This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

Machine Learning Based Suicide Ideation Prediction for Military Personnel

Gen-Min Lin, *Member, IEEE*, Masanori Nagamine, Szu-Nian Yang, Yueh-Ming Tai, Chin Lin, Hiroshi Sato

***Abstract*—Military personnel have greater psychological stress and are at higher suicide attempt risk compared with the general population. High mental stress may cause suicide ideations which are crucially driving suicide attempts. However, traditional statistical methods could only find a moderate degree of correlation between psychological stress and suicide ideation in non-psychiatric individuals. This paper utilizes machine learning techniques including logistic regression, decision tree, random forest, gradient boosting regression tree, support vector machine and multilayer perceptron to predict the presence of suicide ideation by six important psychological stress domains of the military males and females. The accuracies of all the six machine learning methods are over 98%. Among them, the multilayer perceptron and support vector machine provide the best predictions of suicide ideation approximately to 100%. As compared with the BSRS-5 score ≥ 7, a conventional criterion, for the presence of suicide ideation ≥ 1, the proposed algorithms can improve the performances of accuracy, sensitivity, specificity, precision, the AUC of ROC curve and the AUC of PR curve up to 5.7%, 35.9%, 4.6%, 65.2%, 4.3% and 53.2%, respectively; and for the presence of more severely intense suicide ideation ≥ 2, the improvements are 6.1%, 26.2%, 5.8%, 83.5%, 2.8% and 64.7%, respectively.**

***Index Terms*—Machine Learning Techniques, Psychological Stress, Suicide Ideation**

This work was supported by Ministry of Science and Technology, Taiwan, under grant MOST 107-2221-E-899-002-MY3.

Gen-Min Lin is with the Department of Internal Medicine, Hualien Armed Forces General Hospital, Hualien 97144, Taiwan; Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei 11490, Taiwan; Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA. (email: farmer507@yahoo.com.tw or gen-min.lin@northwestern.edu).

Masanori Nagamine is with the Division of Behavioral Science, National Defense Medical College Research Institute, Tokorozawa City, Saitama 359-8513, Japan. (e-mail: nagaminemasanori@gmail.com).

Szu-Nian Yang is with the Department of Psychiatry, Tri-Service General Hospital-Beitou Branch, National Defense Medical Center, Taipei 11490, Taiwan. (e-mail: ysn56725@ms4.hinet.net).

Yueh-Ming Tai is with the Department of Psychiatry, Tri-Service General Hospital-Beitou Branch, National Defense Medical Center, Taipei 11490, Taiwan. (e-mail: Samytai@hotmail.com).

Chin Lin is with the Graduate Institute of Life Sciences, National Defense Medical Center, Taipei 11490, Taiwan. (email: xup6fup0629@gmail.com).

Hiroshi Sato is with the Artificial Intelligence Laboratory, Department of Computer Science, National Defense Academy, Yokosuka, Kanagawa 239-8686, Japan. (e-mail: hsato@nda.ac.jp).

I. Introduction

MILITARY personnel are vulnerable to psychological stress because of tense physical training, multiple deployments and responsibilities. The prevalence of major depression was reported, ranging from 2.0% to 37.4% in the US military [1], and that of combat-related posttraumatic stress disorder (PTSD) was reported 2.0%-17.0% among US military war veterans [2]. A meta-analysis showed consistent results that the worldwide pooled prevalence of PTSD in rescue workers was up to 10.0% [3]. The symptoms of mental disorders developed frequently in those of continued combat exposure and those of repeated deployments [4]. The association between military absenteeism and mental health problems has been discussed in [5]. The rate of suicide attempt among active duty US Army personnel has been increasingly higher than that in the civilians [6]. According to an analysis for 27,501 military participants in [7], 14.3% of survey respondents reported suicide ideation and 3.0% committed suicide. In other words, 21% of those with suicide ideation had suicide attempt. As is known, previous studies have revealed a relationship between suicide ideation and psychological stress [8],[9]. To early predict the presence of suicide ideation and further prevent the behavior of suicide are essential and important in the military.

With the technology improvement and the availability of various kinds of real world big data, artificial intelligence (AI) grows fast accordingly. The academics have made great efforts on the computerized algorithms to deal with big data. Machine learning, a combination of AI and computations, could provide accurate diagnosis of diseases and predict the outcomes [10]-[17]. For instance, the circuits for seizure classification and detection by machine learning are implemented in [18]. Recognition of heart murmurs could be achieved by deep neural networks [19]. In addition, Ambale-Venkatesh *et al.* identified the top-20 risk factors of incident cardiovascular events by the random survival forest which performance was better than the conventional risk calculators [20]. Therefore, using machine learning and deep learning techniques has become an efficient and reliable tool for clinical practice by physicians globally.

High mental stress may cause suicide ideations which are crucially driving suicide attempts. However, traditional statistical methods find merely a moderate correlation between psychological stress and suicide ideation. Machine learning

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

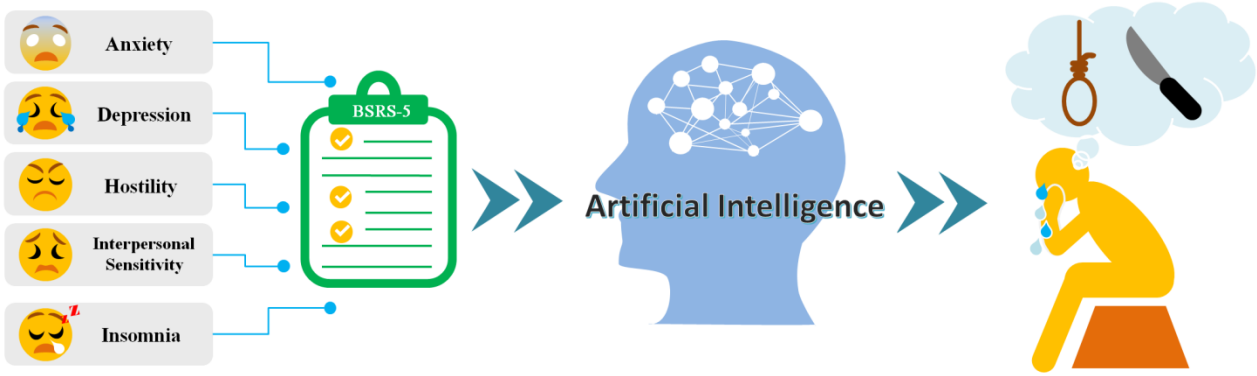


Fig. 1. Schematic diagram of proposed method

could provide better performance of the prediction of suicide ideations. In this paper, we utilize a large sample of the military members for several machine learning techniques by taking the psychological stress dimensions into consideration to predict the presence of suicide ideation.

The schematic diagram of the proposed method in this paper is illustrated in Fig. 1. A binary probabilistic classifier of machine learning algorithm can determine whether the military persons, through their questionnaires, have suicide ideations. Machine learning provides an effective manner for early warning and prevention of suicide by automatic suicide ideation detection.

The rest of this paper is organized as follows. The materials are shown in Section II. In Section III, the proposed algorithms by machine learning techniques are demonstrated. Section IV shows the experimental results. Section V concludes this paper.

II. Materials

This study used a historical cohort of 3,546 military men and women aged 18-50 years, with an average of 29 years of age from the cardiorespiratory fitness and hospitalization events in armed forces (CHIEF) study performed in the Hualien Armed Forces General Hospital, the only one military referral hospital of Eastern Taiwan, in 2014. All participants carried out a formal health examination including a blood routine, biochemical tests and chest X-ray. The participants self-reported a questionnaire for their mental status and experience of substance use as well. The study design of CHIEF study has been described in detail previously [21]-[26].

The questionnaire for the military personnel was the Brief Symptom Rating Scale (BSRS-5). BSRS-5 is composed of five psychopathological domains, i.e. anxiety, depression, hostility, interpersonal sensitivity and insomnia, in which each was measured by a five-point Likert-type scale of 0 to 4 in severity from 0, not at all, to 4, extremely, as shown in Table I. The BSRS-5 score is the sum of the five psychopathological domains scales. Generally, if the BSRS-5 score is lower than or equal to 5, the person is well adjusted. If the BSRS-5 score is within 6 to 9, the person has slight mental stress. Seeking emotional support or talking to friends or families is recommended. However, if the BSRS-5 score is higher than or equal to 10, he or she is under great mental stress. Psychological counseling and medical service are suggested. An additional questionnaire was suicide ideation, which was also measured by a five-point Likert-type scale of 0 to 4, as shown in Table I. The average scales of the BSRS-5 score and the percentage for each of the five psychopathological domains are indicated in Table II. Two outcomes are predicted in this paper. One is for the presence of suicide ideation (suicide ideation ≥ 1), which was defined as no (n =3418, class 0) by Likert-type scale=0 and as any (n =128, class 1) by Likert-type scale ≥ 1. Another case is for more severely intense suicide ideation (suicide ideation ≥ 2), which was defined as no or a little bit (n=3504, class 0) by Likert-type scale ≤ 1 and as more severely intense (n =42, class 1) by Likert-type scale ≥ 2. As shown in Table II, higher average BSRS-5 scores and higher percentages of the existence of the five psychopathological domains in those with suicide ideation present in both of two prediction cases.

TABLE I
Five-Point Likert-Type Scale of Five Psychopathological Domains of the Bsrs-5 Score and Suicide Ideation

	Not at all	A little bit	Moderately	Quite a bit	Extremely
Five-point Likert-type Scale	0	1	2	3	4

TABLE II
Statistics of the Bsrs-5 Score and the Five Psychopathological Domains

		BSRS-5	Anxiety>1	Depression>1	Hostility>1	Interpersonal Sensitivity >1	Insomnia>1
Total (N=3546)		1.91±2.80	5.30%	5.22%	7.16%	4.23%	8.21%
Suicide Ideation ≥1	*Class 0 (N=3418)*	1.66±2.42	3.89%	3.77%	5.71%	3.25%	6.82%
	Class 1 (N=128)	8.71±3.59	42.97%	43.75%	46.09%	30.47%	45.31%
Suicide Ideation ≥2	*Class 0 (N=3504)*	1.81±2.62	4.57%	4.65%	6.65%	3.80%	7.65%
	Class 1 (N=42)	10.45±3.83	66.67%	52.38%	50.00%	40.48%	54.76%

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

The matrices of the suicide ideation severity against the scales of the BSRS-5 score and each psychopathological domain are exhibited in Fig. 2. For the extreme of suicide ideation 4, there is no obviously higher scales in the BSRS-5 score and each psychopathological domain.

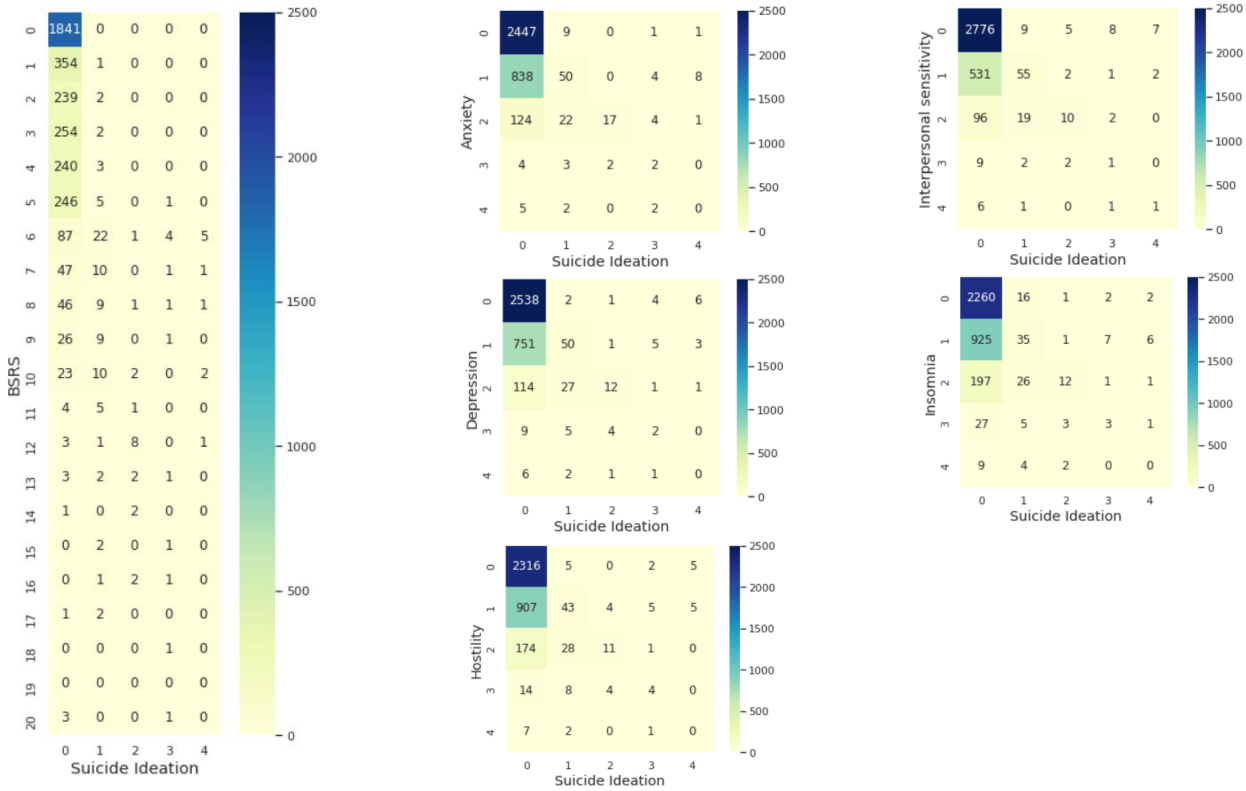


Fig. 2. Data distribution of suicide ideation vs. psychological stress

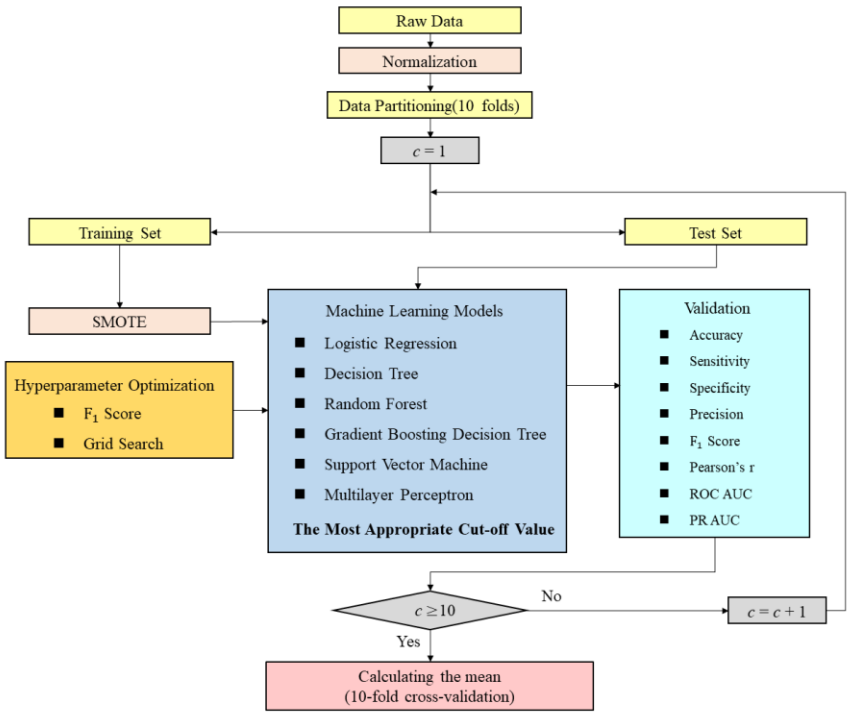


Fig. 3. The flowchart of the proposed method. The input data are pre-processed by mini-max normalization and then partitioned to training and test sets for 10-fold cross validation. The synthetic minority over-sampling technique (SMOTE) for the minority data (those with suicide ideations) in training set is performed to balance the majority data of those without suicide ideations. The six proposed machine learning algorithms are individually trained by optimizing the corresponding hyperparameters.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

III. Proposed Methods

The six input factors of psychological stress for machine learning include BSRS-5 score, anxiety, depression, hostility, interpersonal sensitivity and insomnia. This paper uses six machine learning techniques including logistic regression (LR), decision tree (DT), random forest (RF), gradient boosting decision tree (GBDT), support vector machine (SVM) and multilayer perceptron (MLP) for the prediction of the presence of suicide ideation of the military members. The system diagram of proposed method is illustrated in Fig. 3.

A. Data Pre-Processing

To solve the phenomenon of different dynamic ranges for the six input variables, we apply the normalization of Min-Max scaling [27], [28] to normalize input data into the interval 0~1. Min-Max normalization executes a linear transformation on the original data. Each of the actual data *d* of feature *x* is mapped to a normalized value adjusted in the range of 0 to 1 as Eq.(1).

$$Normalized(d) = d' = \frac{d - min(x)}{max(x) - min(x)}\ , \qquad (1)$$

where *min*(*x*) and *max*(*x*) denote the minimum and maximum values of the input feature *x*, respectively. d' represents the normalized data.

10-fold cross validation is utilized in this paper as shown in Fig. 4. The data numbers illustrated by one fold are detailed in Table III. For the prediction of the presence of any suicide ideation (suicide ideation ≥ 1), the numbers for training and test sets are 3191 (class 0: 3080, class 1: 111) and 355 (class 0: 338, class 1: 17), respectively. For the prediction of more severely intense suicide ideation (suicide ideation ≥ 2), the numbers for training and test sets are 3191 (class 0: 3157, class 1: 34) and 355 (class 0: 347, class 1: 8), respectively. This imbalance in the dataset between class 0 and class 1 is obvious. This problem is addressed by applying the synthetic minority over-sampling technique (SMOTE) [29]. The training data for class 1 are pre-processed by SMOTE to 3080 and 3157 for the two predictions, respectively, as shown in Table III.

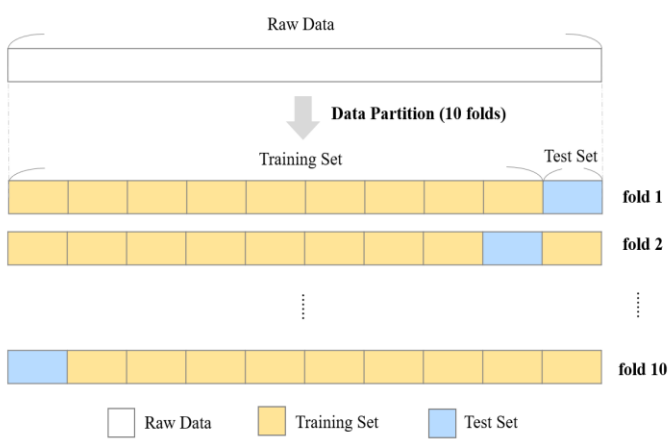


Fig. 4. Data partition of 10-fold cross validation for training and test sets

TABLE III
Numbers of Training and Test Data

		Class 0	Class 1	Total
Suicide Ideation ≥1	*Training Set*	3080	111	3191
	Pre-processed by SMOTE	3080	3080	6160
	Test Set	338	17	355
Suicide Ideation ≥2	*Training Set*	3157	34	3191
	Pre-processed by SMOTE	3157	3157	6314
	Test Set	347	8	355

B. Machine Learning Models

1. Logistic regression

Logistic regression (LR) [30], a classification algorithm used to assign observations to a discrete set of classes, is applied in our method and illustrated in Fig. 5. Logistic regression (LR) is a linear model defined as Eq. (2)-(3).

$$z = w_n x_n + w_{n-1} x_{n-1} + \cdots + w_1 x_1 + w_0\ , \qquad (2)$$

$$y = f_{sigmoid}(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}\ , \qquad (3)$$

where x_1~x_n denote the input variables and w_0 ~ w_n are the weights being learned. n is 6 for BSRS-5 and its five psychopathological domains. Logistic regression transforms its output y using the logistic sigmoid function to return a probability value.

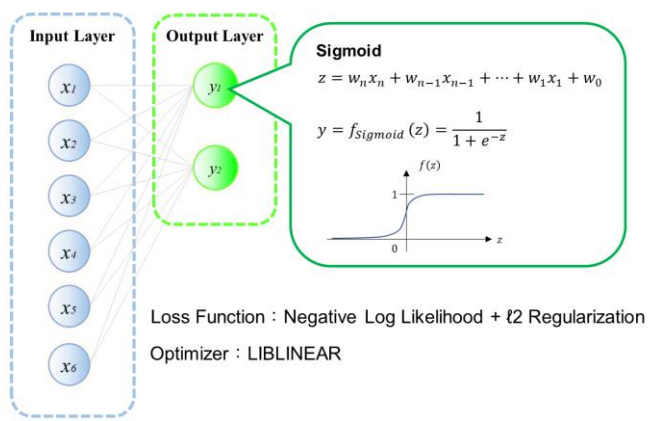


Fig. 5. Illustration of logistic regression used in the proposed method

The loss function consists of loss term and regularization term. The loss term for learning the weight vector is negative log-likelihood as Eq. (4)-(5). m denotes the sample size. X_i indicates the vector of features and Y_i means the class. The regularization term ℓ2-norm is used in our method as Eq. (6) to penalize large weights in order to prevent overfitting. The optimized hyperparameter α is chosen in our algorithm by grid search. We minimize $J_{\ell 2}(w)$ to find the optimal weight vector w by LIBLINEAR (A Library for Large Linear Classification) [31].

$$L(w) = \prod_{i=1}^{m} \mathrm{P}(Y_i = 1|X_i, w)^{Y_i}(1 - \mathrm{P}(Y_i = 1|X_i, w))^{1-Y_i}\ ,$$
$$m = 6160, 6314\ \ for\ \ Suicide\ Ideation \geq 1, 2\ , \qquad (4)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

$$J_0(w) = -\frac{1}{m}\log\big(L(w)\big)\,, \tag{5}$$

$$J_{\ell 2}(w) = J_0(w) + \alpha\sum_{i=0}^{n} {w_i}^2\,. \tag{6}$$

2. Decision tree

Decision tree (DT) [32], which establishes classification models in the form of a tree structure, with CART (Classification and Regression Tree) [33] algorithm is used in our method and shown in Fig. 6.

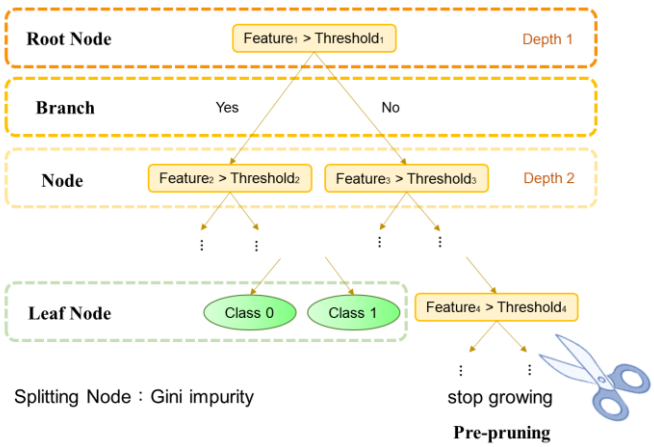


Fig. 6. Illustration of decision tree used in the proposed method

Each root node represents one of the six features and a split point on that feature. DT partitions the data (the parent node) into two subsets (the child nodes) by minimizing *Gini(T)* impurity shown as Eq. (7)-(8) to create decision points for classification task.

$$Gini(T) = \sum_{i=1}^{2}\frac{N_i}{N}gini(D_i)\,, \tag{7}$$

$$gini(D_i) = 1 - \sum_{j=1}^{2}{p_j}^2\,, \tag{8}$$

where T indicates one of the six features. N denotes the total sample number of the parent node, N_i represents the sample number falling on the i-th subset D_i (child node) and p_j indicates the percentage of each category j in D_i after T classification. The two resulting subsets change into the new parent nodes, which are subsequently separated further into two child nodes. This procedure continues until all leafs are classified. The leaf nodes of the tree contain an output prediction (class 0 or class 1). A simpler tree is built by pre-pruning processing to shorten the branches of the tree and avoid over-fitting. The maximum tree depth is the hyperparameter optimized in our algorithm.

3. Random forest

Random forest (RF) [34], an ensemble machine learning technique, constructs multiple decision trees and collects them together for classification. The training algorithm adopted in our method for random forest is the bootstrap aggregating (bagging) technique. RF builds multiple CART models with different samples and different initial variables. In each decision tree, a random subset of the features is taken into consideration for splitting a node. The individual trees are not correlated with each other and thus the trees in random forest of our method are not pruned. The final prediction result is according to the majority-votes model from the multiple DTs. RF combines the merits of feature selection and bagging. The decision tree number is the hyperparameter to be optimized.

4. Gradient boosting decision tree

Gradient boosting decision tree (GBDT) [35] is also an ensemble machine learning method and constructs multiple additive decision tree models. The DTs fitting the gradient on pseudo residuals of previous cumulative models are repeatedly trained to minimize mean squared error. This sequential stepwise manner combines the performance of weak learners (i.e., DT here) in an iterative fashion into a single strong learner to increase the accuracy of prediction. Our algorithm uses the maximum tree depth as the hyperparameter to be optimized to avoid over-fitting.

5. Support vector machine

Support vector machine (SVM) [36] with linear kernel (Linear SVM) is used for our proposed method. A data point is viewed as a 6-dimensional vector and we separate such points with a hyperplane. This linear SVM constructs the maximum-margin hyperplane so that the distance from it to the nearest training data point of any class (class 0 or class 1) is maximized.

If the training set is not linearly separable, soft-margin SVM allows the fat decision margin and some outliers are inside or on the wrong side of the margin. Our method adopts soft-margin SVM, which minimizes training error traded off against margin. Regularization strategy with a constraint by regularization term aims to fit training set data and avoid over-fitting. ℓ2-norm is utilized in SVM for our method. The regularization hyperparameter is optimized in our algorithm to control overfitting.

6. Multilayer perceptron

Multilayer perceptron (MLP) [37] consists of an input layer, hidden layers and an output layer for our algorithm as illustrated in Fig. 7. In fully connected MLP, each node in one layer connects with a certain weight to every node in the following layer as shown in Eq. (9)-(10).

$$v_i = u_{in}x_n + \cdots + u_{i1}x_1 + u_{i0}\,, \tag{9}$$

$${h_i}^k = f_{Relu}(v_i)\,, \tag{10}$$

where n represents the number of input or the number of the neuron in previous hidden layer; v represents the weighted sum of the connections. ${h_i}^k$ denotes the output of the i-th node (neuron) of the k-th hidden layer. The activation function rectifier linear unit (Relu) is used for each node in hidden layers. The output layer y is determined by the logistic regression function.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

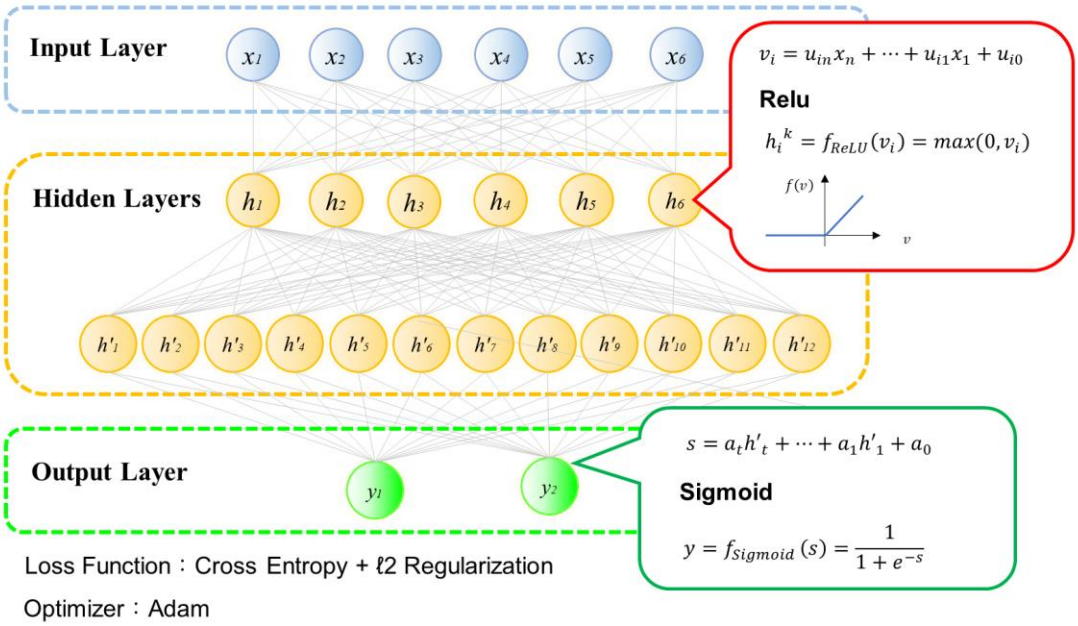


Fig. 7. Illustration of multilayer perceptron used in the proposed method

TABLE IV
HYPERPARAMETER OPTIMIZATION

Model	Hyperparameter	Beginning value	Ending value	Interval	Optimum value Suicide Ideation ≥1	Optimum value Suicide Ideation≥2
Logistic Regression	Regularization	0.01	15	0.01	14.99	9.01
Decision Tree	Maximum Tree Depth	2	15	1	14	6
Random Forest	Number of Trees	50	200	1	65	158
Gradient Boosting Decision Tree	Maximum Tree Depth	2	10	1	6	8
Support Vector Machine	Regularization	1	10	1	9	9
Multilayer Perceptron	Regularization	0.01	1	0.01	0.42	0.08
	Number of Hidden Layers	-	-	-	2	2
	Number of Neurons	-	-	-	6, 12	6, 12
	Number of Iterations	-	-	-	1000	1000

The hyperparameters that are not described in this table are set to the default values used in the Scikit-learn library.

In the forward propagation, the signal flow moves from the input layer through the hidden layers to the output layer. Learning is carried out through backward propagation. The loss function consists of cross entropy and ℓ2-norm regularization to prevent over-fitting. The optimizer Adam is adopted in our method. Besides regularization hyperparameter, the numbers of hidden layers, neurons and iterations are also used as the hyperparameters to be optimized in our MLP method.

Optimized hyperparameters for the six machine learning methods are shown in Table IV. The hyperparameter optimization is executed by grid search of F_1 score as Eq. (11). The precision and recall (sensitivity) are defined in next Section.

$$F_1\ score = \frac{2\times Precision \times Recall}{Precision + Recall}. \tag{11}$$

IV. EXPERIMENTAL RESULTS

A. Assessment Measurement

The most appropriate test cut-off values [38] for the proposed six machine learning methods are determined by F_1 score as described in Eq. (11). The performances of the applied models are evaluated by computing the diagnostic test characteristics, including accuracy, sensitivity, specificity, precision, F_1 score, r value, the area under the receiver operating characteristic (ROC) curve and the area under the precision-recall (PR) curve [39],[40].

The accuracy, sensitivity, specificity and precision defined by true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) are denoted in Eq. (12) – (15).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN}, \tag{13}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{14}$$

$$Precision = \frac{TP}{TP + FP}. \tag{15}$$

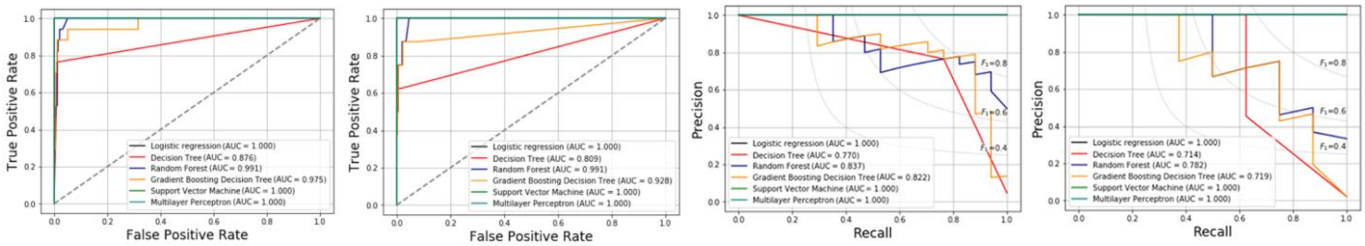
B. Discussion

The performance comparisons of each machine learning technique with the BSRS-5 score ≥7 [9] assessed by accuracy, sensitivity, specificity, precision, F_1 score, r value, are under curve(AUC) of ROC curve and the AUC of PR curve are shown

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

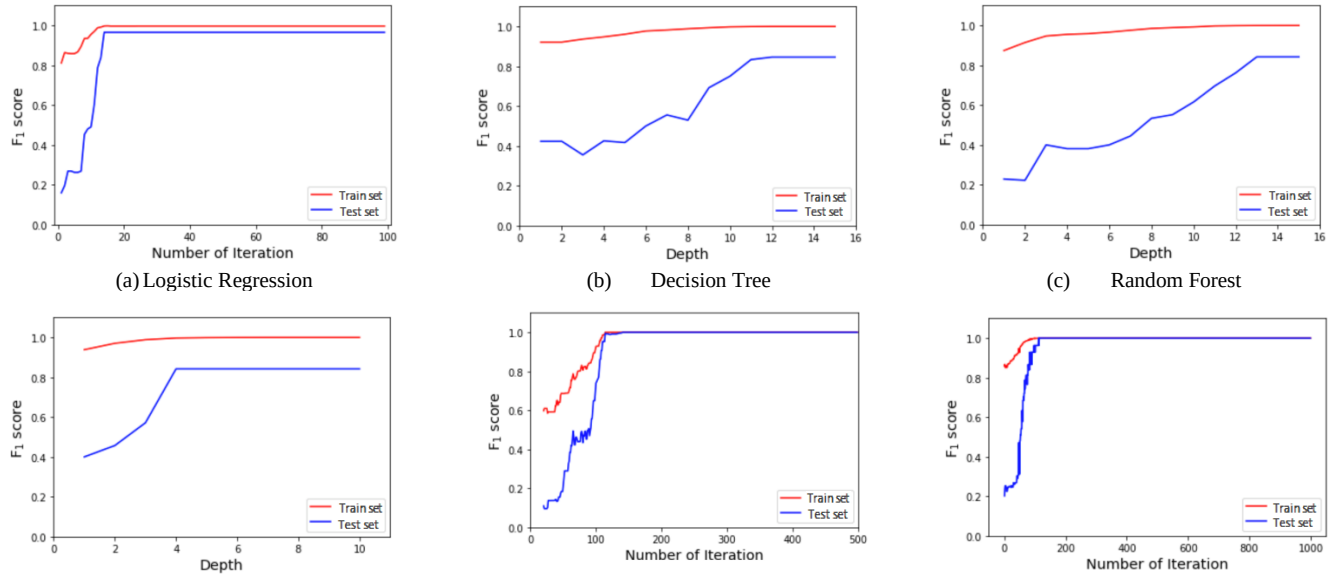
TABLE V
PERFORMANCE COMPARISONS OF SIX MACHINE LEARNING METHODS AND PREVIOUS WORK FOR THE BSRS-5 SCORE ≥7

		Accuracy	Sensitivity	Specificity	Precision	F_1 score	r	ROC AUC	PR AUC
Suicide Ideation ≥1	*Logistic Regression*	99.8%	100%	99.8%	96.3%	98.0%	81.9%	99.9%	97.9%
	Decision Tree	98.4%	77.7%	99.2%	80.7%	78.2%	71.7%	88.4%	79.6%
	Random Forest	98.7%	87.4%	99.1%	80.6%	82.8%	74.1%	97.7%	83.9%
	Gradient Boosting Decision Tree	99.0%	87.7%	99.4%	86.5%	86.6%	77.3%	98.2%	81.5%
	Support Vector Machine	100%	100%	100%	100%	100%	88.7%	100%	100%
	Multilayer Perceptron	100%	100%	100%	100%	100%	87.4%	100%	100%
	BSRS-5 Score ≥7 [9]	94.3%	64.1%	95.4%	34.8%	44.4%	42.4%	95.7%	46.8%
Suicide Ideation ≥2	*Logistic Regression*	99.9%	100%	99.9%	94.1%	96.7%	88.0%	99.9%	93.1%
	Decision Tree	98.9%	78.8%	99.2%	56.4%	61.2%	53.9%	90.4%	43.4%
	Random Forest	99.5%	90.8%	99.6%	71.3%	77.9%	69.2%	99.5%	71.6%
	Gradient Boosting Decision Tree	99.6%	90.8%	99.2%	80.1%	84.7%	69.2%	98.7%	81.3%
	Support Vector Machine	99.9%	100%	99.9%	96.6%	98.1%	87.1%	99.9%	93.3%
	Multilayer Perceptron	99.9%	100%	99.9%	96.7%	98.0%	90.8%	99.9%	97.9%
	BSRS-5 Score ≥7 [9]	93.8%	73.8%	94.1%	13.2%	22.1%	42.4%	97.1%	33.2%



(a) Suicide Ideation ≥1 (b) Suicide Ideation ≥2 (c) Suicide Ideation ≥1 (d) Suicide Ideation ≥2

Fig. 8. Receiver Operating Characteristic Curves and Precision Recall Curves



(a) Logistic Regression (b) Decision Tree (c) Random Forest

(d) Gradient Boosting Decision Tree (e) Support Vector Machine (f) Multilayer Perceptron

Fig. 9. A set of learning curves for each of the six machine learning methods

as the average of the 10-fold cross validation in Table V. The *p* values for all techniques are <0.0001. The accuracies for all of the six machine learning methods are over 98%. The results show that the MLP and SVM provide the best predictions. The second is the LR, and then the three tree-based methods. The RF and GBDT yield overall better results than that of the DT. The prediction results for suicide ideation ≥ 1 are better than those of suicide ideation ≥ 2.

The ROC curves and the PR curves and the corresponding AUCs are shown in Fig. 8, for one of the 10-fold cross validation. The areas under ROC and PR curves reach to 1 for MLP and SVM. The tendency is similar with the results shown in Table V. The learning curves for the six proposed methods for suicide ideation ≥ 1 are shown in Fig. 9. The learning curves depict an improvement in F_1 score when there are changes in the number of iteration (logistic regression, support vector machine and multilayer perceptron) or tree depth (decision tree, random forest, gradient boosting regression tree). All curves for the six proposed methods exhibit convergence well.

The distributions of variable importance for six input parameters except MLP are shown in Fig. 10. The average values of feature importance of the 10-fold cross validation are

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.2988393, IEEE Journal of Biomedical and Health Informatics

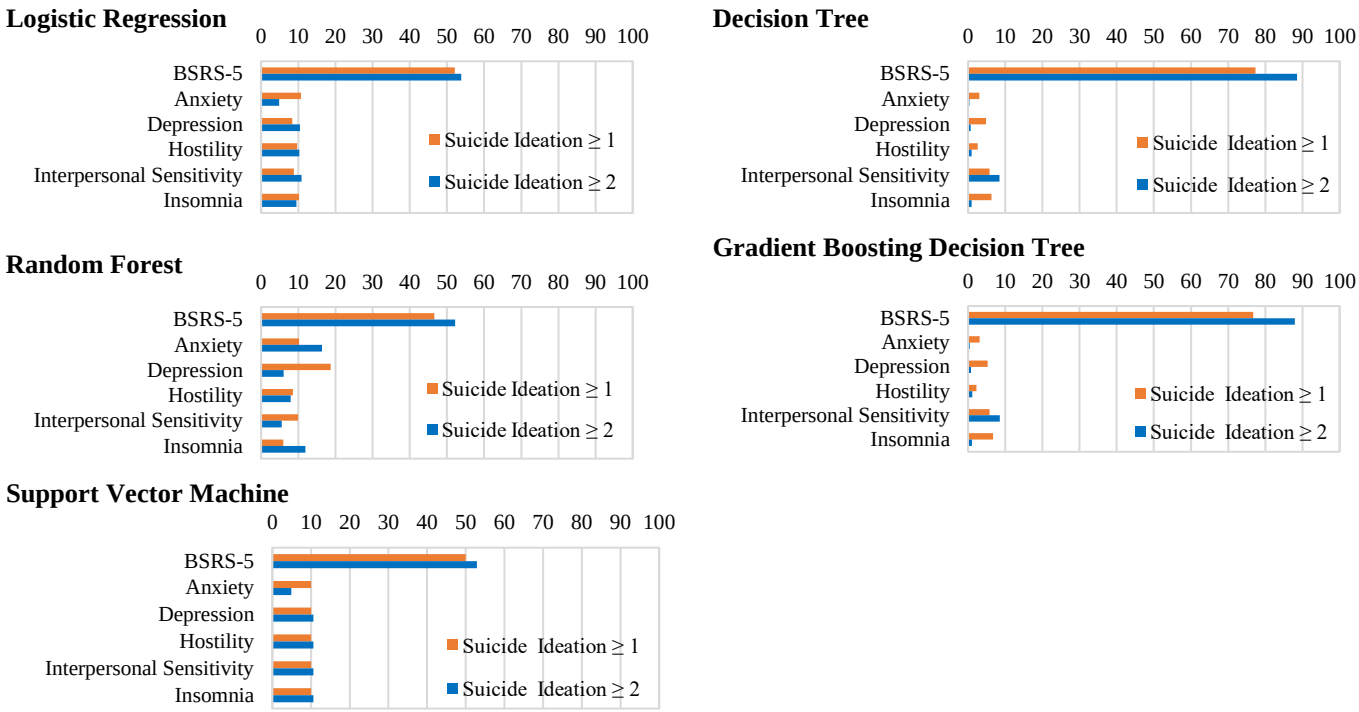


Fig. 10. Feature Importance for the machine learning methods in the proposed algorithm

TABLE VI
FEATURE IMPORTANCE FOR SIX INPUT PARAMETERS OF SIX MACHINE LEARNING METHODS

		BSRS-5	Anxiety	Depression	Hostility	Interpersonal Sensitivity	Insomnia
Suicide Ideation ≥1	*Logistic Regression*	52.15%	10.71%	8.41%	9.66%	8.83%	10.24%
	Decision Tree	77.33%	3.08%	4.88%	2.59%	5.83%	6.29%
	Random Forest	46.60%	10.19%	18.71%	8.54%	9.98%	5.98%
	Gradient Boosting Decision Tree	76.76%	3.11%	5.29%	2.25%	5.83%	6.76%
	Support Vector Machine	50.00%	10.00%	10.00%	10.00%	10.00%	10.00%
Suicide Ideation ≥2	*Logistic Regression*	53.89%	4.93%	10.44%	10.26%	10.89%	9.59%
	Decision Tree	88.54%	0.41%	0.67%	0.93%	8.46%	0.99%
	Random Forest	52.19%	16.36%	6.05%	7.95%	5.52%	11.93%
	Gradient Boosting Decision Tree	87.98%	0.47%	0.81%	1.11%	8.55%	1.08%
	Support Vector Machine	52.86%	4.86%	10.57%	10.57%	10.57%	10.57%

listed in Table VI. All of the five machine learning techniques take the BSRS-5 score as the most important variable and the proportion is even larger for the case of suicide ideation ≥ 2. For both of DT and GBDT, the anxiety dimension has less importance compared to other three methods.

For the screening instrument to predict suicide ideation proposed in [8], the sensitivities for psychiatric group, community group and general medical group are 83.76%, 21.57%, and 10.57%, respectively, and the specificities for the three groups are 72.17%, 99.49%, and 99.88%, respectively. As compared with the BSRS-5 score ≥7 in [9], a conventional criterion, for the presence of suicide ideation ≥ 1, the proposed algorithms can improve the performances of accuracy, sensitivity, specificity, precision, the AUC of ROC curve and the AUC of PR curve up to 5.7%, 35.9%, 4.6%, 65.2%, 4.3% and 53.2%, respectively; and for the presence of more severely intense suicide ideation ≥ 2, the improvements are 6.1%, 26.2%, 5.8%, 83.5%, 2.8% and 64.7%, respectively. Instead of only considering the BSRS-5 score in screen of suicide ideation like [8] and [9], our algorithm additionally takes the five psychopathological domains which are related to the BSRS-5 score, i.e., anxiety, depression, hostility, interpersonal sensitivity and insomnia as the input variables, and our schemes incorporating machine learning techniques provide better results than those of [8] and [9].

In addition, we add several critical physiological data including age, sex, body height, body weight, waist circumference, heart rate, systolic blood pressure, diastolic blood pressure and physical activity on the initial inputs of the BSRS-5 score and related five psychopathological domains in our proposed models. We find that the performances are only improved by incorporating these nine physiological data into the model of logistic regression. All of the performances of logistic regression regarding accuracy, sensitivity, specificity and precision reach 100% for suicide ideation ≥ 1 and 99.9% for suicide ideation ≥ 2. However, for the other five machine learning methods, the performances are not getting better with additional inputs of these physiological data.

As the incidence of suicide attempts is relatively low, a meta-analysis reveals that the utility of suicide ideation for predicting later suicide is limited by low positive predictive value and modest sensitivity [41]. Machine learning techniques for the BSRS-score and related five psychopathological domains can be aimed for later suicide attempts in future work.

V. Conclusion

Our study uses machine learning techniques for several psychological stress dimensions training on the prediction of suicide ideation. This paper utilizes six machine learning techniques to predict the presence of any or more severely intense suicide ideation of military personnel. Normalization of input data and imbalanced classification strategy facilitate the prediction of machine learning methods. The experimental results show that the techniques of MLP and SVM provide the best performance for the two predictions. Most of the proposed machine learning techniques take the BSRS-5 score to be the most pivotal variable and the five psychopathological dimensions are also adopted as the features to improve the screening. As compared with the prior study using the BSRS-5 score only, the machine learning techniques can improve the performances of predicting suicide ideation. This work can substantially help to screen out the military personnel at high suicide risk for suicide prevention.

References

[1] A. M. Gadermann, C. C. Engel, J. A. Naifeh, M. K. Nock, M. Petukhova, P. N. Santiago, W. Benjamin, and A. M. Zaslavsky, and R. C. Kessler, "Prevalence of DSM-IV major depression among U.S. military personnel: meta-analysis and simulation," *Military Medicine*, vol. 177, no. 8, pp.47-59, August 2012.

[2] L. K. Richardson, B. C. Frueh, and R. Acierno, "Prevalence estimates of combat-related PTSD: critical review," *Australian and New Zealand Journal of Psychiatry*, vol. 44, no. 1, pp.4-19, January 2010.

[3] W. Berger, E. S. F. Coutinho, I. Figueira, C. Marques-Portella, M. P. Luz, T. C. Neylan, C. R. Marmar, and M. V. Mendlowiz, "Rescuers at risk: a systematic review and meta-regression analysis of the worldwide current prevalence and correlates of PTSD in rescue workers," *Social Psychiatry and Psychiatric Epidemiology*, vol. 47, no. 6, pp.1001-1011, June 2012.

[4] N. T. Fear, M. Jones, D. Murphy, L. Hull, A. CIversen, B. Coker, L. Machell, J. Sundin, C. Woodhead, N. Jones, N. Greenberg, S. Landau, C. Dandeker, R. J. Rona, M. Hotopf, S. Wessely "What are the consequences of deployment to Iraq and Afghanistan on the mental health of the UK armed forces? A cohort study," *The Lancet*, vol. 375, no. 9728, pp.1783-1797, May 2010.

[5] N.S. Tzeng, C. K. Chen, T. S. Wang, H. A. Chang, Y. C. Kao, H. W. Yeh, W. S. Chiang, and S. Y. Huang, "Forensic psychiatric evaluation for military absenteeism in Taiwan," *Journal of the American Academy of Psychiatry and the Law*, vol.44, pp.352-358, September 2016.

[6] K. E. Bachynski, M. Canham-Chervak, S. A. Black, E. O. Dada, A. M. Millikan, and B. H. Jones, "Mental health risk factors for suicides in US Army, 2007–2008," *Injury Prevention*, vol. 18, pp.405-412, December 2012.

[7] K. L. Zuromski, S. L. Bernecker, P. M. Gutierrez, T. E. Joiner, A. J. King, H. Liu, J. A. Naifeh, M. K. Nock, N. A. Sampson, A. M. Zaslavsky, M. B. Stein, R. J. Ursano, and R. C. Kessler, "Assessment of a risk index for suicide attempts among US Army soldiers with suicidal ideation," *JAMA Network Open*, vol.2, no.3, e190766, March 2019.

[8] F.W. Lung and M.B. Lee, "The five-tem brief symptom rating scale as a suicidal ideation screening instrument for psychiatric inpatients and community residents," *BMC Psychiatry*, vol.8, no.53, July 2008.

[9] C. C. Ma, and Y. M. Tai, "Cut-off values of five-item brief symptom rating scale in evaluating suicidality among military recruits," *Taiwanese Journal of Psychiatry*, vol. 28, no. 2, pp.109-120, June 2014.

[10] L. Han, S. Luo, J. Yu, L. Pan, and S. Chen, "Rule extraction from support vector machines using ensemble learning approach: An application for diagnosis of diabetes," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 728–734, March 2015.

[11] M. Shahin, B. Ahmed, S. T. B. Hamida, F. L. Mulaffer, M. Glos, and T. Penzel, "Deep learning and insomnia: assisting clinicians with their diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1546 - 1553, November 2017.

[12] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 1, pp. 173-183, January 2018.

[13] B. Lei, Peng Yang, Y. Zhuo, F. Zhou, D. Ni, S. Chen, X. Xiao, and T. Wang, "Neuroimaging retrieval via adaptive ensemble manifold learning for brain disease diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol.23, no.4, pp.1661-1673, July 2019.

[14] G. M. Lin, M. J. Chen, C. H. Yeh, Y. Y. Lin, H. Y. Kuo, M. H. Lin, and M. C. Chen, S. D. Lin, Y. Gao, A. Ran, and C. Y. Cheung, "Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy," *Journal of Ophthalmology*, vol. 2018, Article ID 2159702, September 2018.

[15] X. Du, R. Tang, S. Yin, Y. Zhang, and S. Li, "Direct segmentation-based full quantification for left ventricle via deep multi-task regression learning network," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 942-948, May 2019.

[16] A. Akay and H. Hess, "Deep learning: current and emerging applications in medicine and technology," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 906-920, May 2019.

[17] M. J. Chen, P. H. Yang, M. T. Hsieh, C. H. Yeh, C. H. Huang, C. M. Yang, and G. M. Lin, "Machine learning of PM2.5 and PM10 concentrations to relate with outpatient visits for upper respiratory tract infections in Taiwan: a nationwide analysis," *World Journal of Clinical Cases*, vol.6, no.8, pp.200-206, August 2018.

[18] L. Feng, Z. Li and Y. Wang, "VLSI design of SVM-based seizure detection system with on-chip learning capability," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 171-181, February 2018.

[19] J. P. Dominguez-Morales, A. F. Jimenez-Fernandez, M. J. Dominguez-Morales, and G. Jimenez-Moreno, "Deep neural networks for the recognition and classification of heart murmurs using neuromorphic auditory sensors," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 24–34, February 2018.

[20] B. Ambale-Venkatesh, X. Yang, C. O. Wu, K. Liu, W. G. Hundley, R. McClelland, A. S. Gomes, A. R. Folsom, S. Shea, E. Guallar, D. A. Bluemke, and J. A. C. Lima, "Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis," *Circulation Research*, vol. 121, no. 9, pp. 1092-1101, October 2017.

[21] K. W. Chen, F. C. Meng, Y. L. Shih, F. Y. Su, Y. P. Lin, F. Lin, J. W. Lin, W. K. Chang, C. J. Lee, Y. H. Li, C. B. Hsieh, and G. M. Lin, "Sex-Specific association between metabolic abnormalities and elevated alanine aminotransferase levels in a military cohort: The CHIEF study," *International Journal of Environmental Research and Public Health*, vol. 15, no. 3, pp. 545, March 2018.

[22] Y. J. Chen, K. W. Chen, Y. L. Shih, F. Y. Su, Y. P. Lin, F. C. Meng, F. Lin, Y. S. Yu, C. L. Han, C. H. Wang, J. W. Lin, T. Y. Hsieh, Y. H. Li, and G. M. Lin, "Chronic hepatitis B, nonalcoholic steatohepatitis and physical fitness of military males: CHIEF study," *World Journal of Gastroenterology*, vol. 23, no. 25, pp. 4587-4594, July 2017.

[23] G. M. Lin, Y. H. Li, C. J. Lee, J. C. Shiang, K. H. Lin, and K. W. Chen, Y. J. Chen, C. F. Wu, B. S. Lin, Y. S. Yu, F. Lin, F. Y. Su, and C. H. Wang, "Rationale and design of the cardiorespiratory fitness and hospitalization events in armed forces study in eastern Taiwan," *World Journal of Cardiology*, vol. 8, no. 8, pp. 464-471, August 2016.

[24] K. Z. Tsai, J. W. Lin, F. Lin, F. Y. Su, Y. H. Li, Y. P. Lin, Y. K. Lin, C. L. Han, C. B. Hsieh, and G. M. Lin, "Association of betel nut chewing with exercise performance in a military male cohort: the CHIEF study," *Journal of the Royal Army Medical Corps*, vol. 164, pp. 399-404, November 2018.

[25] J. W. Lin, K. Z. Tsai, K. W. Chen, F. Y. Su, Y. H. Li, Y. P. Lin, C. L. Han, F. Lin, Y. K. Lin, C. B. Hsieh, and G. M. Lin, "Sex-specific association between serum uric acid and elevated alanine aminotransferase in a military cohort: the CHIEF study," *Endocrine, Metabolic and Immune Disorders - Drug Targets*, vol 19, no. 3, pp. 333-340, 2019.

[26] W. H. Chao, F. Y. Su, F. Lin, Y. S. Yu, and G. M. Lin, "Association of electrocardiographic left and right ventricular hypertrophy with physical fitness of military males: the CHIEF study," *European Journal of Sport Science*, vol. 19, no. 9, pp. 1214-1220, October 2019.

[27] Y. K. Jain and S. K. Bhandare, "Min max normalization-based data perturbation method for privacy protection," *International Journal of Computer and Communication Technology*, vol. 3, no. 4, pp. 45-50, 2014.

[28] Y. Suresh, L. Kumar and S. K. Rath, "Statistical and machine learning methods for software fault prediction using CK metric suite: A comparative analysis", *ISRN Software Engineering*, pp. 1-15, March 2014.
[29] N. V. Chawla, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "SMOTE: synthetic minority oversampling technique," *Journal of Artificial Intelligence Research*, vol.16, pp. 321–357, June 2002.
[30] D. W. Hosmer, and S. Lemeshow, "Applied logistic regression," *John Wiley & Sons*, 2004.
[31] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: A library for large linear classification," *Journal of Machine Learning Research*, vol. 9, pp. 1871-1874, 2008.
[32] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp.81-106, 1986.
[33] L. Breiman, "Classification and regression trees," *Routledge*, 2017.
[34] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp.5-32, October 2001.
[35] J. H. Friedman and J. J. Meulman, "Multiple additive regression trees with application in epidemiology," *in Proceedings of the 8th Biennial CDC and ADSTR Symposium on Statistical Methods Issues Associated with Complicated Designs and Data Structures*, vol. 22, no. 9, pp.1365-1381, April 2003.
[36] B. Scholkopf, "Learning with kernels: support vector machines, regularization, optimization, and beyond," *MIT Press*, 2001.
[37] J. Patterson and A. Gibson, "Deep learning: a practitioner's approach," *O'Reilly Media*, 2017.
[38] F. Habibzadeh, P. Habibzadeh, and M. Yadollahie, "On determining the most appropriate test cut-off value: The case of tests with continuous results," *Biochemia Medica*, vol. 26, no. 3, pp. 297–307, October 2016.
[39] J. Davis and M. Goadrich, "The relationship between precision recall and roc curves," *in Proceedings of International Conference of Machine Learning*, pp. 233–240, June 2006.
[40] K. Hajian-Tilaki, "Receiver operating characteristic curve analysis for medical diagnostic test evaluation," *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627-635, Spring 2013.
[41] C. M. McHugh, A. Corderoy, C. J. Ryan, I. B. Hickie, and M. M. Large, "Association between suicidal ideation and suicide: meta-analyses of odds ratios, sensitivity, specificity and positive predictive value," *BJPsych Open*, vol. 5, no. 2, pp. e18, March 2019.