Machine Learning Algorithm for Development of Enhanced Support Vector Machine Technique to Predict Stress

Dr. R. Rajesh Kanna

Received: 6 December 2019 Accepted: 4 January 2020 Published: 15 January 2020

Abstract
Stress is a common risk factor for many diseases. A correct and efficient prediction model is required to predict stress levels for targeted prevention and intervention in the personal healthcare domain. Before preventing the event of stress-related diseases, stress should be detected and managed early. However, surveys are used to evaluate an individual's stress condition with ease of measurement and requiring little time. However, anything that puts high demands on a person makes it stressful. This includes positive events such as getting married, buying a house, going to college, or receiving a promotion. Of course, not all stress is caused by external factors. Stress can also be internal or self-generated, when a person worries excessively about something that may or may not happen, or have irrational, pessimistic thoughts about life. This article aims to develop a predictive model to find the interruption of stress using an efficient way. One of the successive machine learning algorithm is SVM. This paper proposed to enhance the parameters of SVM which is used to improve the efficiency for predicting stress. This article proposed an Enhanced Support Vector Machine classifier to predict Stress. The stress dataset is downloaded from the Kaggle repository with 951 instances and 21 attributes.

Index terms— stress, classification, SVM, KNN, machine learning

1 Introduction

tress or depression may lead to mental disorders. Work pressure, working environment, traveling distance, height, weight, food habits, etc. are some of the major reasons behind building stress among the people. Many researchers had tried to predict stress interruption using machine learning techniques including Decision Tree, Naïve Bayes, Random Forest, KNN and SVM, etc.

The primary objective of the chapter is to develop an enhanced Support Vector Machine (SVM) classifier for Stress prediction.

The research work of this article implements the machine learning algorithm for predicting whether a person is interrupted by stress or not. The implementation for the stress dataset has been developed by Enhanced Support Vector Machine, and its performance is compared with KNN and SVM.

2 II.

3 Literature Study

The below table 1 shows that the performance of existing machine learning techniques [23] to predict the accuracy. The literature study was conducted by reviewing 23 articles which were published in reputed journals . According to the existing study the highest accuracy is obtained by J48 (i.e) Decision Tree. So the proposed system concentrates on to develop a model which provides highest accuracy than the existing works.

4 III.

5 Objectives

The primary objective of the chapter is to develop an enhanced Support Vector Machine (SVM) classifier for Stress prediction. Support Vector Machine is enhanced for this research by tuning its Hyperparameters. The Hyperparameter for SVM is its kernel function. This research uses the RBF kernel function, which is used as a way of computing the dot product of two vectors x and y in some (very high dimensional) feature space.

RBF is tuned with its parameters; "Gamma" and "C' complexity parameter. "Gamma" can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. "C" parameter is used to increase the complexity level of "gamma". The accuracy level is increased when the RBF kernel is tuned with "Gamma" and "C" parameters. The concerns received from the existing study are resolved by the proposed research work(i.e) Enhanced SVM when using RBF kernel functions. Finally, the efficiency is measured by the performance obtained by the Enhanced SVM classifier.

6 IV.

7 The Research Flow for Stress Prediction

Research framework involves the steps taken to implement SVM to predict Stress through the research. This section presents the Enhanced SVM methodology used by the research work (i.e) model to predict stress. The following Figure **??**.1 shows that the methodology used in this research work. It has several steps.

The firststep is collecting the dataset. Dataset for this research work is downloaded from the Kaggle repository which contains 951 instances and 21 attributes.

The second step of the research, the dataset is applied for Data preprocessing which makes the data to be nominal values. This preprocessing work is done by using WEKA tool using by "Discretize" filter.

8 Figure 1: The Research flow for Stress Prediction

The third step is feature selection. In this step of the research is to select the subset of attributes based on certain conditions. This research uses "Correlate Attriburte Eval" from "Attribute Evaluator" and "Ranker" approach in "Search Method". At the end of this step, top ranking attributes are grouped into subset.

The fourth step is developing Enhanced SVM classifier to predict the Stress interruption. Existing SVM classifier is enhanced by tuning the RBF(Radial Basis Function) kernel function with its Hyperparameters. There are two parameters are tuned to increase the efficiency of RBF kernel function. 1. Gamma 2. C-Complexity parameter.After tuning these two parameters, SVM works efficiently than any other method performedto predict Stress interruption. After implementing the Enhanced SVM classifier, the expected output is either 'Yes-1' or 'No-0'.

Finally the performance is evaluated in terms of Accuracy, Precision, Recall and F-Measure with existing methodologies.

9 a) Data Collection

The data for the research is taken from Kaggle repository. The below table **??**.3 The above table 2 shows that the dataset which is related to Stress of working people. There are several reasons for the working people to be stressful.

10 b) Data Pre-Processing

The data set is pre-processed with a machine learning tool WEKA. In this step the data values are converted into nominal values. Dataset may contain numeric data but classifier handles only nominal values.

In that case research needs to discretize the data, which can be done with the following filters: weka.filters.supervised.attribute.Discretize

The "Discretize" filter is stored in the package "weka.filters.supervised.attribute". Here Weka is the root package for all other sub packages.

11 c) Feature Selection

In Machine Learning, feature selection also known as attribute selection or variable subset selection. It is the process of selecting a subset of relevant features for model construction. Feature selection techniques are used for the research is Feature Selection involves two steps. In the first step "Attribute Evaluator" will be chosen. In the second step suitable "Search method" will be selected for "AttributeEvaluator" to select the highly relevant attributes from the dataset. This research work uses the "Correlation Attribute Eval" approach in "Attribute Evaluator" to choose the relevant attributes for the subset. To find the relevant attributes for the subset generation "Ranker" method is chosen in the "Search Method" which gives a ranking for the correlated values. An efficient machine learning technique required only top ranking i.e. dominant attributes for prediction of stress accurately. Because, the top ranking attributes are only highly relevant attributes for predicting the class. To choose the top ranking value, "Ranker" method is tuned with "Threshold" value.

Threshold value for ranking: In ranker ”Threshold” is its property which takes number as values. Threshold value is used to select the subset of ranked attributes either from positive or negative by given its initial rank value. This research work uses threshold value is 0, which uses only positive ranked values for feature selection. The above Figure **??** shows that the list of attribute in the subset after ”Threshold” value is assigned to the ”Ranker” method. Figure **??**.2 shows that both positive and negative ranked values. To remove the negative values, set Threshold=0. It filters the attributes which are negatively ranked. Finally, out of 18 attributes from subset, only 10 attributes are chosen for new subset after applying ”Threshold” value. After completion of feature selection, the new subset will be given as input for the proposed classifier, SVM.

12 V. Enhanced Support Vector Machine for Predicting Stress

This research work is carried out to enhance SVM features for the prediction of Stress interruption accurately. To reach the objective, SVM is enhanced with RBF (Radial Basis Function) kernel function and with tuning parameters of RBF.

This research uses the RBF kernel function to map the data. RBF kernel works by mapping the data to a higher dimensional feature space using an appropriate kernel function and a maximum margin is found for separating hyperplane in feature space [15].

The accuracy problem is usually represented by the proportion of correct classifications. A soft margin can be obtained in two different ways. It is important to add a constant factor to the kernel function output whenever the given input vectors are identical.

And, the magnitude of the constant factor to be added to the kernel or the bound size of the weights controls the number of training points that the system misclassifies. The setting of this parameter depends on the specific data at hand.

To completely specify the support vector machine it requires to specify two parameters; a) the kernel function and b)the magnitude of the penalty for violating the soft margin. Hence, to improve the accuracy of SVM, the RBF kernel function is applied in this research; this is the best criterion used for achieving better results. The next section discussed the procedure for Enhanced SVM methodology. a) Enhanced SVM Algorithm Algorithm 6.2 explains the necessary steps to be followed to improve the performance of Support Vector Machine. Step 1: Collect Stress dataset S

Step 2: Pre-process the data using ”Discretize”

Step 3: Select the subset of attributes using ”CorrelationAttributeEval” and ”Ranker” method

13 C

Step 4: Eliminate the minimum ranked attributes by using ”Threshold”. Set Threshold=0

Step 5: Update the subset after eliminating minimum ranked value.

Step 4: Implement the classifier Enhanced SVM on subset

Step 5: Tune the parameters of SVM

Step 5.1: Select RBF (Radial Basis Function) kernel function

Step 5.2: Use the ”Gamma” parameter. Set ”Gamma” =1

Step 5.3: Tune the ”Gamma” by ”C ”Complexity parameter. Set C=0

Step 6: Evaluate the performance

Step 7: End This article is proposed by applying the RBF kernel function with gamma factor and complexity factor C in Support Vector Machine algorithm. This parameter tuning helps to improve the efficiency of Support Vector Machine Algorithm in proposed work.

14 b) Kernel Function

Kernel functions are used to linearly or nonlinearly map the input data to a high-dimensional space (feature space). The idea of the kernel function is to enable operations to be performed in the input space rather than the potentially high dimension feature space. Hence the inner product does not need to be evaluated in the feature space This research work chooses RBF kernel function in SVM for searching values in feature space.

The RBF kernel on two samples x and x’, represented as feature vectors in some input space, is defined as where ||x?x?||2||x?x?||2 is the squared Euclidean distance between two data points x and x?. SVM classifier using an RBF kernel has two parameters: gamma and C.

15 c) Gamma Parameter

Gamma is a parameter of the RBF kernel and can be thought of as the ’spread’ of the kernel and therefore the decision region. When gamma is low, the ’curve’ of the decision boundary is very low and thus the decision region is very broad. When gamma is high, the ’curve’ of the decision boundary is high, which creates islands of decision-boundaries around data points.

When Gamma = 0.01, low gamma like 0.01, the decision boundary is not very ’curvy’, rather it is just one big sweeping arch. When Gamma = 1.0, the big difference in curve when increase the gamma to 1. Now the decision boundary is starting to better cover the spread of the data. So, the research chooses the best Gamma parameter is 1.0 after experimenting successive incremental of ”Gamma” parameter.

16 d) C-Complexity Parameter

The C parameter in support vector machine trades off correct classification of training examples against maximization of the decision functions margin. The only thing will change by the C is the penalty for misclassification.

Larger value of C will be accepted and the decision function will be working better at classifying all training points correctly. Therefore, the complexity parameter is increased from 1 to 10 in this research work.

When C = 1, the classifier is clearly tolerant of misclassified data point. When C = 10, the classifier is highly tolerant of misclassified data point. From the above table 3, it is observed that the accuracy is increasing up to certain level of Gamma factor and Complexity parameter. The most dangerous and common effect of increasing gamma parameter is overfitting. The experiment starts from the Gamma =0.01 and the Complexity parameter C is not specified. But it is produced low accuracy and the time taken is also very low.

To increase the accuracy and also to choose misclassification values, the Complexity parameter C is applied as 10 after experimenting the C value in the research. The accuracy is 82% when "Gamma=0.01" and "C=10". It is better than when "C=0". So the research work decided to increase the "Gamma" factor for the constant "C" parameter. The highest accuracy (96%) is produced by enhanced SVM when Gamma = 1 and Complexity parameter =10.

This study also analyzed the performance of RBF Kernel with Polynomial and Linear Kernel functions by using Accuracy and Execution Time. This section implemented the parameter tuning in Enhanced Support Vector Machine, and the efficiency will be measured by evaluating its performance with existing methodology SVM and KNN.

17 VI.

18 Performance Evaluation

For experimental work, the open source Machine Learning tool WEKA is used.

The following metrics are used to evaluate the performance of proposed Machine Learning Algorithm which is discussed detail in Research Methodology.

19 Result and Discussion

Various experiments are conducted with Stress datasets to evaluate the performance of the proposed Enhanced Support Vector Algorithm. To assess the performance of the proposed algorithm, the results are compared with the earlier studies results (i.e) SVM and KNN. Figure 5 shows that precision rate in Enhanced SVM, KNN and SVM. Proposed SVM algorithm achieves better precision 93% which is higher than the other techniques KNN (90%) and SVM (90%) in the Stress data set. Figure **??** summarized the comparison of all the performance metrics, which is used in stress dataset. Among the different category machine learning algorithms, Enhanced SVM produces better results when compared to exiting machine learning algorithms such as SVM and KNN.

20 VIII.

21 Conclusion

In this research, an Enhanced SVM which improves the efficiency of the machine learning algorithm to prediction of Stress. The performance of enhanced SVM is compared with the existing SVM and KNN method. Those techniques are studied and evaluated using Stress dataset. It has been analyzed that tuning the RBF kernel with Gamma and Complexity parameter, Enhanced SVM can outperform than KNN and earlier works. Proposed SVM algorithm achieves better accuracy i.e. 96% when compared to other techniques like KNN(91%) and SVM (92%) in the Stress data set with minimum execution time. This research work also recommends that the significantly evaluated classifier Enhanced SVM can be used for real-time prediction of stress and early-stage heart failure can be avoided. However, more training data whether from hospitals or from domain-experts can be added for increasing the prediction performance of the classifiers. [1] [2]

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

2

Figure 1: Figure 2 :

[1] ( ) C © 2020 Global Journals

[2] © 2020 Global Journals

1

Classifier	Accuracy	Precision	Recall
Bayes Net	88.59%	0.824	0.834
Multilayer perceptron	85.43%	0.836	0.867
Naive Bayes	84.2105%	0.717	0.890
Logistic regression	84.9649%	0.824	0.838
J48	86.42%	0.871	0.879
Random Forest	83.333%	0.833	0.825

Figure 2: Table 1 :

2

Feature Selection
1. Attribute Evaluator: CorrelationAttributeEval 2. Search Method: Ranker

Preprocessing Discretization	Enhanced SVM Classifier Kernel Function: RBF kernel Paramete Complexity parameter Performance Evaluation		
Stress Dataset			
Accuracy	Precision	Recall	F-measure

Figure 3: Table 2 :

3

S. No.	Gamma value	Complexity parameter	Accuracy	Execution Time (in seconds)
1	2	10	92.76	0.98
2	1	10	96.33	0.33
3	0.9	10	91	0.30
4	0.07	10	90.1	0.28
5	0.05	10	88.19	0.21
6	0.01	10	82.13	0.17
7	0.01	1	62.01	0.16

Figure 4: Table 3 :

4

Kernel function	Accuracy (%)	Execution Time (in seconds)
RBF Kernel	96.33	0.33
Polynomial Kernel	91.69	0.71
Linear Kernel	85	0.323

It is observed from the above table 4
that SVM with RBF kernel performance is higher than
that of the polynomial kernel and linear kernel in
prediction of stress. The SVM with RBF kernel produced
96% accuracy compared to the polynomial kernel.

Figure 5: Table 4 :

5

S.No.	Stress dataset Techniques	Accuracy	Precision	Recall
1	Enhanced SVM	96.33%	92.63%	90.26%
2	SVM	91.69%	89.96%	88.25%
3	KNN	90.78%	89.68%	87.21%

Figure 6: Table 5 :

[IEEEInternational Symposium on IT in Medicine and Education] , 978-1-4244-2511-2/08©2008 Crown. *IEEEInternational Symposium on IT in Medicine and Education*

[Abdullah and Rajalaxmi ()] 'A Data mining Model for Predicting the Coronary Heart Disease Using Random Forest Classifier'. Sheik Abdullah , Rajalaxmi . *International Journal of Computer Applications'* 2019. p. .

[Abdallahkassem and Hamad ()] 'A Smart Device for the Detection of Heart Abnormality using R-R Interval'. Mustapha Abdallahkassem , Hamad . *28th IEEE International conference on Microelectronics(ICM)*, 2016. (Chady El Moucary and ElieFayad)

[Dursundelen et al. ()] 'An analytic approach to better understanding and management of coronary surgeries'. Asil Dursundelen , Leman Oztekin , Tomak . *Decision Support Systems* 2012. 52 p. .

[Vijiyarani] 'An Efficient Classification Tree Technique for Heart Disease Prediction'. S Vijiyarani . *International Conference on Research Trends in Computer Technologies (ICRTCT -2013) Proceedings published in International Journal of Computer Applications*, IJCA. p. .

[Noh et al. ()] *Associative Classification Approach for Diagnosing Cardiovascular Disease*, Kiyong Noh , Heongyu Lee , Ho-Sun Shon , Ju Bum , Keun Lee , Ho Ryu . 2006. Springer. 345 p. .

[Gjoreski et al. ()] 'Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers'. Martin Gjoreski , Anton Gradis ?ek , Matjaz? Gams , Monika Simjanoska , Ana Peterlin , Gregorpoglajen . *13th International IEEE Conference on Intelligent Environments*, 2017.

[Kumar Yadav and Agarwal ()] 'Clustering of Lung Cancer Data Using Foggy K-Means'. Akhilesh Kumar Yadav , Divyatomar , Sonali Agarwal . *International Conference on Recent Trends in Information Technology (ICRTIT)*, 2013. 21 p. .

[Olson and Delen ()] 'Comparative analysis of data mining methods for bankruptcy prediction'. David L Olson , Dursun Delen , Yanyanmeng . *Decision Support Systems* 2012. 52 p. .

[Sayali et al. ()] 'Comparative Study of KNN, Naive Bayes and Decision Tree Classification Techniques'. D Sayali , H P Jadhav , Channe . ID: NOV153131. *International Journal of Science and Research* 2016. 5 (1) .

[Guru and Dahiya (2007)] 'Decision Support System for Heart Disease Diagnosis Using Neural Network'. Niti Guru , Anil Dahiya . *Delhi Business Review* January -June 2007. 8 (1) .

[Reshma and Kinariwala] 'Detection and Analysis of Stress using Machine Learning Techniques'. Supriya Reshma , Kinariwala . *International Journal of Engineering and Advanced Technology (IJEAT)* 2249 -8958. (9) .

[Balasundar et al. ()] 'Development of a Data Clustering Algorithm for Predicting Heart'. V Balasundar , T Devi , N Saravan . *International Journal of Computer Applications* 2012. 48 p. .

[Tahiramahboob and Bazelahghaffar ()] 'Evaluating Ensemble Prediction of Coronary Heart Disease using Receiver Operating Characteristics'. Ridairfan Tahiramahboob , Bazelahghaffar . *IEEE Internet Technologies and Application*, 2017.

[Ketutagungenriko et al. (2018)] 'Heart Disease Diagnosis System with k-Nearest Neighbors Method Using Real Clinical Medical Records'. Muhammad Ketutagungenriko , Dadanggunawan Suryanegara , Al . *4th International Conference*, June 2018.

[Sai et al. ()] 'Heart Disease Prediction Using ANN Algorithm in Data Mining'. P Sai , Chandrasekhar Reddy , Jaya Puneetpalagi . *IJCSMC* 2016. 6 p. .

[Chaitrali et al. (2012)] 'Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques'. S Chaitrali , Sulabha S Dangare , Apte . *International Journal of Computer Applications* 0975 888. June 2012. 47 (10) .

[Lathaparthiban and Subramanian ()] 'Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm'. R Lathaparthiban , Subramanian . *International Journal of Biological, Biomedical and Medical Sciences* 2008. 3 (3) .

[Parthiban and Subramanian ()] 'Intelligent heart disease prediction system using CANFIS and genetic algorithm'. Latha Parthiban , R Subramanian . *International Journal of Biological, Biomedical and Medical Sciences* 2008. 3 (3) .

[Palaniappan (2008)] 'Intelligent Heart Disease Prediction System Using Data Mining Techniques'. Sellappan Palaniappan , Rafiahawang . *IJCSNS)* August 2008. 8 (8) .

[Wang ()] 'Medical Knowledge Acquisition through Data Mining'. Hai Wang . *Proceedings* 2008.

[Lee et al. (2007)] *MiningBiosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV*, Hongyu Lee , Ki Yong Noh , Keun Ho Ryu . May 2007. p. . (LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining)

[Madhurapatil et al. (2019)] 'Prediction and Analysis of Heart Disease Using SVM Algorithm'. Rima Madhurapatil , Jadhav , Vishakhapatil , Geetachillarge Aditibhawar . *International Journal for Research in Applied Science & Engineering Technology* Jan 2019. 7.

[Singh et al. ()] 'Prediction of Heart Diseases Using Associative Classification'. Jagdeep Singh , Amit Kamra , Harbhag Singh . *5th International Conference on Wireless Networks and Embedded System*, 2016.