ORIGINAL RESEARCH

Machine Learning Methods to Evaluate the Depression Status of Chinese Recruits: A Diagnostic Study

Mengxue Zhao[1]
Zhengzhi Feng[2]

[1]Department of Military Psychology, Faculty of Medical Psychology, Army Medical University, Chongqing, People's Republic of China; [2]Faculty of Medical Psychology, Army Medical University, Chongqing, People's Republic of China

Purpose: Traditional questionnaires assessing the severity of depression are limited and might not be appropriate for military personnel. We intend to explore the diagnostic ability of three machine learning methods for evaluating the depression status of Chinese recruits, using the Chinese version of Beck Depression Inventory-II (BDI-II) as the standard.

Patients and Methods: Our diagnostic study was carried out in Luoyang City (Henan Province, China; 10/16/2018–12/10/2018) with a sample of 1000 Chinese male recruits selected using cluster convenient sampling. All participants completed the BDI and 3 questionnaires including the data of demographics, military careers and 18 factors. The participants were randomly selected as the training set and the testing at 2:1. The machine learning methods tested for assessing the presence or absence of depression status were neural network (NN), support vector machine (SVM), and decision tree (DT).

Results: A total of 1000 participants completed the questionnaires, with 223 reporting depression status and 777 not. The highest sensitivity was observed for DT (94.1%), followed by SVM (93.4%) and NN (93.1%). The highest specificity was observed for NN (60.0%), followed by SVM (58.8%) and DT (43.3%). The area under the curve (AUC) of the SVM was the largest (0.862) compared with NN (0.860) and DT (0.734). The regression prediction error and error volatility of the SVM were the smallest.

Conclusion: The SVM has the smallest prediction error and error volatility, as well as the largest AUC compared with NN and DT for assessing the presence or absence of depression status in Chinese recruits.

Keywords: depression, questionnaire, military, machine learning, diagnosis

Introduction

Major depressive disorder (MDD) is characterized by persistent low mood, lack of positive affect, and loss of interest in usually pleasurable activities, which accounts for approximately 20% of people in their lifetime.[1,2] The identified risk factors for MDD include family or personal history of MDD, chronic medical illness, alcohol or substance use, post-traumatic stress syndrome (PTSD), job change or financial difficulty, domestic abuse or violence, female, low income and unemployment, and disability.[1,3]

The population of military personnel are at high risk of developing or exacerbating MDD owing to working in highly stressful conditions and environment, and a high incidence of PTSD.[4–8] These mental health problems are associated with high costs and decreases in productivity. It is reported that

Correspondence: Zhengzhi Feng
Faculty of Medical Psychology, Army Medical University, Chongqing, People's Republic of China
Tel +86 13228688828
Fax +86 23-68752341
Email fzz@tmmu.edu.cn

9–20% of military personnel returning from deployment suffer from mental illness, with 50% of them requiring mental health services.[9–11] The stressors are limited to not only those living in the combat environment but also those returning from deployment.[12,13] The depressive status of military personnel is unique because of the violent training in the military, while they do not have to bear or witness violence in the real world.[14,15] Recruits can be apprehensive toward this violence and resentful at mandatory military service.[16,17] According to a research on the prevalence of depression and associated factors among military personnel, having siblings, military work type, smoking, a sick person at home, and problematic relations with fathers, co-workers, supervisors, subordinates, and relatives are the main factors associated with an increased likelihood of depression.[18]

Numerous tools are available for assessing recruits' depression severity and are useful for screening MDD or for susceptibility to MDD. Those classic tools include the Zung Self-rating Depression Scale (SDS),[19–21] Symptom Checklist 90 (SCL-90),[22] and Beck Depression Inventory (BDI),[23,24] among others. Those questionnaires are valuable for the diagnosis of MDD, but they have limitations such as cultural adaptation and the fact that an intelligent individual can "cheat" and answer what the interviewer wants to hear. In addition, they are not adapted to the harsh working conditions of the military.[25,26] BDI is a measure of current depressive symptoms. It is not a tool for predicting future depressive traits. A previous study showed that pre-deployment mental health screening might reduce psychiatric disorders during and after deployment.[27] Soldiers who scored high for hope, optimism, confidence, and resilience before deployment were less likely to develop psychiatric illness.[28] Another study developed a questionnaire administered early in soldiers' carriers, and can predict future mental health problems.[5]

New methods based on machine learning methods are now being developed for the assessment of emotional status.[29–34] Nevertheless, the existing studies have contradictory results and deficiencies, such as using various machine learning methods for evaluating depression status and the selection of the calibration questionnaires. Many studies have used the SDS, and few used the BDI for calibration. In addition, there were no studies focusing on the evaluation methods and results of the depression status in the population of Chinese soldiers, particularly in Chinese recruits.

Therefore, this study aimed to explore the diagnostic ability of three machine learning methods, neural networks, support vector machines (SVM), and decision trees, so as to evaluate the depression status of Chinese recruits, using the BDI as the standard. The results could provide additional tools for the detection of the depression status and its susceptibility to this specific population.

Methods
Study Design and Participants
This study was carried out in Luoyang City (Henan Province, China) between October 16 and December 10, 2018. A sample of 1000 Chinese male soldiers were selected using the method of cluster convenient sampling. This study was approved by the Medical Ethics Committee of the Army Medical University of the Chinese People's Liberation Army. The study was anonymous. The soldiers consented to this study and participated voluntarily. The soldiers were informed about the purpose of the study, and it was conducted in accordance with the Declaration of Helsinki.

The inclusion criteria were: 1) age ≥18 years; 2) male; 3) <1 year of military service; 4) junior high school education and above; and 5) residing in Luoyang City, Henan Province. The exclusion criteria were: 1) subjects with severe physical illness or schizophrenia, bipolar disorder, cerebral organic diseases, epilepsy, substance abuse, and other mental disorders; 2) non-cooperation or serious data loss; or 3) could not complete the investigation due to withdrawal, distribution, deployment, injuries, etc.

Survey Tools
The BDI was used as the calibration questionnaire. The BDI has 21 items mainly used to assess the severity of depression. Each item corresponds to a symptom category, such as pessimism, suicidal intention, sleep disorders, and social withdrawal. The severity of depressive symptoms assessed by the BDI ranges from none to extremely severe and is classified into four levels, with a value of 0–3. A total score of 4 points or less represents no depressive state, 5–13 points a mild depressive state, 14–20 points a moderate depressive state, and 21 points or higher a severe depressive state.[35] Anyone with a score below 4 is considered not depressed.[35] The results of the BDI assessment were converted into binary categories, that was, with or without depression.

The Chinese version of Beck Depression Inventory-II is validated among depression patients.[36] The Cronbach

coefficient of BDI is 0.83. The odd and even split-half reliability coefficient was 0.86 (Spearman-Brown correlation coefficient was 0.93). For the retest consistency, the stability coefficient of this scale for retesting within a few weeks was usually 0.70–0.80. For the convergent validity, the BDI was significantly associated with clinical depression assessment, with a correlation coefficient of 0.60–0.90, which varied with sample size. For discriminant validity, the BDI had a greater correlation with clinical depression assessment (0.59) than that of anxiety (0.14).[23,37,38]

Eighteen factors from three questionnaires were used as a part of the input features. Military mental health status questionnaire (MMHSQ) included 7 factors: psychosis, depression, suicidal tendency, post-traumatic stress, sleep problems, social phobia and antisocial tendency. The military mental health ability questionnaire (MMHAQ) included 6 factors: emergency dealing ability, stress tolerance ability, calm ability, adapting ability, physical control ability and cooperation ability. There are 5 factors in the mental quality questionnaire for army-men, including intelligence, loyalty, courage, frustration, and confidence. All these scales have good content validity, structure validity and criterion validity.[39–41]

Observation Indicators and Data Collection and Processing

The input features of three machine learning models include not only demographic data but also psychological self-reports data. And the psychological self-reported data came from the 3-dimension 3-rank model (3D3RM), which is thought to be representative of China.[42] The baseline data collected in this study include age, duration of military service, serving place, sex, and personnel classification. Each participant completed the BDI scale assessment and the evaluation of 18 factors of three questionnaires. So, the input features of three machine learning models were age, duration of military service, serving place, sex, personnel classification and 18 factors from 3 questionnaires.

The research subjects were randomly selected as the training set and the testing set at a ratio of 2:1 (N: 667, 333). The scores of mental health ability and mental health quality were reversely scored. The scores of 18 factors were calculated, which were the original score of the factor. The BDI score was calculated, which was the original score of the calibration. The extremum standardization method was used to normalize the original score of the factor and the original score of the calibration, and the data were converted into a score using the centesimal system, which was the standard score. The k-means method was used to perform binary clustering of the standard scores of BDI, and to determine the category of each sample. The standard scores of durations of military service, age, personnel category, and 18 factors were used as the input features of machine learning. The BDI centesimal score was used as the outcome variable for model training and testing. The goodness and badness of the model were presented by parameters such as sensitivity, specificity, accuracy, and area under the curve (AUC).

Machine Learning Methods

Neural network: This study used a backpropagation neural network (BPNN) with only one hidden layer to handle the model calculation, including the processes of network initialization, hidden layer output calculation, output layer output calculation, error calculation, weights updating, thresholds updating, and determining whether the algorithm iteration was ended. Figure 1 shows the schematic diagram of neural network training.

SVM: This study used SVM to find the most suitable hyperplane based on the indicators. Figure 2 shows the flowchart of SVM parameter optimization.

Decision tree: This study used the C4.5 decision tree algorithm, which divided the characteristics with the information gain rate and discretized the value space. It could convert continuous values and missing values into discrete attributes and then perform calculations, overcoming the shortcomings of using information gain to select characteristics, and had pruned in the process of constructing the tree, and improved the accuracy. Figure 3 shows the local schematic diagram of the two-level regression tree model.

Statistical Analysis

All data were analyzed using Pycharm Community Edition (Version 2019.1.3, JetBrains, Czech Republic) and SPSS 24.0 (IBM, Armonk, NY, USA). All continuous variables were tested for normal distribution using the Kolmogorov–Smirnov test. The data that conformed to the normal distribution were expressed as means ± standard deviations; otherwise, they were presented as medians (ranges). Categorical variables were expressed as frequencies (percentages) and analyzed using the chi-square test. A grid search algorithm was used for parameter selection of the SVM classification model and regression model to obtain the optimal error penalty factor C and the parameter gamma of the kernel function. The comparison of model
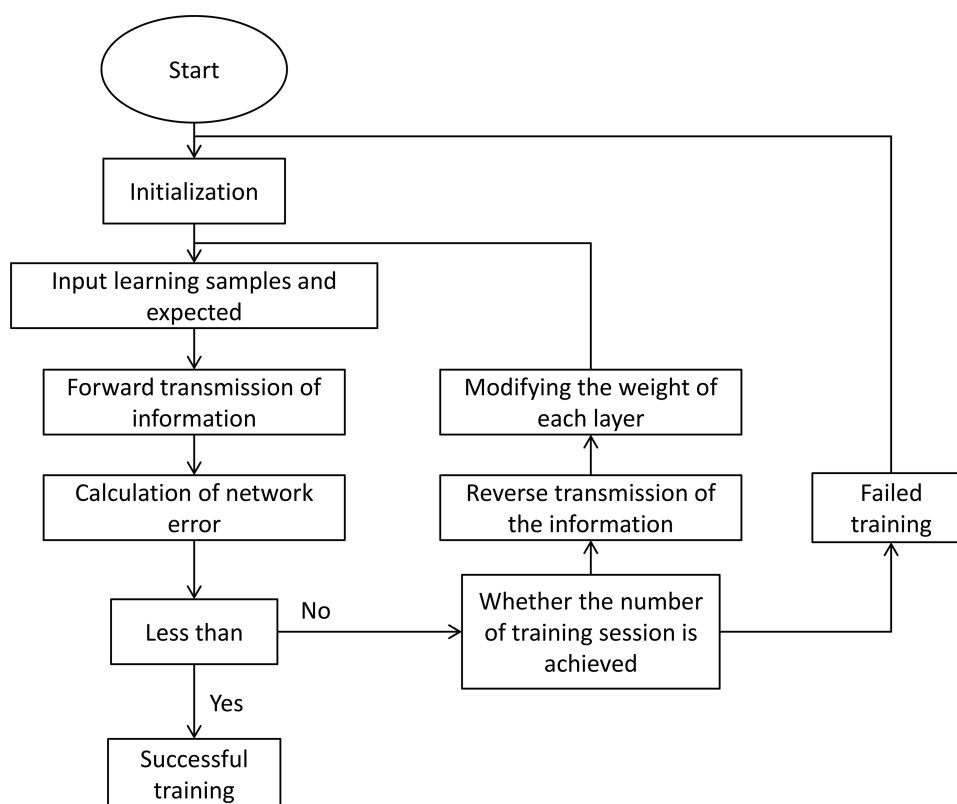
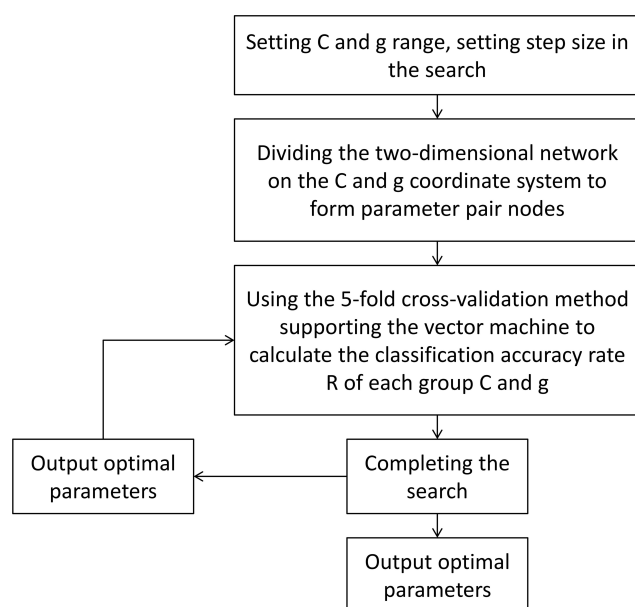**Figure 1** Schematic diagram of neural network training.

Figure 2 Support vector machine diagram.

classification results was presented using receiver operating characteristic (ROC) curves. The comparisons of the differences in regression results were calculated by non-parametric tests. In the regression model, the dependent variables were the combined scores of the three combined weighted factors after converting to the centesimal system. The parameters used to compare the regression accuracy of the three machine learning models were obtained, including the mean square error (MSE), root means square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). $P<0.05$ was considered statistically significant.

Results
Characteristics of the Participants

In this study, a total of 1000 questionnaires were distributed, and 1000 valid questionnaires were retrieved. The effective retrieval rate was 100%. At the Luoyang training site, 1000 military soldiers were selected using the method of cluster convenient sampling. All of them were males, aged 18–24 years, with an average of 19.3±1.4 years. The median age was 19 years old, and the mode was 18 years old. The military service in all samples was <1 month. There are no missing BDI items. According to the BDI results, anyone with a score below 4 is considered not depressed.[35] Seven hundred and seventy-seven participants had no depressive symptoms, 171 samples had mild depressive symptoms, 40 samples had moderate

X[5] <=12.5
entropy=110.144
samples=666
value=6.365

X[5] <=7.5
entropy=53.048
samples=617
value=4.696

X[20] <=55.0
entropy=352.372
samples=49
value=27.381

X[8] <=12.5
entropy=18.533
samples=375
value=2.222

X[14] <=14.5
entropy=82.337
samples=242
value=8.531

X[13] <=18.0
entropy=194.967
samples=47
value=24.734

X[12] <=29.0
entropy=17.347
samples=2
value=89.585

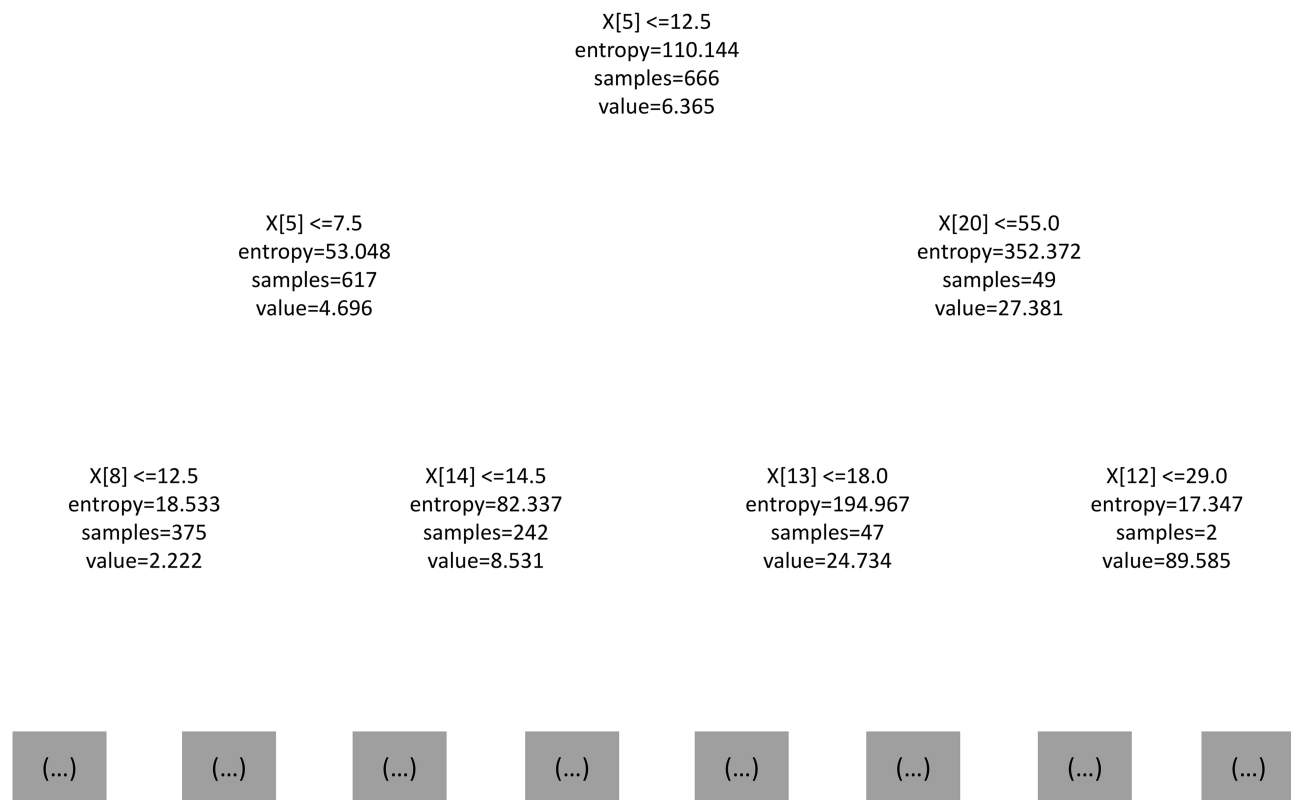(...) (...) (...) (...) (...) (...) (...) (...)

Figure 3 Local schematic diagram of the two-level regression tree model.

depressive symptoms, and 12 participants had severe depressive symptoms. In other words,777 participants had no depressive symptoms, and 223 had depressive symptoms. There were 667 participants in the training set and 333 in the validation set after randomization.

The Classification Effect of the Three Machine Learning Methods

The highest sensitivity for MDD was observed for the decision tree (94.1%), followed by the SVM (93.4%) and the neural network (93.1%). The highest specificity for MDD was observed for the neural network (60.0%), followed by the SVM (58.8%) and the decision tree (43.3%). (Table 1) Regarding the ROC curve, the AUC of the SVM was the largest (0.862), compared with that of the neural network (0.860) and that of the decision tree (0.734) (Figure 4).

Regression Analysis of Three Machine Learning Methods

The comparison results of the regression model parameters showed that SVM had the smallest regression prediction error and the error volatility (Table 2), thus presenting the best regression results.

Discussion

Traditional questionnaires for the detection of MDD or susceptibility to MDD have disadvantage and might not be appropriate for military personnel.[25,26] Therefore, this study aimed to explore the predictive and diagnostic ability of three machine learning methods for evaluating the depression status of Chinese recruits, using the BDI as the standard. The highest sensitivity was observed for the decision tree, followed by the SVM and neural network. The highest specificity was observed for the neural network, followed by the SVM and decision tree. The AUC of the SVM was the largest compared with the neural

Table 1 The Sensitivity, Specificity, and AUC of Three Machine Learning Methods (Neural Network, Support Vector Machine, and Decision Tree) for Evaluating the Depression Status of Chinese Recruits

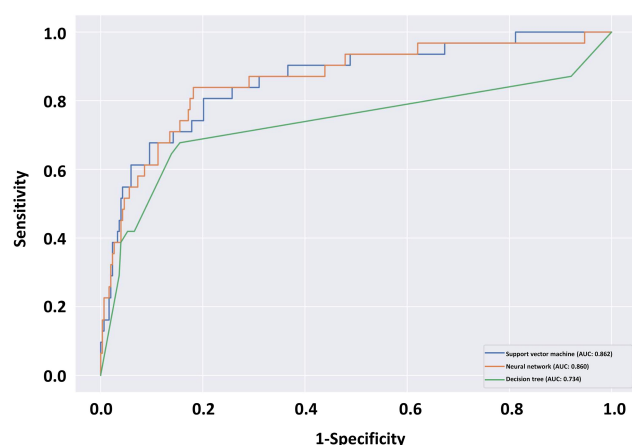|	Neural Network	Support Vector Machine	Decision Tree
Sensitivity	0.931	0.934	0.941
Specificity	0.600	0.588	0.433
AUC	0.860	0.862	0.734

Abbreviation: AUC, area under the curve.

Figure 4 Receiver operating characteristics (ROC) curve for depression in Chinese recruits from three machine learning methods: support vector machine (blue line), neural network (orange line), and decision tree (green line).

network and decision tree. The regression prediction error and error volatility of the SVM were the smallest. Therefore, an SVM had the best prediction error, error volatility, and AUC compared with the neural network and decision tree for the detection of depression status in Chinese recruits.

Sensitivity and specificity are important indicators of machine learning. In this study, high sensitivity and medium specificity may indeed result in high false positive. This result may be attributed to the dichotomy problem with BDI, which is a rigorous screening criterion for depression. However, the sensitivity of this model is significant for the identification of depressed state in Chinese military recruits. Based on the results of this model and combined with clinical interviews, the new recruits can be selected and eliminated without missing a single person. Furthermore, feature selection in machine learning models has the potential to solve the specificity problem. We will go further in the follow-up study.

Table 2 Comparison of Regression Model Parameters of Three Machine Learning Methods (Neural Network, Support Vector Machine, and Decision Tree)

	Neural Network	Support Vector Machine	Decision Tree
R^2	0.465	0.544	0.477
MSE	64.884	60.087	68.043
RMSE	8.055	7.752	8.249
MAE	4.821	4.659	5.243
MAPE	0.897	1.030	0.822

Abbreviations: R^2, coefficient of determination; MSE, mean square error; RMSE, root mean square error; MAE, mean absolute error; MAPE, mean absolute percentage error.

The military is supposed to commit to the well-being of their troopers. This not only ensures an optimal fighting force but also contributes to the unity of the army. Taking care of military personnel is conducive to military fitness and operational readiness.[43] In the present study, the frequency of having depressive symptoms was 22.3%, higher than that in the general Chinese population.[44,45] The reason might be that all participants were recruits of less than 1 month, and they were still adapting to military life.[16,17] The recruits in the study were between 18 and 24 years old, but the onset of MDD tends to be greater in later adolescence/earlier adulthood, so the comparison with the entire population may not be all that fair.

Different countries have different methods of screening for depression. In Canada, screening mostly relies on interviews, while in the United States, screening tools are used.[14,15] Nevertheless, self-reported data on the psychological dimension improve model performance. A machine-learning technique was applied in Britain in a study of 13,690 current or former servicemen and found out that self-report could effectively distinguish those with PTSD.[34] The US military improved the accuracy of machine-learning models from 17.5% to 29.4% (67.9% improvement) by adding self-report into management data.[30] In the present study, the input features of three machine learning models include not only demographic data but also psychological self-reports. Additionally, the psychological self-reported data came from the 3-dimension 3-rank model (3D3RM), which is thought to be the representative of China.[42] The 3D3RM focuses on the mental health of military soldiers, which are three dimensions of psychological stability, psychological capability, and psychological quality.[42] Therefore, the use of self-reported data in the present study probably increases the value and reliability of the machine learning models.

Previous studies examined machine learning methods for the prediction of MDD and related conditions. Batterham et al[46] elaborated on a decision tree model that could assess the 4-year risk of MDD, and that was better than logistic regression. A decision tree allows for the inclusion of high-order interaction terms than a logistic regression model and allows breaking down a population into categories with clinical usefulness. Karstoft et al[31] used a Markov boundary feature selection algorithm for generalized local learning and achieved a model that could predict PTSD with high accuracy in Danish military personnel. Gradus et al[47] elaborated random forest models for the detection of trauma and suicidal ideation after

deployment in Iraq and Afghanistan. The results showed that their model achieved a high degree of replicability that could be used as a screening tool among veterans. Moreover, Rosellini et al[48] used super-learning to develop models to show that they could detect MDD, suicidal ideation, and anxiety with high accuracy, identifying soldiers that might benefit from not being deployed. Nevertheless, those studies only examine one machine learning method. Leightley et al[34] compared SVM, random forests, artificial neural networks, and bagging for the detection of PTSD in British military personnel. In their study, random forests achieved the highest accuracy (97%), while SVM achieved the highest sensitivity (70%). In our study, sensitivity was >93% for all three machine learning methods (neural network, SVM, and decision tree) as BDI is the standard. SVM achieved the highest AUC and the lowest errors. Nevertheless, further studies are necessary to refine the model. Nevertheless, a strength of the present study was the use of BDI as the standard compared with previous studies that used the SDS, since a head-to-head comparison of the BDI and SDS showed that the BDI was superior in terms of psychometric properties.[49]

Nevertheless, comparing deep learning models can be difficult because each model has its own advantages and disadvantages. Furthermore, different input variables may be the primary factors influencing output. The fuzzy neural network was used in the mental health assessment of college students.[50] The gradient enhancement model was better than the random forest and the regression in predicting mortality of patients admitted to the hospital.[51] The risk of suicide in American soldiers was evaluated using the Naive Bayes, random forest, SVM, and elastic net penalty regression, of which the sensitivity was optimal for an inelastic mesh classifier.[32] In another study of suicide risk in soldiers, the AUC of logistic regression was 0.62, while that of the Super Learner model was 0.83.[52] In a study by Ding et al, 144 patients with MDD and 204 matched healthy controls were recruited. They were required to watch a series of affective and neutral stimuli under monitoring using EEG, eye tracking, and galvanic skin response; then, three machine learning algorithms including random forests, logistic regression, and SVM were trained to build dichotomous classification model.[53] The results showed that the highest classification f1 score was obtained by logistic regression algorithms (79.6% accuracy, 76.7% precision, 85.2% recall, and 80.7% f1 score).[53] The SVM is a supervised learning method to classify the annotated results with only one

solution. The SVM is considered as one of the most effective classification algorithms available. The advantage of SVM is that it suits in building classifiers with a small number of samples for each category and tries to adapt the nonlinear discriminant function to achieve more accurate classification.[29] Meanwhile, the generalization performance of SVM is not stable, and BPNN has different advantages in sensitivity and specificity compared with SVM.[54,55] The utility of these machine learning models also differs based on their input variables.

A recent systematic review has pointed out that the current available studies have methodological limitations, but that research is on the right way.[56] In particular, there is a need to recapitulate the results in multiple centers. Another research avenue could also be the inclusion of neuroimaging biomarkers, which have been identified for PTSD.[57–61] Indeed, the combination of self-reported data and neuroimaging features could provide a complete model of the risk of depressive status in military personnel. A study revealed that using functional MRI data in an SVM could identify patients with severe depression, but did not perform well for milder depression.[62] Hence, inclusion of self-reported data could help distinguish all grades of depression. Indeed, self-reported data can be used to stratify the patients effectively using machine learning.[63] Additionally, self-reported and predeployment demographic/clinical data are much cheaper than imaging data. The utility of the imaging data might be limited if it does not improve the predictions. A review highlights that the common challenges of the learning machine studies in depressive disorders are the small sample size, feature reduction, overfitting, classification methods, and cross-validation.[58] Those points should be taken into consideration in future studies.

This study has limitations. First, the sample size was small and limited to a single station. BDI has good reliability and validity, and there are many clinical and applied studies using the most common scale in China.[23,24,36] BDI can differentiate subtypes of depression and distinguish depression from anxiety.[64] Nevertheless, BDI is a severity rating scale, rather than a screening scale, and is often used as a self-scoring questionnaire. There are limitations that only one tool was used to estimate the depression status and to screen MDD without a clinical interview and the machine learning model cannot overcome the limitations of the BDI scale. The use of multiple tools could be explored in the future to improve the sensitivity and specificity of the models. Moreover, routine blood biochemistry data could also be

included.[65,66] In addition, this study lacks data-driven feature selection for depression, and model generalization ability has not been clarified. Second, machine learning methods have their own limitations, such as not considering some relationships that might be bidirectional. Third, the predictive power of the model has not been verified by prospective studies. The present study included cross-sectional data and should be validated using longitudinal data. Fourth, only men were recruited, and the models should be validated in women. Moreover, a large number of complex socio-demographic variables and career variables have not been included into the model. Finally, the mediators of depression were not explored.

Conclusion

In conclusion, an SVM had the smallest prediction error and error volatility, as well as the largest AUC compared with the neural network and decision tree for the detection of MDD in Chinese recruits. This tool could be used to estimate the depression or susceptibility to MDD in recruits and identify those who could benefit from interventions to prevent MDD. In addition, this tool could be particularly valuable since military personnel is often unwilling to disclose psychiatric conditions by themselves.[67]

Ethics Approval and Informed Consent

This study was approved by the Medical Ethics Committee of the Army Medical University of the Chinese People's Liberation Army. The soldiers consented to this study and were not ordered to participate. The soldiers were informed about the purpose of the study, and that it was conducted in accordance with the Declaration of Helsinki.

Acknowledgments

This study was supported by the Foundation for Humanities and Social Science (2018XRW001).

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare that they have no conflicts of interest.

References

1. American Psychiatric Association Steering Committee on Parctice Guidelines. *Practice Guidelines for the Treatment of Patients with Major Depressive Disorder*. 3rd ed. Philadelphia: American Psychiatric Association; 2010.
2. World Health Organization. *Depression and Other Common Mental Disorders. Global Health Estimates*. Geneva: World Health Organization; 2017.
3. Otte C, Gold SM, Penninx BW, et al. Major depressive disorder. *Nat Rev Dis Primers*. 2016;2:16065.
4. Shen YC, Arkes J, Kwan BW, Tan LY, Williams TV. Effects of Iraq/Afghanistan deployments on PTSD diagnoses for still active personnel in all four services. *Mil Med*. 2010;175(10):763–769. doi:10.7205/MILMED-D-10-00086
5. Shen YC, Arkes J, Lester PB. Association between baseline psychological attributes and mental health outcomes after soldiers returned from deployment. *BMC Psychol*. 2017;5(1):32. doi:10.1186/s40359-017-0201-4
6. Thomas JL, Wilk JE, Riviere LA, McGurk D, Castro CA, Hoge CW. Prevalence of mental health problems and functional impairment among active component and National Guard soldiers 3 and 12 months following combat in Iraq. *Arch Gen Psychiatry*. 2010;67(6):614–623. doi:10.1001/archgenpsychiatry.2010.54
7. Shen YC, Arkes J, Williams TV. Effects of Iraq/Afghanistan deployments on major depression and substance use disorder: analysis of active duty personnel in the US military. *Am J Public Health*. 2012;102(Suppl1):S80–S87. doi:10.2105/AJPH.2011.300425
8. Milliken CS, Auchterlonie JL, Hoge CW. Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. *JAMA*. 2007;298(18):2141–2148. doi:10.1001/jama.298.18.2141
9. Hoge CW, Auchterlonie JL, Milliken CS. Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *JAMA*. 2006;295(9):1023–1032. doi:10.1001/jama.295.9.1023
10. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *N Engl J Med*. 2004;351(1):13–22. doi:10.1056/NEJMoa040603
11. Wells TS, LeardMann CA, Fortuna SO, et al. A prospective study of depression following combat deployment in support of the wars in Iraq and Afghanistan. *Am J Public Health*. 2010;100(1):90–99. doi:10.2105/AJPH.2008.155432
12. Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *J Consult Clin Psychol*. 2000;68(5):748–766. doi:10.1037/0022-006X.68.5.748
13. Steenkamp MM, Boasso AM, Nash WP, Larson JL, Lubin RE, Litz BT. PTSD symptom presentation across the deployment cycle. *J Affect Disord*. 2015;176:87–94. doi:10.1016/j.jad.2015.01.043
14. Thériault FL, Garber BG, Momoli F, Gardner W, Zamorski MA, Colman I. Mental health service utilization in depressed Canadian armed forces personnel. *Can J Psychiatry*. 2019;64(1):59–67. doi:10.1177/0706743718787792
15. Thériault FL, Gardner W, Momoli F, et al. Mental health service use in depressed military personnel: a systematic review. *Mil Med*. 2020;185(7–8):e1255–e1262. doi:10.1093/milmed/usaa015
16. Williams RA, Hagerty BM, Yousha SM, Hoyle KS, Oe H. Factors associated with depression in navy recruits. *J Clin Psychol*. 2002;58(4):323–337.
17. Bin Zubair U, Mansoor S, Rana MH. Prevalence of depressive symptoms and associated socio-demographic factors among recruits during military training. *J R Army Med Corps*. 2015;161(2):127–131. doi:10.1136/jramc-2014-000253

18. Al-Amri M, A-A MD. Prevalence of depression and associated factors among military personnel in the air base in Taif region. *Am J Res Comm*. 2013;1(12):21–45.

19. Biggs JT, Wylie LT, Ziegler VE. Validity of the Zung self-rating depression scale. *Br J Psychiatry*. 1978;132:381–385. doi:10.1192/bjp.132.4.381

20. Jokelainen J, Timonen M, Keinänen-Kiukaanniemi S, Härkönen P, Jurvelin H, Suija K. Validation of the Zung self-rating depression scale (SDS) in older adults. *Scand J Prim Health Care*. 2019;37(3):353–357. doi:10.1080/02813432.2019.1639923

21. Lee HC, Chiu HF, Wing YK, Leung CM, Kwong PK, Chung DW. The Zung self-rating depression scale: screening for depression among the Hong Kong Chinese elderly. *J Geriatr Psychiatry Neurol*. 1994;7(4):216–220. doi:10.1177/089198879400700404

22. Prinz U, Nutzinger DO, Schulz H, Petermann F, Braukhaus C, Andreas S. Comparative psychometric analyses of the SCL-90-R and its short versions in patients with affective disorders. *BMC Psychiatry*. 2013;13:104. doi:10.1186/1471-244X-13-104

23. Richter P, Werner J, Heerlein A, Kraus A, Sauer H. On the validity of the beck depression inventory. A review. *Psychopathology*. 1998;31(3):160–168. doi:10.1159/000066239

24. García-Batista ZE, Guerra-Peña K, Cano-Vindel A, Herrera-Martínez SX, Medrano LA. Validity and reliability of the Beck Depression Inventory (BDI-II) in general and hospital population of Dominican Republic. *PLoS One*. 2018;13(6):e0199750. doi:10.1371/journal.pone.0199750

25. Gahm GA, Lucenko BA. Screening soldiers in outpatient care for mental health concerns. *Mil Med*. 2008;173(1):17–24. doi:10.7205/MILMED.173.1.17

26. Searle AK, Van Hooff M, McFarlane AC, et al. Screening for depression and psychological distress in a currently serving military population: the diagnostic accuracy of the K10 and the PHQ9. *Assessment*. 2019;26(8):1411–1426. doi:10.1177/1073191117745124

27. Warner CH, Appenzeller GN, Parker JR, Warner CM, Hoge CW. Effectiveness of mental health screening and coordination of in-theater care prior to deployment to Iraq: a cohort study. *Am J Psychiatry*. 2011;168(4):378–385. doi:10.1176/appi.ajp.2010.10091303

28. Krasikoiiva DV, Lester PB, Harms PD. Effects of psychological capital on mental health and substance abuse. *J Leadership Organiz Stud*. 2015;22(3):280–291. doi:10.1177/1548051815585853

29. Wu TC, Zhou Z, Wang H, et al. Advanced machine learning methods in psychiatry: an introduction. *Gen Psychiatry*. 2020;33(2):e100197. doi:10.1136/gpsych-2020-100197

30. Bernecker SL, Rosellini AJ, Nock MK, et al. Improving risk prediction accuracy for new soldiers in the U.S. Army by adding self-report survey data to administrative data. *BMC Psychiatry*. 2018;18(1):87. doi:10.1186/s12888-018-1656-4

31. Karstoft KI, Statnikov A, Andersen SB, Madsen T, Galatzer-Levy IR. Early identification of posttraumatic stress following military deployment: application of machine learning methods to a prospective study of Danish soldiers. *J Affect Disord*. 2015;184:170–175. doi:10.1016/j.jad.2015.05.057

32. Kessler RC, Stein MB, Petukhova MV, et al. Predicting suicides after outpatient mental health visits in the army study to assess risk and resilience in servicemembers (Army STARRS). *Mol Psychiatry*. 2017;22(4):544–551. doi:10.1038/mp.2016.110

33. Kessler RC, Warner CH, Ivany C, et al. Predicting suicides after psychiatric hospitalization in US army soldiers: the army study to assess risk and rEsilience in servicemembers (Army STARRS). *JAMA Psychiatry*. 2015;72(1):49–57. doi:10.1001/jamapsychiatry.2014.1754

34. Leightley D, Williamson V, Darby J, Fear NT. Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort. *J Ment Health*. 2019;28(1):34–41. doi:10.1080/09638237.2018.1521946

35. Xiangdong W. *Manual of Mental Health Rating Scale (Updated Version)*. China Mental Health Magazine; 1999.

36. Wang Z, Huang J, Li ZH, Chen J, Zhang HY, Fang YR. Reliability and validity of the Chinese version of Beck depression inventory-II among depression patients. *Chin Ment Health J*. 2011;25(6):476–480.

37. Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck depression inventory for primary care. *Behav Res Ther*. 1997;35(8):785–791. doi:10.1016/S0005-7967(97)00025-9

38. Beck AT, Steer RA. Internal consistencies of the original and revised Beck Depression Inventory. *J Clin Psychol*. 1984;40(6):1365–1367. doi:10.1002/1097-4679(198411)40:6<1365::AID-JCLP2270400615>3.0.CO;2-D

39. Wei L. *Study on the Assessment of Military Mental Health Status and the Eye Movement Identification*. Chongqing: Army Medical University; 2018.

40. Feng H. *Development of Military Mental Health Ability Questionnaire and Analysis of Its Demographic Characteristics*. Chongqing: Army Medical University; 2018.

41. Feng W, Zhengzhi F, Yaqin L. The compilation of mental quality questionnaire for armymen. *J Prey Med Chin PEA*. 2007;02:101–104.

42. Feng Z. Mental health assessment for military men: theories and model. *J Third Militaru Med Univ*. 2015;22.

43. Murphy D, Ashwick R, Palmer E, Busuttil W. Describing the profile of a population of UK veterans seeking support for mental health difficulties. *J Ment Health*. 2019;28(6):654–661. doi:10.1080/09638237.2017.1385739

44. Lee S, Tsang A, Huang YQ, et al. The epidemiology of depression in metropolitan China. *Psychol Med*. 2009;39(5):735–747. doi:10.1017/S0033291708004091

45. Gu L, Xie J, Long J, et al. Epidemiology of major depressive disorder in mainland china: a systematic review. *PLoS One*. 2013;8(6):e65356. doi:10.1371/journal.pone.0065356

46. Batterham PJ, Christensen H, Mackinnon AJ. Modifiable risk factors predicting major depressive disorder at four year follow-up: a decision tree approach. *BMC Psychiatry*. 2009;9:75. doi:10.1186/1471-244X-9-75

47. Gradus JL, King MW, Galatzer-Levy I, Street AE. Gender differences in machine learning models of trauma and suicidal ideation in veterans of the Iraq and Afghanistan Wars. *J Trauma Stress*. 2017;30(4):362–371. doi:10.1002/jts.22210

48. Rosellini AJ, Stein MB, Benedek DM, et al. Predeployment predictors of psychiatric disorder-symptoms and interpersonal violence during combat deployment. *Depress Anxiety*. 2018;35(11):1073–1080. doi:10.1002/da.22807

49. Campbell MH, Maynard D, Roberti JW, Emmanuel MK. A comparison of the psychometric strengths of the public-domain Zung Self-rating Depression Scale with the proprietary Beck Depression Inventory-II in Barbados. *West Indian Med J*. 2012;61(5):483–488. doi:10.7727/wimj.2010.145

50. Mental health evaluation model based on fuzzy neural network. International Conference on Smart Grid and Electrical Automation (ICSGEA); 2016.

51. Brajer N, Cozzi B, Gao M, et al. Prospective and external evaluation of a machine learning model to predict in-hospital mortality of adults at time of admission. *JAMA Netw Open*. 2020;3(2):e1920733. doi:10.1001/jamanetworkopen.2019.20733

52. Bernecker SL, Zuromski KL, Gutierrez PM, et al. Predicting suicide attempts among soldiers who deny suicidal ideation in the army study to assess risk and resilience in servicemembers (Army STARRS). *Behav Res Ther*. 2019;120:103350. doi:10.1016/j.brat.2018.11.018

53. Ding X, Yue X, Zheng R, Bi C, Li D, Yao G. Classifying major depression patients and healthy controls using EEG, eye tracking and galvanic skin response data. *J Affect Disord*. 2019;251:156–161. doi:10.1016/j.jad.2019.03.058

54. Hai ND. *Comparison Between SVM and Back Propagation Neural Network in Building IDS*. Vol. 240. Dordrecht: Springer; 2013.

55. Lee MC. Comparison of support vector machine and back propagation neural network in evaluating the enterprise financial distress. *Intl J Artific Intel Appl*. 2010;1(3):31–43.

56. Claude LA, Houenou J, Duchesnay E, Favre P. Will machine learning applied to neuroimaging in bipolar disorder help the clinician? A critical review and methodological suggestions. *Bipolar Disord*. 2020;22(4):334–355. doi:10.1111/bdi.12895

57. Zilcha-Mano S, Zhu X, Suarez-Jimenez B, et al. Diagnostic and predictive neuroimaging biomarkers for posttraumatic stress disorder. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5 (7):688–696. doi:10.1016/j.bpsc.2020.03.010

58. Gao S, Calhoun VD, Sui J. Machine learning in major depression: from classification to treatment outcome prediction. *CNS Neurosci Ther*. 2018;24(11):1037–1052.

59. Janssen RJ, Mourão-Miranda J, Schnack HG. Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2018;3(9):798–808. doi:10.1016/j.bpsc.2018.04.004

60. Schnyer DM, Clasen PC, Gonzalez C, Beevers CG. Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. *Psychiatry Res Neuroimaging*. 2017;264:1–9. doi:10.1016/j.pscychresns.2017.03.003

61. Guo H, Qin M, Chen J, Xu Y, Xiang J. Machine-learning classifier for patients with major depressive disorder: multifeature approach based on a high-order minimum spanning tree functional brain network. *Comput Math Methods Med*. 2017;2017:4820935. doi:10.1155/2017/4820935

62. Ramasubbu R, Brown MR, Cortese F, et al. Accuracy of automated classification of major depressive disorder as a function of symptom severity. *NeuroImage Clin*. 2016;12:320–331. doi:10.1016/j.nicl.2016.07.012

63. Kessler RC, van Loo HM, Wardenaar KJ, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry*. 2016;21(10):1366–1371. doi:10.1038/mp.2015.198

64. Aaron T, Steer RA. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin Psychol Rev*. 1988;8 (1):77–100. doi:10.1016/0272-7358(88)90050-5

65. Zheng H, Zheng P, Zhao L, et al. Predictive diagnosis of major depression using NMR-based metabolomics and least-squares support vector machine. *Clin Chim Acta*. 2017;464:223–227. doi:10.1016/j.cca.2016.11.039

66. Dinga R, Marquand AF, Veltman DJ, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry*. 2018;8(1):241. doi:10.1038/s41398-018-0289-1

67. Vannoy SD, Andrews BK, Atkins DC, Dondanville KA, Young-McCaughan S, Peterson AL. Under reporting of suicide ideation in US Army population screening: an ongoing challenge. *Suicide Life Threat Behav*. 2017;47(6):723–728. doi:10.1111/sltb.12318