 (68.8%)	78.56	< 0.001
	Yes	311 (19.1%)	997 (31.2%)		

3.2 Results of Feature Selection

The optimal parameter combination for the model was identified using grid search and 5-fold cross-validation: bootstrap=False, max_depth=20, min_samples_leaf=1, min_samples_split=2, and n_estimators=200. With these parameters, the sixteen features were ranked by importance(Figure 1). The ten most important features—"Age," "Pain Location," "Sleep Duration," "Nap Duration," "Social Score," "Maximum Exercise Intensity," "Pension Receipt Frequency," "Number of Chronic Diseases," "Drinking Frequency," and "Marital Status"—were selected for constructing the BP neural network.

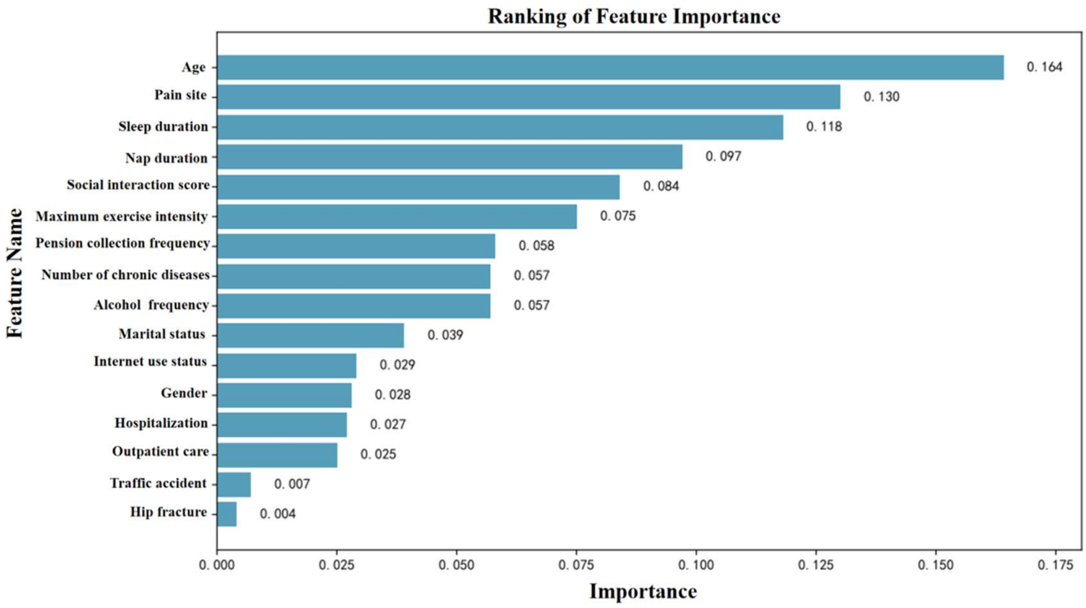

Figure 1 Ranking of feature importance for depression prediction based on the Random Forest algorithm.

3.3 Model Construction and Performance Evaluation

The BP model consists of two hidden layers with 10 and 5 neurons, respectively. The activation function is ReLU, the adaptive learning rate is 0.001, and the maximum number of iterations is set to 1000. During training, the model continuously adjusts the weights based on the input data to minimise prediction error. The final performance metrics of the model on the training and test sets are shown in Table 2. The model's accuracy on the training set is 0.72, and on the test set it is 0.67. The test set performance is slightly lower than the training set performance, but the difference is not significant, indicating that the model has some generalisation ability. Precision and Recall show similar trends on both the training and test sets, indicating that the model is robust to different datasets.

Table 2　Performance metrics of the BP Neural Network model.

Dataset	Precision	Recall	F1-Score	Accuracy	Dataset
Train data	0.71	0.72	0.69	0.72	Train data
Test data	0.65	0.67	0.64	0.67	Test data

3.4 Receiver Operating Characteristic (ROC) Curve

As shown in Figure 2, the model's AUC (Area Under the Curve) value on the training set is 0.62, while on the test set it is 0.59, indicating a slight decrease in performance when generalising to the test set. The upper-left corner of the training set curve is higher, indicating that, below a certain threshold, the model can identify positive examples well. The test set curve is generally close to the diagonal, indicating that the model's discriminative ability on the test set is weaker, especially at lower false-positive rates, where the improvement in the true-positive rate is not significant. Overall, the model's performance in controlling false positives is slightly inferior to that on the training set, but it still possesses a certain degree of discriminative ability.

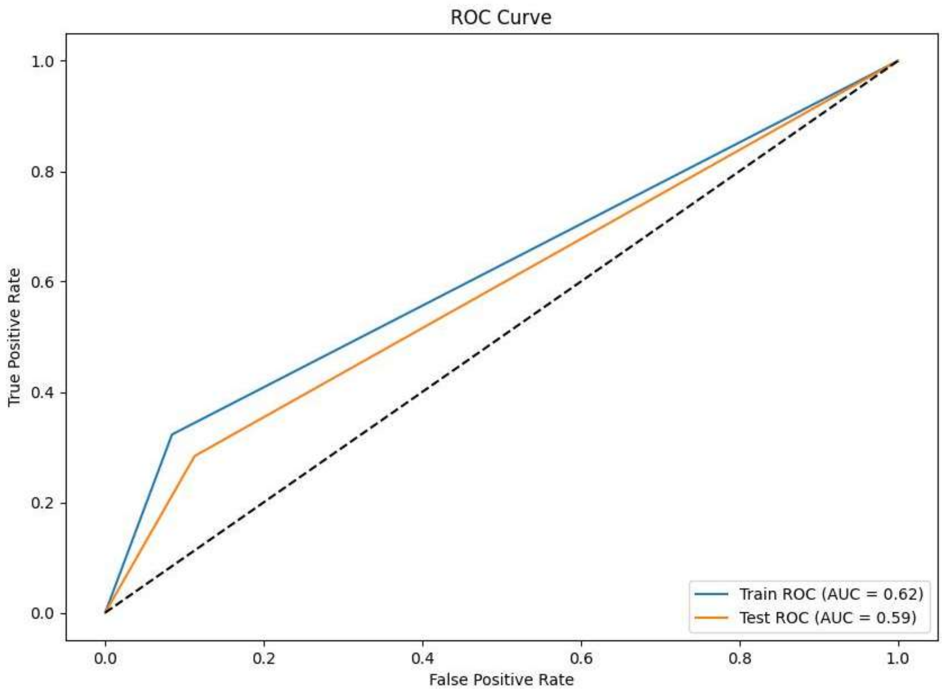

Figure 2 Ranking of feature importance for depression prediction based on the Random Forest algorithm.

4. Discussion

This study constructed a depression prediction model for elderly diabetic patients using CHARLS data, ultimately including 4,827 eligible respondents. We used a hybrid algorithm combining random forest (RF) feature selection and backpropagation neural network (ANN) to train the model. The observed prevalence of depression in the elderly diabetic population was 33.7%. The study also explored the relationship between depressive symptoms and predictive factors across various dimensions through univariate analysis. Results showed that gender, age group, marital status, number of chronic diseases, sleep duration, and number of pain locations were significantly associated with depressive symptoms ($P < 0.05$). Furthermore, the BP neural network model built based on non-invasive features demonstrated good generalisation ability in predicting depression, achieving a test set accuracy of 0.67. Our study shows that physical pain and sleep duration are the most critical factors for predicting depressive status in elderly diabetic patients.

Based on the above findings, this study confirms that physical pain and lifestyle are the most predictive indicators in non-invasive models, rather than traditional sociodemographic characteristics. The number of pain sites ranked highly in the random forest ranking, and the cumulative effect of these sites was significant, consistent with results from several previous epidemiological studies. The risk of depression in patients with chronic diseases increased linearly with the increase in the number of body pain sites, and chronic pain significantly influenced the generation of negative emotions in diabetic patients compared to other physiological indicators [13, 14, 15].

Extensive physical pain in elderly diabetic patients can trap them in a feedback loop of "pain-loneliness-depression" by limiting physical activity and worsening sleep quality [16, 17]. This means that in clinical practice, psychological screening for diabetic patients should not only focus on "whether there is pain," but should also observe the location, intensity, and potential association effects of pain to identify individuals at high risk of psychological distress. Univariate analysis of sleep characteristics showed a significant correlation between sleep duration and depression, validating the importance of sleep as an interventionable behaviour. However, it is noteworthy that although "nap duration" was not significant in the univariate test ($P = 0.093$), it was identified as a key predictive feature in the nonlinear neural network model. This suggests that napping has an indirect impact on the risk of depression in elderly diabetic patients. Prolonged or forced naps can lead to mild inflammation, which can result in daytime fatigue [18, 19]. Napping may also serve as a compensatory mechanism for nighttime sleep deprivation [20, 21], acting as a confounding factor for health risks in complex models, information that is easily overlooked in traditional model analyses.

Finally, we would like to point out that although there is still room for improvement in the algorithmic AUC (0.59) and test accuracy (0.67) of our model, its high negative predictive value (NPV) remains valuable for population screening. In practical applications, this means that the model can exclude most "low-risk" individuals with high confidence. For community physicians, this triage capability is more practically valuable than accurate diagnosis, ensuring improved accessibility of psychological intervention services for elderly diabetic patients in a depressive state, even with cost constraints.

5. Limitations

Our study still has the following limitations. First, the AUC of 0.59 reflects the inherent challenges in modelling structured medical data. The reliance on the CES-D 10 self-report scale for depression diagnosis, with its subjectivity and memory bias leading to "label noise," limits the theoretical upper limit of predictions. Furthermore, the CHARLS 2020-based cross-sectional design makes it impossible to determine the temporal sequence of pain, sleep disturbances, and depression; future models should incorporate longitudinal tracking data to capture dynamic evolution.

Conclusions

This study confirms that non-invasive characteristics centred on physiological pain distribution, sleep behaviour, and social participation are key early warning signals of depression risk in elderly diabetic patients. This model not only technically achieves non-linear classification but also provides a low-barrier, scalable screening tool in practice. By integrating this predictive logic into community chronic disease management systems, it is hoped that high-risk groups can be identified early, thereby improving the overall quality of life and adherence to management for elderly diabetic patients.

Funding

This Research is funded by 2024 school level research project in Neusoft Institute Guangdong. Project number: 2024DRXJ030

References

Yang, K., Fang, Y., He, J. and Li, J. (2025) ‘Prevalence and risk factors of depression in patients with diabetes mellitus: a systematic review and meta-analysis’, Frontiers in Endocrinology, 16, 1660478. doi: 10.3389/fendo.2025.1660478.

Luo, C., Yu, X.M., Zeng, M.Q., Duan, C.Z., Xu, S.Y., Zhu, C.Y., Zheng, Z.G., Sun, D., Fang, J. and He, D.J. (2025) ‘Breaking the diabetes-depression cycle: Exploring shared mechanisms, neuroinflammation, and emerging interventions for metabolic-mood comorbidities’, World

Journal of Diabetes, 16(7), 107406. doi: 10.4239/wjd.v16.i7.107406.

Lei, A.A., Phang, V.W.X., Lee, Y.Z., Kow, A.S.F., Tham, C.L., Ho, Y.C. and Lee, M.T. (2025) ‘Chronic Stress-Associated Depressive Disorders: The Impact of HPA Axis Dysregulation and Neuroinflammation on the Hippocampus—A Mini Review’, International Journal of Molecular Sciences, 26, 2940. doi: 10.3390/ijms26072940.

Oh, D.J., et al. (2024) ‘Association of depression with the progression of multimorbidity in older adults: A population-based cohort study’, The American Journal of Geriatric Psychiatry, 32(8), pp. 957-967.

Ju, C., et al. (2025) ‘Interactive effect of sleep duration, lifestyle factors and comorbidity on depressive symptoms: Insights from the China health and retirement longitudinal study’, Journal of Affective Disorders, 379, pp. 900-912.

Li, Z., et al. (2025) ‘Social support mediates the relationship between depression and subjective well-being in elderly patients with chronic diseases: Evidence from a survey in Rural Western China’, PLoS ONE, 20(6), e0325029.

Chekroud, A.M., et al. (2021) ‘The promise of machine learning in predicting treatment outcomes in psychiatry’, World Psychiatry, 20(2), pp. 154-170.

Ke, H., Xu, A., Zhou, H., Chen, J., Wu, W., He, Q. and Cao, H. (2025) ‘Machine learning models of depression in middle-aged and older adults with cardiovascular metabolic diseases’, Journal of Affective Disorders, 387, 119494. doi: 10.1016/j.jad.2025.119494.

Murri, M.B., Cattelani, L., Chesani, F., Palumbo, P., Triolo, F. and Alexopoulos, G.S. (2022) ‘Risk prediction models for depression in community-dwelling older adults’, The American Journal of Geriatric Psychiatry, 30, pp. 949–960.

Speiser, J.L., et al. (2019) ‘A comparison of random forest variable selection methods for classification prediction modeling’, Expert Systems with Applications, 134, pp. 93-101.

Marcus, G. (2018) ‘Deep learning: A critical appraisal’, arXiv preprint, arXiv:1801.00631.

Dadgar, S.M.H. and Kaardaan, M. (2017) ‘A hybrid method of feature selection and neural network with genetic algorithm to predict diabetes’, International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), 7(24), pp. 3397-3404.

Tanik, N., Sarp, U., Ucar, M., Celikbilek, A., Balbaloğlu, Ö., Ak, H., Atalay, T., Arik, H.O., Okyay, M. and İnan, L. (2016) ‘Pain, depression and sleep disorders in patients with diabetic and nondiabetic carpal tunnel syndrome: A vicious cycle’, Arquivos de Neuro-Psiquiatria, 74, pp. 207-211. doi: 10.1590/0004-282X20160020.

Miao, E., Wu, Q. and Cai, Y. (2024) ‘Mediating effect of depressive symptoms on the relationship of chronic pain and cardiovascular diseases among Chinese population: evidence from the CHARLS’, Journal of Psychosomatic Research, 180, 111639.

Stubbs, B., et al. (2016) ‘The epidemiology of back pain and its relationship with depression, psychosis, anxiety, sleep disturbances, and stress sensitivity: Data from 43 low-and middle-income countries’, General Hospital Psychiatry, 43, pp. 63-70.

Martin, D.J., et al. (2012) ‘The impact of chronic musculoskeletal pain on physical activity in older adults: implications for self-management’, Journal of Aging and Physical Activity, 20(Supplement), pp. S294-S295.

Christiansen, J., et al. (2021) ‘Loneliness, social isolation, and chronic disease outcomes’, Annals of Behavioral Medicine, 55(3), pp. 203-215.

Leng, Y., et al. (2014) ‘Daytime napping, sleep duration and serum C reactive protein: a population-based cohort study’, BMJ Open, 4(11), e006071.

Hu, M., et al. (2021) ‘Sleep, inflammation and cognitive function in middle-aged and older adults: a population-based study’, Journal of Affective Disorders, 284, pp. 120-125.

Han, X., et al. (2016) ‘Long sleep duration and afternoon napping are associated with higher risk of incident diabetes in middle-aged and older Chinese: the Dongfeng-Tongji cohort study’, Annals of Medicine, 48(4), pp. 216-223.

Souabni, M., Souabni, M.J., Hammouda, O., Romdhani, M., Trabelsi, K., Ammar, A. and Driss, T. (2022) ‘Benefits and risks of napping in older adults: A systematic review’, Frontiers in Aging Neuroscience, 14, 1000707.