Instructions to replicate ML classification results.

Machine Learning (ML) Models were implemented using WEKA software Version 3.9.

The software can be free downloaded at this link: http://www.cs.waikato.ac.nz/ml/weka/downloading.html . A complete documentation about WEKA software is available at this link: http://www.cs.waikato.ac.nz/ml/weka/ .

Classification results that are reported in the paper were obtained following these steps:

1) Open WEKA 3.9 and choose the *Explorer* application.
2) From the *Preprocess Panel* open the file "**ML training set.arff**" (data from 1628 participants). This file contains the data of the 15 features that the authors used in the paper to classify subjects as high vs low level of perceived stress (see "Data analysis mehodology" section in the manuscript). The variable named "LOW VS HIGH STRESS" is the class to predict.
3) Switch in *Classify Panel* and choose Supplied test set as *Test option*. Upload the file "**ML test set.arff**" as supplied test set.
4) From the *Classifier* list choose CostSensitiveClassifier and, within the classifier options change the costMatrix, setting it as a 2 X 2 cost matrix with the following weights:
   0.0   4.0
   1.0   0.0
   Finally, select the classifier (Logistic, SMO, Random Forest, Naïve Bayes) and run the model. The analyses were carried out using the default parameters of classifiers. The details on the default ML classifiers parameters are reported in **Annex 1**.

Annex 1: details on ML classifiers parameters

The parameters of the ML classifiers run in WEKA 3.9 software are here reported. It should be noted that these are also the default parameters automatically chosen by the software to run these algorithms.

Logistic:
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model. = 4
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size. = 100
- debug -- Output debug information to the console. = FALSE
- ridge -- Set the Ridge value in the log-likelihood. = 1.0E-8
- useConjugateGradientDescent -- Use conjugate gradient descent rather than BFGS updates; faster for problems with many parameters. = FALSE
- maxIts -- Maximum number of iterations to perform. = -1
- doNotCheckCapabilities -- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime). = FALSE

SVM (SMO):
- buildCalibrationModels -- Whether to fit calibration models to the SVM's outputs (for proper probability estimates). = FALSE
- numFolds -- The number of folds for cross-validation used to generate training data for calibration models (-1 means use training data). = -1
- randomSeed -- Random number seed for the cross-validation. = 1
- c -- The complexity parameter C. = 1.0
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model. = 2
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size. = 100
- kernel -- The kernel to use. Polykernel –C 250007 -E 1.0
- checksTurnedOff -- Turns time-consuming checks off - use with caution. = FALSE
- debug -- If set to true, classifier may output additional info to the console. = FALSE
- filterType -- Determines how/if the data will be transformed. = Normalized training data
- toleranceParameter -- The tolerance parameter (shouldn't be changed). = 0.001
- calibrator -- The calibration method to use. = Logistic
- doNotCheckCapabilities -- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime). = FALSE
- epsilon -- The epsilon for round-off error (shouldn't be changed). 1.0E-12

Naïve Bayes:
- useKernelEstimator -- Use a kernel estimator for numeric attributes rather than a normal distribution. = FALSE
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model. = 2

- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size. = 100
- debug -- If set to true, classifier may output additional info to the console. = FALSE
- displayModelInOldFormat -- Use old format for model output. The old format is better when there are many class values. The new format is better when there are fewer classes and many attributes. = FALSE
- doNotCheckCapabilities -- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime). = FALSE
- useSupervisedDiscretization -- Use supervised discretization to convert numeric attributes to nominal ones. = FALSE

Random Forest:
- seed -- The random number seed to be used. = 1
- storeOutOfBagPredictions -- Whether to store the out-of-bag predictions. = FALSE
- numExecutionSlots -- The number of execution slots (threads) to use for constructing the ensemble. = 1
- bagSizePercent -- Size of each bag, as a percentage of the training set size. = 100
- numDecimalPlaces -- The number of decimal places to be used for the output of numbers in the model. = 2
- batchSize -- The preferred number of instances to process if batch prediction is being performed. More or fewer instances may be provided, but this gives implementations a chance to specify a preferred batch size. = 100
- printClassifiers -- Print the individual classifiers in the output. = FALSE
- numIterations -- The number of iterations to be performed. = 100
- debug -- If set to true, classifier may output additional info to the console. = FALSE
- outputOutOfBagComplexityStatistics -- Whether to output complexity-based statistics when out-of-bag evaluation is performed. = FALSE
- breakTiesRandomly -- Break ties randomly when several attributes look equally good. = FALSE
- doNotCheckCapabilities -- If set, classifier capabilities are not checked before classifier is built (Use with caution to reduce runtime). = FALSE
- maxDepth -- The maximum depth of the tree, 0 for unlimited. = 0
- calcOutOfBag -- Whether the out-of-bag error is calculated. = FALSE
- numFeatures -- Sets the number of randomly chosen attributes. If 0, int(log_2(#predictors) + 1) is used. = 0