UNIVERSITY^{OF} BIRMINGHAM

Review on classification techniques used in biophysiological stress monitoring

Iqbal, Talha; Elahi, Adnan; Shahzad, Atif; Wijns, William

[Link to publication on Research at Birmingham portal](#)

Review on Classification Techniques used in Biophysiological Stress Monitoring

Talha Iqbal[1], Adnan Elahi[2], Atif Shahzad[1,3], William Wijns[1,4]

[1]Lambe Institute of Transitional Research, University of Galway, H91 TK33, Ireland
[2]Electrical and Electronics Engineering, University of Galway, H91 TK33, Ireland
[3]Centre for Systems Modelling and Quantitative Biomedicine (SMQB), University of Birmingham, Birmingham, United Kingdom
[4]CÚRAM Center for Research in Medical Devices, H91 W2TY Galway, Ireland
*Corresponding Author: Talha Iqbal (t.iqbal1@nuigalway.ie)

Abstract—**Cardiovascular activities are directly related to the response of a body in a stressed condition. Stress, based on its intensity, can be divided in two types i.e. Acute stress (short term stress) and Chronic stress (long term stress). Repeated acute stress and continuous chronic stress may play a vital role in inflammation in circulatory system and thus leads to a heart attack or to a stroke. In this study, we have reviewed commonly used machine learning classification techniques applied to different stress indicating parameters used in stress monitoring devices. These parameters include Photoplethysmograph (PPG), Electrocardiograph (ECG), Electromyograph (EMG), Galvanic Skin Response (GSR), Heart Rate Variation (HRV), skin temperature, respiratory rate, Electroencephalograph (EEG) and salivary cortisol, used in stress monitoring devices. This study also provides discussion on choosing a classifier, which depends upon number of factors other than accuracy, like number of subjects involve in an experiment, type of signals processing and computational limitations.**

Index Terms—**bio-physiological stress, classification techniques, classifiers, stress parameters, heart attack, physical stress, mental stress, stress monitoring and recognition.**

I. INTRODUCTION

Stress can be of two types, Physical Stress and Mental Stress. Physical stress is often caused by a poor diet, sleep deprivation, too much work or may be due to illness. Mental stress is triggered due to worrying about illness of a loved one, death of closed ones, retirement or money or being fired from work [1]. Generally, most of the stress comes from our daily responsibilities. Work pressure and obligations, which are mental and physical, are not always noticeable to us. In response to stress, incurred in daily life, our body automatically alters our blood pressure, respiration, heart rate, blood flow to muscles and our metabolism. The response tries to help our body to react fast, yet efficiently, to a high pressure situation [2].

Stress situation can become threat to our well-being and health, if no adjustments are made in time to cope with its effects. It is very important to realize that all the external events. It does not matter how we are perceiving those events. These events can cause stress and may cause you to feel 'being out of control'.

Fatigue, headaches, sleeping problems, digestive problems, and muscle tension are some most common effects of stress. A long term and un-managed stress may cause heart disease, high blood pressure, diabetes, and obesity [3]. Stress may also cause anxiety, restlessness and inability to focus. These mental and physical changes may effect our weight loss progress as eating junk food can be used to cope stress situation [4]. Several people gets addicted to tobacco or illegal drugs as they use it as a stress management mechanism. If stress state period goes for very long time, it does increases the chances of having heart attack, hypertension or stroke in a person [5].

According to American Psychological Association (APA) stress is linked to 6 leading causes of deaths including heart disease, depression, anxiety disorder, diabetes etc. Center for Disease Control and Prevention reported that 110 million people die every year as a direct result of stress i.e. 7 people in every 2 seconds.

We need to realize that without a proper monitoring and management of stress, the situation will get more difficult to contain. People are getting depressed, are very easily angered, and they have started to withdraw to themselves. As stress is becoming one of the movers, performance of people has declined. People thinks this economic recession will stay like this and there is no way to fight it. All these situations can be avoided if people knew how they can fight and win from the effects of stress.

Recognition of high stress state is very essential. For this purpose, one can monitor stress using physiological indicators of stress such as increasing heart rate, blood pressure, respiratory rate, sweaty hands and fast pulse rate. Unfortunately, most people, cannot recognize these physical symptoms. Such people can use stress monitoring devices that will inform them in-time about increase in their stress level and thus they will be able to control it before hand by either doing meditation or by exercise. Literature suggests that following parameters, individually or with different combinations, can be considered for stress monitoring and are discussed below:

- **PPG:** A Photoplethysmograph (PPG) is light based plethysmogram which is used to detect changes in volume of the blood running in micro-vascular system of the

body. Usually, a pulse oximeter is used to obtain PPG. A pulse oximeter illuminates skin using light emitting diode (LED) and calculates change in absorption of light through photo-diode. As flow of blood to the skin is modulated with different types of other physiological systems, thus, we can use PPG to monitor hypovolemia, breathing and conditions of other circulatory system [6]. For monitoring of the stress, PPG signal obtained from pulse oximeter is used to calculate the oxygen level within the body of the human, often called oxygen saturation. For normal person this oxygen saturation level is in-between 95 to 100%. Saturation level below 90% is considered abnormal and can lead to a clinical emergency [7]. The waveform of PPG signal differs patient to patient and to the location as well as the manner in which pulse oximeter is attached.

- **ECG:** Electrocardiography (ECG, also known as EKG) is a method used to record the electrical activities of the heart with respect to time by placing electrodes on the skin. These electrodes detects small electrical changes under the skin that are generated by depolarization and re-polarization of heart muscles in electrophysiologic pattern for each heartbeat. ECG is commonly used to detect cardiac problems. In stress monitoring application, an ECG signal is used to calculate Heart Rate Variation(HRV). At stress state, the heart rate varies significantly thus can be monitored by using HRV.

- **EMG:** Elecromyography (EMG) is a technique for recording and evaluation of the electrical activities produced by muscles of skeleton. Instrument called electromyograph is used to generate a record called electromygram. EMG is the recording of the muscle cells whenever these cells are activated electrically or neurologically. Medical abnormalities such as Polymyositis and Muscular dystrophy, recruitment order, and activation levels can be analyzed using EMG signals. Moreover, we can also analyze biomechanics of animal or human movement.

- **GSR:** Galvanic Skin Response (GSR) is used to measure skin's electrical conductance. Sympathic nervous system can be triggered through strong emotion, which results in more sweat being released by sweat glands. The signal is obtain using two electrodes attached to two fingers of same hand and often used to monitor the quality of sleep.

- **EEG:** Electrophysiological method for monitoring and recording electrical activities of the brain is called electroencephalography (EEG). It is, most of the time, non-invasive method with electrodes located on the scalp but sometimes we can use invasive electrodes. Basically, voltage fluctuations are measured through EEG that causes ionic current in the neurons of brain. These readings are analysed with respect to a period of time and diagnosis are made on spectral content or event-related potentials.

- **Respiratory Rate:** Respiration rate is defined as number of times someone takes breath in 60 seconds, mostly calculated by observing the rise and fall of the subject's chest. Measuring the respiration rate while person is sleeping is the most difficult task and cannot be done using general lab devices. Respiratory rate can also be measured from blood volume pulses. There are two ways to do so; by calculating time change in-between two successive heart beats and change in the amplitude of the blood volume. The infrared sensors on the ring takes samples at 250 Hz from arteries and capillaries of finger and thus does not disturbs sleeping subject [8]. The collected samples shows inter-beat-interval, yielding data on heart rate variability, heart rate and respiratory rate.

- **Salivary Cortisol:** Cortisol is a steroid hormone, which is produced in adrenal gland by zona fasciculata in response to stress. For a normal person, concentration of salivary cortisol ranges from 10.2 to 27.3 with $+/- 0.8 \ nmol/L$ in morning and ranges from 2.2 to 4.1 with $+/- 0.2 \ nmol/L$ in night [9]. High level of cortisol circulating the human body shows sustained stressed condition of the human and may create allostatic load. This allostatic load can cause various physical changes in the body. Changed level of cortisol can be observed in form of mood disorders, anxiety disorder, illness, stomach pain, fear and other physiological and psychological disorders [10]. Collection protocols and approved collection method of salivary cortisol are defined by Salimetrics USA [11].

Stress can be triggered by using a questionnaire or can be caused due to physio- and sociological factors. Different classification algorithms are used to recognize stress state. In this study, we will focus more on different machine learning techniques as these algorithms are more accurate and popular as well as state-of-the-art way of monitoring and recognizing the stress in humans.

II. SEARCH METHODOLOGY

Studies only using machine learning algorithms for classification of mental and physical stress with an accuracy of more than 50% were selected. All the search was done using Google Scholar, Pub Med, IEEE digital library and Cochrane library. The search terms used are as follow:

- Stress monitoring using medical devices
- "stress" monitoring wearable systems
- medical devices for "stress" monitoring
- devices to measure mental and physical stress
- Different techniques to measure physiological and mental stress and
- stress detection health devices

Search database yielded 105 studies, with no duplicate study. Studies published between 2000 to March 2019 were taken into the account for literature search. Selection of literature for inclusion in this paper was divided into three steps.

- In first step, studies related to recognition of mental and physical stress using machine learning algorithms were included by reviewing paper's title and abstract.
- In the second step, studies that did prediction of stress using machine learning techniques, directly, were selected.
- In the final step, a deep study was done to acquire information about input data, methods of measurements, classification methodology, tools and number of subjects

used for experimentation, accuracy achieved and conclusion drawn by the authors of the papers.

Most frequently used machine learning algorithms in these 105 papers are selected for discussion, which are 21 different techniques. For result comparison, in first stage, 23 studies were removed on the bases of information in their title and abstract. In second stage, after thorough studies, 58 studies were excluded due to lack of any or multiple information required, described in previous paragraph. Finally, only 24 studies met the selection criteria and are included in this review.

III. QUESTIONNAIRE METHODS

A questionnaire method is mostly used to measure mental stress. In this method, stress is triggered in the subject using some questions and recording their response time along with PPG and EEG signals. These response signals are then fed to a neural network to classify the stress states into low, normal and highly stressed states. Such assessment of the stress is often called subjective assessment. Most of the research community uses a "Stroop test" to measure the mental health of the subject. Stroop test is basically a colour naming activity, mostly designed on computer as a game. In this game, subjects are ask to call name of the colour irrespective of what is written with that colour. For example, word BLUE is written with PURPLE colour, so subject have to say PURPLE as an answer to this question. Stroop test designed by Nagananda et. al. in [12] uses five colours; BLUE, YELLOW, GREEN, RED and PURPLE and classifies stress into low, medium and high stress levels using simple neural network.

Beside stroop test, many researches designs their own test questionnaires. As Kallus et. al. [13] designed a RESTQ that measured the frequency of stress and activities related to stress recovery. The authors designed five different versions of RESTQ based on the types of subjects one want to use. RESTQ-Basic for general usage and had seven stress scales. RESTQ-Sport for athletes with five recovery scales. RESTQ-Coach for coaches, RESTQ-CA for adolescents and RESTQ-Work for subject's work context. Every version had its own time frame of three days/nights or seven days and nights. The output is indicated on the scale of 0 to 6 i.e. never to always, respectively.

Boynton et. al. [14] presented a very interesting study about selection, design and development of a self defined questionnaire. The authors argued that anyone can design a list of questions and print it but designing a well effective and generalized questionnaire needs creative imaginations and careful planning. Authors also discussed different perspectives of a questionnaire that should be taken into account while designing or developing it. A new questionnaire often fails to provide a high quality generalized data, thus, whenever possible one should use previously validated questionnaires and rephrase them appropriately for their targeted audiences and information they require. Authors concludes that a nicely explained and carefully designed questionnaire will always lead to improved response rates.

IV. MACHINE LEARNING ALGORITHMS

Recently, development of different machine learning algorithms has greatly helped to develop tools that assists doctors to support patient care and predict any mental disorders. Machine learning techniques are widely used as a decision boundary making tool in complex data analysis of health. Supervised machine learning algorithms produces general hypotheses from externally supplied labelled feature. This hypotheses is then used to make predictions about new incoming features [15]. Literature included in this review are measured on following criteria:

1) Should be frequently used classification methods.
2) Should have achieved good classification accuracy (>50%).
3) Should be about classification of physical and mental stress states.

Selection of learning algorithm is a critical step. Usually, an algorithm is evaluated on the basis of its number of correctly predicted outcomes over total number of prediction attempts (which is called prediction accuracy). There are three ways to testing a classifier. One is by splitting the training set into training and evaluation set. The ratio of split should be 70% and 30%, respectively. Cross validation is the second option to test the performance of the classifier. Here training set is split in mutually exclusive and equal-sized subsets. Classifier is then, for every subset, trained on the union of all subsets. Error rate of every subset is calculated and average error rate determines the classifier performance. There is a special type of cross validation called Leave-one-out. Computationally this method of cross validation is expensive but is used whenever we require greater accuracy in terms of error rate of a classifier. Error rate does depends upon numbers of parameters like size of training set, dimension of problem, hyper-parameters tuning and use of relative features of the problem. Third and final way to measure the performance of the classifier is statistical comparison of classifier's accuracies when trained on specific datasets [16].

Supervised machine learning algorithms can be divided into four major types; Logic-based Algorithms, Perceptron-based Methods, Statistical learning Techniques and Support Vector Machines (SVM). Each type of algorithm have different sub-learning algorithms. Some of these learning algorithms are illustrated as follows:

A. Decision Tree Classifier

Decision tree does classification by sorting input instances on the basis of feature values [17]. Each node of the decision tree shows a classified feature from an input instance while each branch shows an assumed nodal value. Classification of instances starting from root and is sorted depending upon their feature values. The best divisor of input training data becomes root node of the decision tree.

Fig. 1 shows steps involved in classification using a decision tree. This divide and conquer strategy is efficiently fast and can be an important classifier if we have hundreds of thousands of input instances. A pseduo-code for designing a decision tree is presented in [18].

All weights are updated in such a way that it brings the outcome (result of classifier) nearer to the desired output. The most popular algorithm for updating weights is known as Back Propagation (BP) algorithm and is defined as:

$$\Delta W_{ji} = \eta \, \delta_j \, O_i \qquad (1)$$

C. Bayesian Network Classifier

To represent probability relationships of input instances (features) in form of graphs, Bayesian Network is used. Structure of Bayesian Network (BN) is Directed Acyclic Graph (DAG) and there is one-to-one correspondence between its nodes. Arcs in DAG shows influence of different features on each other. Conditional independences can be detected if there is no arc representing casual influences in-between features or there is no descendants nodes from this node (feature). Typically, learning Bayesian network is two-fold task: first, learn DAG structure and how BN structure is created using DAG, then determine BN parameters. Pseudo-code for training BN is presented in [18]. Fig. 3 shows structure of general Bayesian network.
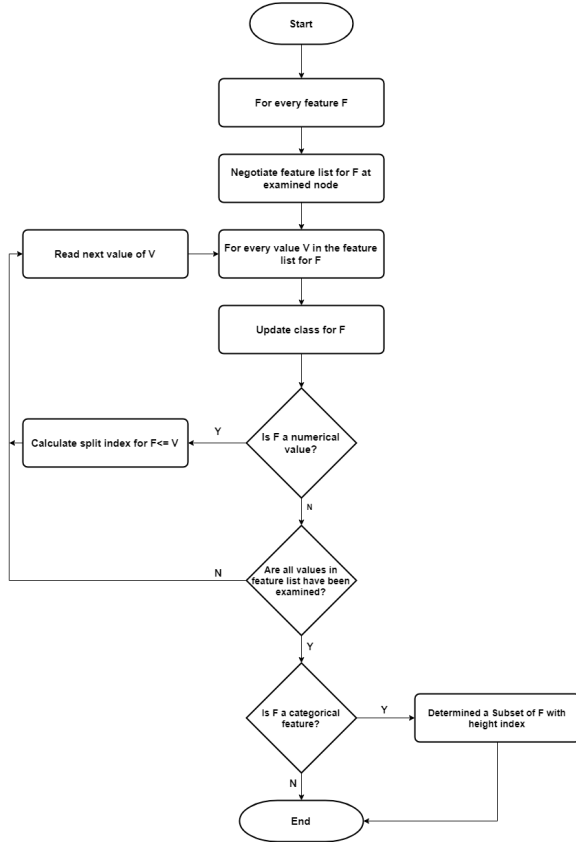

Fig. 1. Flowchart of C.45 decision tree algorithm

B. Artificial Neural Network Classifier

Artificial Neural Network (ANN) are used for classification whenever instances in training dataset can not be linearly separated. An overview of Artificial Neural Network (ANN) is provided in [22]. ANN is created by connection of many neurons (units) patterned as shown in Fig. 2.
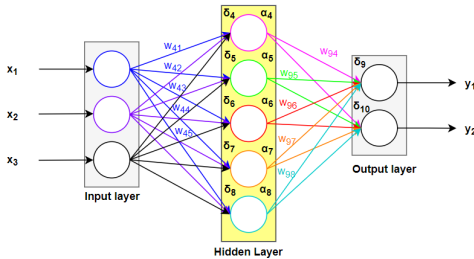

Fig. 2. Feed forward ANN

Neurons of the network are divided into three layers : input layer, which receives incoming information from training dataset; output layer, that is gives us processed result (most of the time probabilities); and hidden layer, which is in-between input and output layer. If there is one-way communication between neurons of the network i.e. only from input to output, then the network is called feed forward network. Outcome of an ANN depends upon three factors: architecture of the network, weights associated with each neuron in the network, and input along with activation functions used for each neuron.

Fig. 3. Structure of BN

D. Naive Bayesian Classifier

Naive Bayesian classifier is very simpler form of Bayesian network. Naive Bayesian (NB) has only one parent node in its DAGs, which is an unobserved node, and have many children nodes, representing observed nodes (Fig. 4). NB works with a strong assumption that all the child nodes are independent from its parent node and thus, one may say that Naive Bayesian classifier is a type of estimator. Mathematically,

$$R = \frac{P(i|X)}{P(j|X)} = \frac{P(i)P(X|i)}{P(j)P(X|j)} = \frac{P(i)\prod P(X_r|i)}{P(j)\prod P(X_r|j)} \qquad (2)$$

From equation 2, one can conclude that larger probability value will indicate that class label assign to a feature (child

node) is its actual label. The threshold for classification is as; if value of $R > 1$ then predict i otherwise predict j.

Fig. 4. Structure of NB

E. k-Nearest Neighbour Classifier

k-Nearest Neighbour (kNN) is one of the simplest instance based learning algorithm. Working of kNN is as; it classifies all the close proximity instances, in a database, into a single group and then wh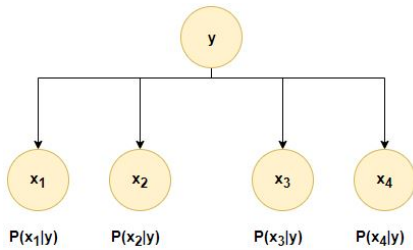en new instance (feature) comes, the classifier observes the properties of the instance and place it into the closest matched group (nearest neighbour). Fig. 5 shows flowchart for working of kNN classifier. For accurate classification, initializing a value to k is most critical step in kNN classifier.

F. Nearest Cluster Classifier

Nearest Cluster Classifier is a new classification technique proposed to reduce the training set of k-Nearest Neighbour (kNN) and enhance its performance by using clustering method, (proposed in [29]). The main goal of this method is to classify the given test samples according to their nearest neighbour tag.

This algorithm first clusters the given (training) set to a number of different partitions. After getting these partitions, different clustering algorithms are executed to eliminate large number of clusters from those partitions. Then, central label of previously produced clusters is calculated using majority vote method between patterns of class labels in the cluster. A set of most accurate clusters are used as training set of the final 1-NN classifier. In last, the class label of upcoming test sample is calculated according to class label of his nearest cluster centre [30].

G. Learning Vector Quantization Classifier

The Learning Vector Quantization (LVQ) algorithm is artificial neural network based algorithm which enable us to choose number of training features and learns how those features should look like. LVQ is more likely a collection of codebook vectors. These vectors are consists of a list of numbers which have same input as well as output features as their training set.

In form of neural network, every vector of the codebook is considered as a neuron, each feature on the codebook vector
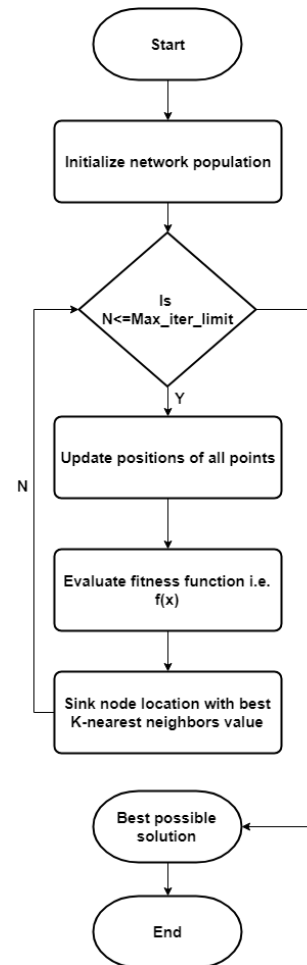
Fig. 5. Flowchart of kNN classifier

is considered weight and collection of vectors of codebook makes a network [32]. Prediction procedure of LVQ is same as used in kNN (k-Nearest Neighbour) algorithm. For prediction, all the vectors in the codebook is searched and most similar K is found. Then output is summarized for those selected K instances. By default, the value of K is considered as 1 and best matched vectors in codebook are called Best Matching Unit (BMU). Fig. 6 shows general architecture of LVQ classifier.
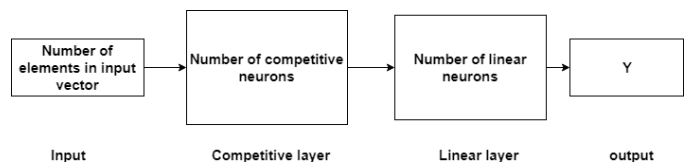
Fig. 6. LVQ network architecture

H. Kohonen Self-Organizing Map Classifier

The Kohonen Self-Organizing Map (KSOM) learning algorithm is originally presented in 1982, and uses vector quantization with similarities of patterns [34]. This algorithm

is generally used for clustering problems having very complex dataset. Sizes of topology map and learning are the two parameters that should be taken into account before designing a KSOM based classifier.

Fig. 7. Steps in original KSOM learning algorithm

Fig. 8. Result of (a) Original KSOM (b) Revised KSOM algorithm on Iris dataset

Steps used for designing KSOM network for classification is shown in Fig. 7. Complete dataset is repeatedly trained using different sized maps and a suitable size is found for accurate clu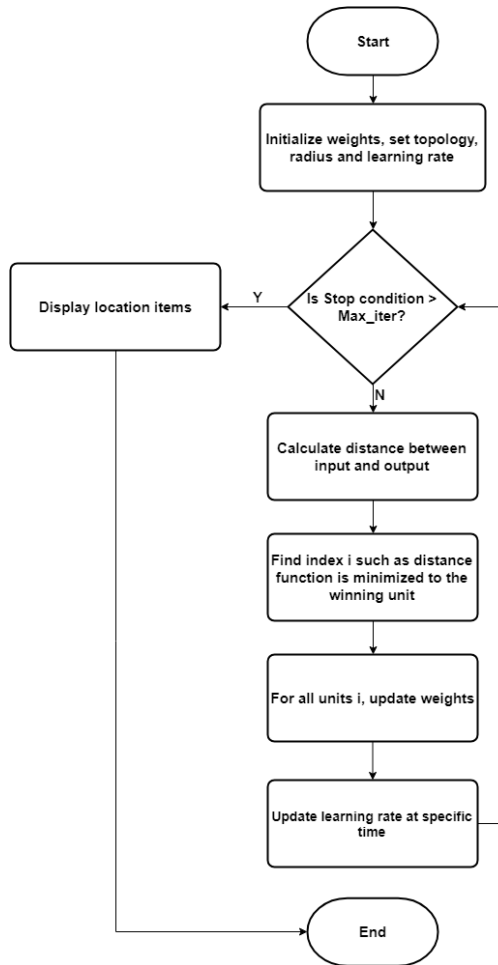ster classification. Euclidean Distance model is used to calculate distance between two nodes. Even grouping and clustering done by KSOM is quite accurate but as every clustered group lies close to another group, sometimes, leads to overlapping of cluster and problem of non-linear separation. To tackle this problem, [6] presents a variant of original KSOM classification algorithm. They used different approach for distance calculation and Fig. 8 shows result comparison of original and revised KSOM classifier.

I. Principal Component Analysis

Principle Component Analysis (PCA) is generally used to reduce dimensions of high-dimension dataset to a small subspace before using it to train any learning or classification algorithms. PCA transforms data into a new co-ordinate system having low dimension subspace. In this co-ordinate system, first axis represents first principle component that represent greatest aggregate of variance in a dataset [36].
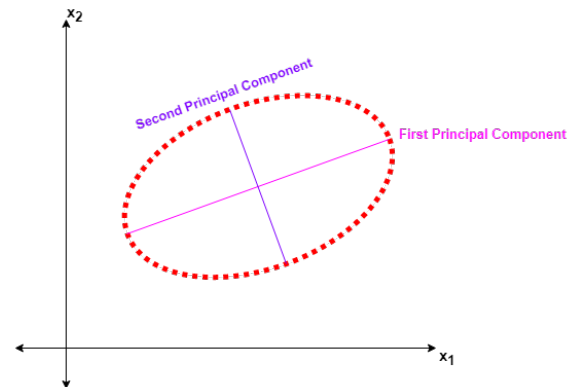
Fig. 9. Two Principle Components of dataset having two variables x_1 and x_2

From fig. 9, we can see that by calculating two principle components, we can cover the variance of whole dataset and all captured elements are independent of each other.

J. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is most commonly used dimensionality reduction algorithm. LDA is used during pre-processing step in a pattern classification and other machine learning applications.

LDA is calculated in five steps [38]; compute d-dimensional mean vectors for each class in a given dataset, compute the in-between-class scatter matrix and within-class scatter matrix, then compute Eigenvectors and their corresponding eigenvalues for both scatter matrices, select linear discriminants for the new feature subspace by sorting the eigenvectors in descending order using eigenvalues, and on final step, transforming the samples onto new subspace by simply doing matrix multiplication. Generally, LDA calculate an extra axis that shows maximum separation between the two classes, as shown in fig. 10.

K. Logistic Regression

Logistic Regression is one of the simplest machine learning algorithm mostly used for binary classification problem. This algorithm can be implemented easily and is used a baseline
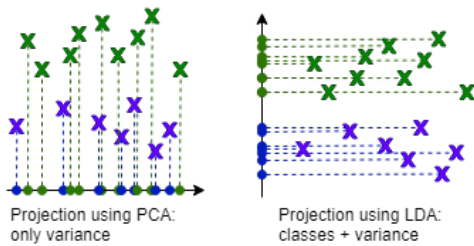
Fig. 10. Comparison of PCA and LDA

algorithm for other two-class classifiers. There are three types of logistic regression algorithms [40]:

1) If the targeted features has only two outcomes like Spam or Not spam emails or Diabetic or Not Diabatic, then this problem is solved by using binary logistic regression.
2) If the targeted features has three or more than three nominal categories, then multinominal logistic regression. For example, prediction of type of clothing.
3) If the targeted features has three or more than three ordinal categories, then ordinal logistic regression like rating any product between 1 to 5.

Logistic regression estimates and describes relationship between independent and dependent binary features within a dataset. Fig. 11 shows classification boundary calculated by logistic regression algorithm.
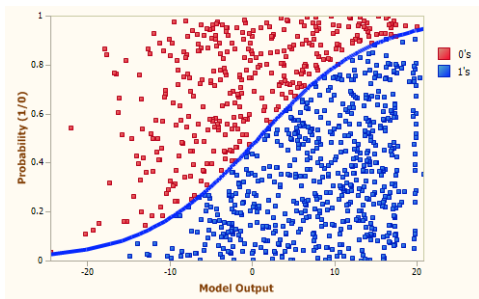

Fig. 11. Classification using Logistic Regression [41].

L. ZeroR and OneR classifier

ZeroR is the simplest classification technique, also known as Zero Rule algorithm, that relies only on target and ignores all predictors. This algorithm only predicts majority class of the given dataset. It is used for determining a baseline performance as a benchmark for other classification methods, although there is no predictability power in ZeroR. This algorithm construct a frequency table for the target and select its most frequent value [43].

OneR is abbreviation of One Rule. It is simple yet accurate classification technique which can produce one rule for every predictor present in the data. Then rule with the smallest error rate is selected and named 'one rule'. Rule for the predictor is created by construction of frequency table having two columns; target and its frequency. OneR has slightly less accuracy as compare to state of the art algorithms used for

classification but is too easy for humans to interpret. Following is the pseudo-code for designing a oneR algorithm [44]:

1) For every predictor,
2) For every value of predictor, make following rule;
   - Count appearance of each target value.
   - Find most frequently appeared class.
   - Make the rule assignment of the selected class to the predictor value.
   - Calculate error rate for each predictor.
3) Choose the predictor with smallest error rate.

M. Multi-Layer Perceptron Classifier

Multilayer perceptron are build using number of neurons connected in different layers, mostly to solve a non-linear classification problem. Each perceptron is used to categorize small linearly discrete regions of input problem while output of each perceptrons is integrated together to generate final output.

Neurons in multilayer perceptrons are organised as input layer, one or more hidden layers and the output layer, as shown in fig. 12. The learning rule for this technique is called back-propagation rule or generalised delta rule.
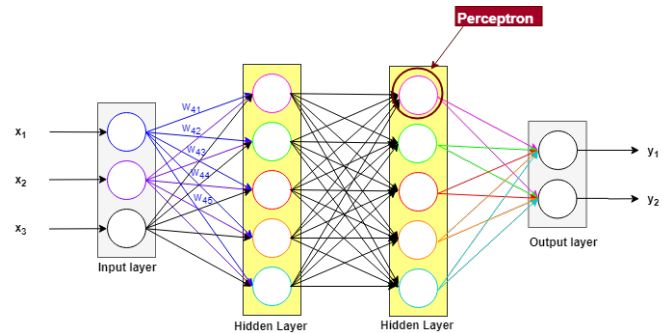
Fig. 12. Structure of MLP Classifier

N. Genetic Algorithm

Genetic Algorithm (GA) is an optimization technique that follows principles of natural selection and genetics. This technique is frequently used for finding an optimal or nearer optimal solution of complex problems that otherwise may take a long time to get solved. Genetic algorithms basically have pool of possible solutions to a given problem. These solution undergoes mutation and recombination just like in natural genetics, making new children, and this process is recurred again and again over different generations. Every individual or candidate solution is provided with a fitness value (label) depending upon their objective function. The fittest individual is given more chance to mate and produce more fitter offspring (individual). Fig. 13 shows general steps involved in Genetic Algorithm.
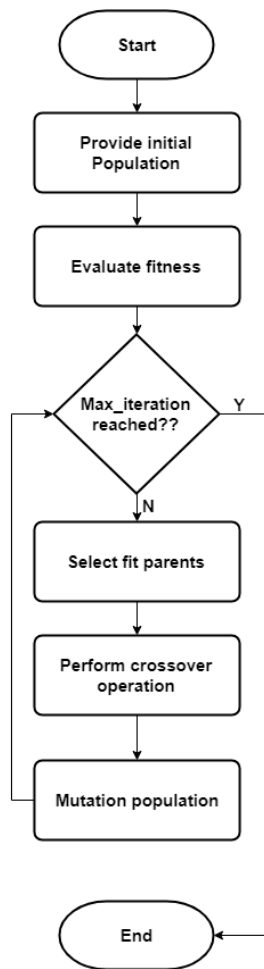
Fig. 13. Steps involved in Genetic Algorithm (GA)

O. Decision Forest

Decision Forest classification technique is a supervised learning technique. This classifier performs best if we have to predict only two target classes at the output. Decision forest is typically implemented using multiple decision tress model. Hyper-parameters of each decision tree model are tuned before training and testing of decision forest classifier. When combined, each decision tree votes for the popular output class, which is mostly used whenever we have ensemble model.

Working of decision forest is as [47]; it creates many individual tress using whole dataset, keeps starting point of each tree different (usually this selection is random), every tree in decision forest gives a non-normalized histogram of frequency labels at the output, these histograms are then summed up and normalized to get probability of each label, and at last, the tree having highest confidence value is given higher weight in the last decision of the forest.

P. Decision Jungle

Decision Jungles are extension of decision forest. A jungle is consists of ensemble decision Directed Acyclic Graph

(DAG). Hyper-parameters to be initialized are [49] resampling method; you should know how you want to create each individual tree (could either be $begging$ or $replicate$ technique), specify how to train your model; either you want to train your model using $single$ parameter or you want to train it via parameter $range$, number of decision DAGs; define maximum number of graphs which will be created, maximum depth of the decision DAGs; determine maximum depth allowed for each graphs, maximum width of the decision DAGs; determine maximum width allowed for each graphs, maximum number of iteration per decision DAGs; determine number of iterations over given dataset to build each DAG.

Decision jungle allows branches of the trees to merge thus this algorithm can be generalized. Moreover, this algorithm can create non-linear boundaries during classification and is robust to noisy feature.

Q. Random Forest

Random Forest is a supervised machine learning algorithm. This algorithm creates a random trees (forest) that are somewhat similar to decision trees and training method selected is always $begging$, as in $begging$ we combine learning models linearly to increase the overall accuracy. While growing new trees, random forest adds more randomness to the existing model. Instead of finding most important target feature for node splitting, this algorithm searches for best feature in the random subset of target features. In this way, we get wide diversity which in-return results in better model. So, as random forest only consider a random subset of features for splitting a node, we can make the trees of the model more random by using random thresholding of every feature rather than looking for the best threshold value [50].

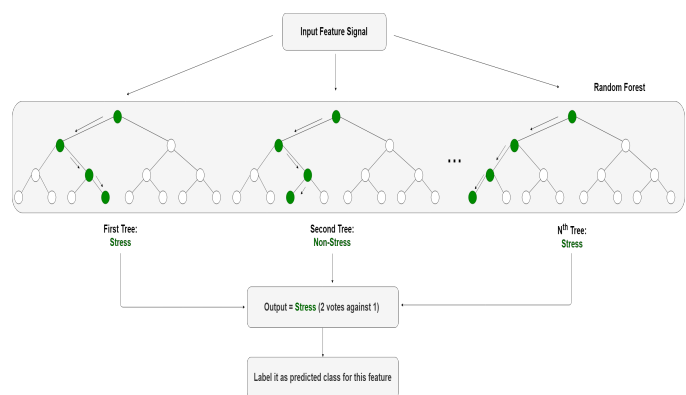Fig. 14. Simplified Random Forest Classification, classifying stress and non-stress

R. One vs All Multiclass Model

This model is used for prediction of three or more than three classes, specifically when the target outcome is dependant on categorical or continuous prediction variables. This algorithm can also be used for binary classification that needs multiple classes at the output [52]. Model is created using binary

classifiers for each of the multiple class output. Every binary classifier is assigned to individual classes as an complement to all other classes present in the model. Final prediction is done by running all the binary classifiers and selecting the prediction having highest probability score. In essence, a group of individual model is made and results are then merged at the end, to create single model which can predict all classes. Therefore, we can choose any binary classifier as basic classifier for one-vs-all model. To configure such model one should know that there are no configurable parameters of this model. So, optimization should be done in the binary classifier which is to be provided at the input for building this model.

Fig. 15. Working of AdaBoost Classifier

S. Ada-boost

Boosting refers to group of techniques that creates a strong classifier using numbers of weak classifiers. To find a weak classifier, different machine learning based algorithm having varied distribution are used. Each learning algorithm generates a new weak classification rule. This process is iterated many times and at the end a boosting algorithm is formed by combining all newly generated weak classifiers rules to make a strong rule for prediction. There are few steps that should be followed for the selection of right distribution [53]:

- Step 1: Give all the distributions to base learner and assign equal weights to every observation.
- Step 2: If first base learner gives any prediction error, then pay more attention to observations causing this prediction error. Then, apply new base learner.
- Step 3: Until base learning limit is reached or desired accuracy is achieved, keep repeating Step 2.

Ada-boost is usually used along with decision trees. Ada-boost model is created successively one after another, weight of every training instance (feature) is updated which affects the overall learning performed by next tree in the line. After the generation of first tree, for each training instance (feature), the performance of tree is weighted i.e. how much consideration the next incoming instance (feature) will get from the new generated tree up-next. After creating all the trees, prediction is made for new data.Fig. 15 shows classification done through Ada-boost technique.

T. Hidden Markov Model

Hidden Markov Model (HMM) is statistical version of Markov model and is assumed to be a Markov method having unobserved (hidden) states. In simple Markov model, observer can see the states directly and that is why Markov model have only parameters related to state transitional probabilities. On the other hand, in HMM the transitional states are not directly visible but output, which is depended on the states, is visible. As, every state has a specific probability distribution over possible output tokens, the sequence of the tokens produced by HMM gives information about the state arrangements. This phenomenon is known as pattern theory. The word hidden do not refers towards the parameters of the model. In fact, it
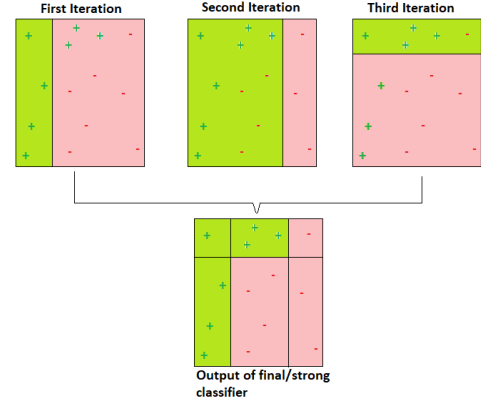
refers towards the sequence of the states through which model passes. Parameters of the Hidden Markov Model are exactly known. HMM framework contains following components [54]:

- States, e.g. labels. Usually denoted by $T = t_1, t_2, ..., t_N$.
- Observations, e.g. words. Usually denoted by $W = w_1, w_2, ..., w_N$.
- Two Special States: t_{start} and t_{end}. These states are not directly associated with observations.

The states and observation related probabilities are an initial probability distribution over states, a final probability distribution over states, a matrix with the probabilities going from one state to another state, called as transition probability, and a matrix with the probabilities of an observation generated from a state, called as emission probability. Fig. 16 shows probabilities incur in HMM.
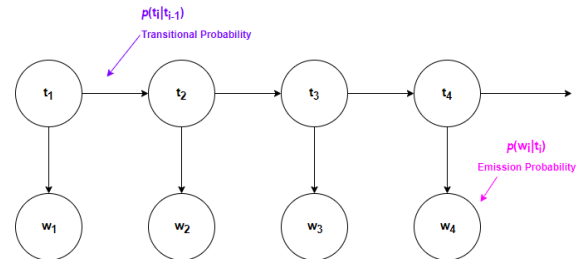
Fig. 16. Transition and Emission probabilities in Hidden Markov Model

U. Support Vector Machine Classifier

Support Vector Machine (SVM) classification technique is the most precise method of solving a classification problem. These are built around a perception of margin i.e. data is separated into two classes, on each side of the hyperplane. SVM classifier is binary classifier, so for multi-class classification problem, multiple machines are trained [56]. SVM aims maximization of margin between instances (features) of the two classes and minimizing generalization error, usually incurred in other classifiers. Fig. 17 shows how two different set of features are classified using SVM.
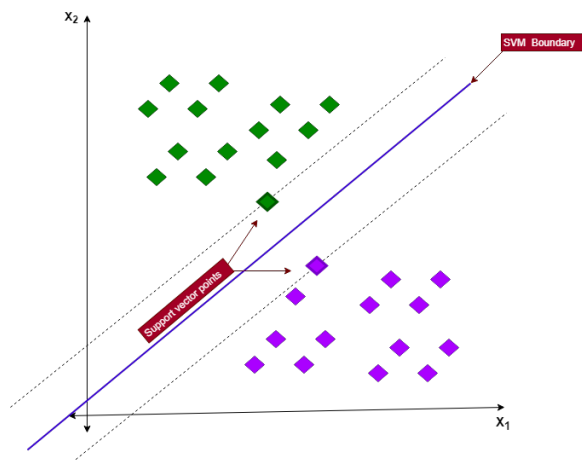
Fig. 17. Maximum margin hyperplane for SVM trained model

Data points that lie on the margin of an optimized hyperplane are called support vector point and linear combination these point form solution to the classification problem, all other points are ignored. Mathematically, SVM uses QP problem with N dimensions, where N shows training samples.

V. REVIEW OF DIFFERENT MACHINE LEARNING ALGORITHMS

A. Decision Tree Classifier

Schmidt et. al. [19] introduced a new, freely available, dataset for stress and its affect detection. They have recorded motion as well as physiological data using sensors attached to wrist and chest of the subjects. They examined 15 subjects in a lab. Their database is consist of three affective states i.e. neutral, stress, and amusement. Using Decision tree classifier for three class classification (neutral, stress, and amusement), they achieved the average accuracy of 53.9% while in the case of considering Decision tree as binary classifier (stress vs non-stress), they achieved average accuracy of 74%.

Castaldo et. al. [20], detected mental stress by using linear as well as non-linear Heart Rate Variability (HRV) features, collected from 42 students of universities, during their oral exam (that defined stress state) and after a vacation (defined as rest state). The authors concluded that decision tree (C4.5) learning algorithm classified the two states i.e. stress and rest with an accuracy of 79%.

For achieving on-line assessment of user's affective states, Barreto et. al. [21] proposed to use some of physiological signals like Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter from user to classify user's affective states by implementing some pattern learning algorithms. Authors designed a computer based "Paced Stroop Test" that acted as stimulus to any emotional stress in subject. They used Naive Bayes, Decision tree and SVM as a learning algorithms. Total of 32 subjects (students) were recruited from Florida International University and 192 feature vectors were generated (three stress; In-congruent Stroop and three non-stress; Congruent Stroop). Eleven (11) features were determined for each segmentation and only 20 samples were used as test

samples, all other samples were used to train the classifiers. Accuracy achieved with Decision tree classifier when all the 11 features were used was 88.02%. Accuracy decreased to 56.25% when authors removed feature of Pupil Diameter and was 88.54% when only Skin Temperature was removed from feature vector.

B. Artificial Neural Network Classifier

Singh et. al. [23] presented a technique based on analysis of co-relation and proposed a mathematical function that estimates stress level of auto-mobile drivers. Stress monitoring was done using physiological parameters of the driver. The result obtained showed a strong co-relation between stress function and driver's stress level. Authors used Artificial Neural Network classifier along with some simple thresholding technique (in pre-processing stage) to classify stress states as "Low Stress", "Moderate Stress" and "High Stress" depending upon different traffic conditions. For studies, they used 10 drivers and achieved an overall accuracy of 82.5% (combining ANN with thresholding technique).

Cogan et. al. [24] analysed those physiological signals that can be monitor easily via wrist wearable device. They used 20 subjects for their experimentation and tried to classify four type of stress states of an individual. These types included; Physical stress, emotional stress , cognitive stress and relaxation. The authors claims to achieve 98.8% of accurate classification of above mentioned four states by utilizing Neural Network classifier.

Verma et. al. [25] proposed a new five-layered fog cloud IoT-based student-centric monitoring of the stress to predict stress of the students in a particular context. The model computes four stressed based parameters. These parameters includes leaf node evidence, context, workload and student health trait. Authors have calculated stress indexes of 55 students and the classification technique used was Bayesian Belief Network (BNN) but for comparison they used Artificial Neural Network classifier along with some other classifiers. The accuracy they achieved via ANN classifier was 85.3%.

C. Bayesian Network Classifier

Calvo et. al. [26] used BN for stress classification and they investigated the relative effects of different parameters on the classification of mental states. These parameters included number of subjects, sampling rate, number of sessions recorded, and compared with eight different classification methods i.e. ZeroR, OneR,FT,Naive Bayes, MLP, LLR, SVM and Bayes Network. The signals choose for the study were ECG, EMG and GSR. There were only three subjects (2 Male and one Female) participated in this experiment and readings were taken in three sessions with sampling rate of 40 Hz and 1KHz. Length of each session was 24 minutes. The accuracy achieved by Bayesian Network classifier was 81.3% for 40Hz sampling rate while for 1KHz sampling rate, accuracy is not mentioned.

D. Naive Bayesian Classifier

As described in section V-C, Calvo et. al. [26] showed the effects of variation of number of subjects, sampling rate, and number of sessions recorded on eight different classification methods. ECG, EMG and GSR signals were monitored and only three subjects (2 Male and one Female) were used in this experiment. The readings were taken in three different sessions having sampling rate of $40Hz$ and $1KHz$. The accuracies achieved by Naive Bayesian classifier were 66.3% and 61.7% for sampling rate of $40Hz$ and $1KHz$, respectively.

Barreto et. al. [21], as illustrated in section V-A, proposed to utilize some of the physiological signals like Blood Volume Pulse, GSR, Skin Temperature (ST) and Pupil Diameter (PD) of the subject to classify their affective states by implementing different pattern learning algorithms. Authors designed a computer based "Paced Stroop Test" that acted as stimulus to emotional stress in the subject. They used Naive Bayes, Decision tree and SVM classifiers as a learning algorithms. Total of 32 subjects (students) were recruited from Florida International University and eleven (11) features were determined for each segmentation. Accuracy of Naive Bayes classifier with all the 11 features included was 78.65% while it was decreased to 53.65% when feature of Pupil Diameter was removed. The accuracy achieved, when only Skin Temperature features was removed from feature vector, was 82.81%.

E. k-Nearest Neighbour Classifier

Mozos et. al. [27] presented a machine learning technique for automatic detection of the social stress by combining data captured using two sensors i.e. physiological and social responses. They compared the overall performance of the model using SVM, AdaBoost and kNN. Authors concluded, from experimentation, that by combining the readings form both the sensors, one can achieve an accurate classification between neutral and stressful states during Trier Social Stress Test (TSST). Eighteen subjects were used for this experimentation and accuracy achieved with kNN classifier when $k = 3$ was 87.3%. Authors also claims that their system can be used in real-time stress detection as a wearable device and allow user to determine his/her stress state at any instance.

Schmidt et. al. [19], as described in section V-A, introduced a freely available dataset for stress and its affect detection. Authors have recorded motion as well as physiological data through different sensors attached to wrist and chest of the subjects. They investigated 15 subjects in a controlled environment. Their database was consisting of three affective states i.e. neutral, stress, and amusement. Using kNN classifier, for three class classification (neutral, stress, and amusement), authors achieved the average accuracy of 50.14% while in the case of kNN as a binary classifier (stress vs non-stress), they achieved average accuracy of 70.5%.

Cogan et. al. [24], reported earlier in section V-B, did analyses on those physiological signals which could be measured easily via wrist wearable device. Authors selected 20 subjects for their experimentation and tried to classify four type of stress states within each individual. These states were Physical stress, emotional stress , cognitive stress and relaxation. The

authors claims to achieve 97.5% of accurate classification by implementing kNN classifier.

Jebelli et. al. [28] proposed a method for automatic worker's stress detection using EEG signals. The Authors gathered EEG signals data of 11 construction field workers and preprocessed it to get high quality signals. Worker's salivary cortisol which is a stress hormone was also collected during their work and low or high stress levels were labelled. Frequency and time domain features were calculated by sliding and fixed window approach. Authors applied kNN and several SVM techniques to recognise worker's stress during their work. They achieved accuracy of 65.8% with fixed window kNN classifier while accuracy of 61.12% with sliding window kNN classifier.

F. Nearest Cluster Classifier

Setz et. al. [31] presented NCC as a stress classifier. Authors analysed the power of electrodermal activity (EDA) to distinguish stress due to cognitive load in work environment. Thirty three subjects were undergone a laboratory experimentation which included light cognitive load along with two stress factors related to workplace. Mental stress was induced by solving complex arithmetic problems under pressure of time, while psychosocial stress was induced by evaluating social threats. Experimentation was done using a wearable device that monitored EDA as measurement of stress of each individual. NCC and LDA classifiers were used to classify three phases (MIST, recovery and MIST+recovery phase) of feature combinations. The average accuracy of NCC classifier, when features were related to each other, was 73% while the achieved accuracy of NCC increased to 76.5%, when features were non-relative.

G. Learning Vector Quantization Classifier

de Vries et. al. [33] performed a large scale study on mental stress detection in an early stages. Authors measured skin conductance, electrocardiogram and respirations rate of the subjects in a semi-controlled environment. Total of 61 subjects (20 male,41 female) participated in the experiment but only 50 participants successfully completed the training, without any technical issues. These participants were divided into three groups namely; $A(AlphaTraining)$ having 18 participants, $B(randomBetaTraining)$ having 12 participants and $C(control, Musiconly)$ having 20 participants. Authors obtained an accuracy of 86.8% by deploying Learning Vector Quantization (LVQ) method to classify stressed state from relaxation.

H. Kohonen Self-Organizing Map Classifier

For stress classification, Ranjan et. al. [35] presented a comprehensive study on extraction and signal processing methodologies used on features of physiological signals. Authors adopted a Self Organising Map (SOM) approach to classify data in topographically distinct classes of low, medium and high stress states. For this study, 20 drivers were selected having vast experience of driving. The data collected was distributed into three stress states i.e. low, medium and high

stress state. Authors analysed signals obtained form GSR, PPG (Respiration rate), ECG and HRV. For classification, KSOM classifier was used and classification was done with the accuracy of 81.6%.

I. Principal Component Analysis

For stress monitoring, Mujeeb et. al. [37] assessed the ability of deep breathing for relaxation using wearable stress monitor. They developed a relationship between various mental stress activities and regular deep breathing patterns. Then, they used heart rate monitor, respiration rate monitor and electrodermal activity sensors to get those parameters which are consistent in ruling sympathetic nervous system. Total of 25 subjects (15 male, 10 female) participated in this study. For visualization of the data, Principle Component Analysis (PCA) was used and dimensions of the feature vector was reduced from six to two. After the whole data analysis, authors found that only 60% of the total subjects indicated increase in the stress level caused by improper breathing and thus, concluded that it was necessary to people training about how they can improve their breathing (deal with stress).

J. Linear Discriminant Analysis

Schmidt et. al. [19], as described in section V-A, introduced a dataset for stress and its affect detection. Authors have recorded physiological and motion related data through different sensors placed on wrist and chest of the subjects. For experiment, 15 subjects were recruited and experimentation was in a controlled environment (lab). The proposed database was consisting of three affective states i.e. neutral state, stressed state, and amusement state. Using LDA as a multi-class classifier (neutral, stress, and amusement), authors achieved the average accuracy of 62.13% while in the case of LDA as a binary classifier (stress vs non-stress), they achieved average accuracy of 78.05%.

Dobbins et. al. in [39] explored some alternative procedure for negative emotions detection that can provide better fidelity from psychophysiological data (such as heart rate) to dynamically labelled data derived from task of driving (like speed or road type). Existing systems mostly relays on labels obtained from individual self-report. Authors used machine learning techniques to generate labels, in two steps. First, they derived labels from individual self-reports and then labelled the data using psychophysiological activities (like Heart Rate (HR) and Pulse Transit Time (PTT), etc.) for generating dynamic labels of low vs high stress for every participant. Labelling techniques were evaluated using Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). Authors did two types of studies. In study 1, there were thirteen participants (seven females and six males) while study 2 included only eight participants (six females and two males). ECG, ear PPG and Accelerometer were used to collect data from the drivers. Total of 525,697,711 instances of data was collected for both the studies. The average accuracy (in terms of Area under the curve) of study 1 while using LDA classifier was 57.4%. Average accuracy (in terms of Area under the curve) of study 2 was 64.2%.

Setz et. al. [31], as explained in section V-F, inspected the power of electrodermal activity (EDA) for discrimination of stress state form neural state caused by cognitive load in work environment. Total of Thirty three subjects underwent a laboratory experiment. This experiment, included lenient cognitive load and two stress factors related to workplace. Mental stress was prompt by solving complex arithmetic problems under specified time. Psychosocial stress was induced by assessing social threats. For experimentation, a wearable device that monitored EDA as a parameter of stress was worn by each individual subject. NCC and LDA classifiers were used to classify three phases (MIST, recovery and MIST+recovery phase) of feature mesh. The achieved average accuracy of LDA classifier with related features was 76.5% while when features were non-relative the accuracy of LDA classifier jumped to 80.2%.

K. Logistic Regression

Subhani et. al. [42] proposed a Machine learning (ML) architecture based on analysis of electroencephalogram (EEG) signals of a stressed subject. The ML framework included EEG features, ROC curve, the Bhattacharya distance, t-test, classification algorithms (Navie Bayes, SVM and Logistic Regression combined) and 10-fold cross validation. Forty two subjects, who were not taking any kind of medication that increases cardiac activities, were selected for this study. Authors achieved accuracy of 94.6% for binary identification of stress while 83.4% for multi-level identification.

L. ZeroR and OneR classifier

As described in section V-C and V-D, the effects of variation of number of subjects, sampling rate, and number of sessions recorded were examined on eight different classification methods in Calvo et. al. [26]. Signals of ECG, EMG and GSR were monitored and three subjects (2 Male and one Female) were investigated in this experiment. The readings were taken in three separate sessions with sampling rate of $40Hz$ and $1KHz$. The accuracies achieved by ZeroR classifier were 12.5% for both $40Hz$ and $1KHz$ sampling rate, while accuracies of OneR classifier were 50.4% and 48.3% for $40Hz$ and $1KHz$, respectively.

M. Multi-Layer Perceptron Classifier

As described in section V-C, Calvo et. al. [26] investigated the effects of change in number of subjects, sampling rate, and number of sessions recorded on eight discreet classifiers. ECG, EMG and GSR signals were observed from only three subjects (2 Male and one Female). The readings were recorded in three different sessions with sampling rate of $40Hz$ and $1KHz$. The accuracies achieved by MLP classifier were 92.9% and 97.5% for $40Hz$ and $1KHz$ sampling rate, respectively.

Assessment of mental fitness while doing a task might be helpful to avoid unnecessary risks. Mohino-Herranz et. al. in [45] gathered information about mental stress by using two physiological signals i.e. ECG and electrical bio-impedance (TEB) signals. Total of 40 subjects were used to

create the dataset. For classification, MLP classifier is used. Authors had designed three scenarios for the measurement. First scenario was activity identification; the proposed model distinguished different kind of activities (neutral, emotional, mental and physical) with the probability error of 21.23%. Second scenario was emotional state; the proposed model distinguished different kind of emotional states like neutral, disgust and sadness. The probability error incurred was 4.8% while third scenario was mental activity; the proposed model distinguished between low mental load and mental overload with the probability error of 32.3%.

N. Genetic Algorithm

Mokhayeri et. al. [46] proposed a new stress detection method that uses GA and fuzzy SVM techniques to classify stress. They analysed Pupil diameter (PD), ECG and PPG signals using soft-computation and extracted most relevant features. These features were then optimized using Gentic Algorithm (GA) and passed to Fuzzy SVM for classification between stress and relaxed states of the subject. Dataset was composed of data collected from sixty healthy peoples (25 female and 35 male). Genetic Algorithm (GA) was used to construct a template of selected eye region using average image of 100 images.

O. Decision Forest

Akbulut et. al. [48] designed a smart wearable device that provided continuous medical monitoring and could also create a profile showing risk of disease over time. The authors named this device as Cardiovascular Disease Monitoring (CVDiMo) device. There were two groups of 30 participants in this experiment. Different systematic tests were performed on six bio-signals collected through six different sensors i.e. ECG, Body Temperature, Pulase Oximeter, GSR, Blood Pressure and Glucometer. Beside above bio-signals, physical activity and stress levels deduced from emotional state analysis was also done to increase the performance of risk estimation. By using decision forest as a multi-class classifier, the authors achieved 96% of accuracy.

P. Decision Jungle

Akbulut et. al. [48], as described in V-O, proposed a design of a smart wearable device which provided continuous medical monitoring and also creates a profile displaying disease risk over the time. Authors called this device a Cardiovascular Disease Monitoring (CVDiMo) device. Participants were distributed into two groups and total 30 people participated in this experiment. Several systematic tests were run on six bio-signals gathered form six different sensors, namely; ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure and Gluco-meter. Along with these bio-signals, physical activity and stress levels infer from emotional state analysis were also included to increase the performance of estimation of the risk. By using decision jungle as a multi-class classifier, the overall accuracy achieved was 66%.

Q. Random Forest

Schmidt et. al. [19], as described in section V-A, introduced a new and freely available dataset for stress detection. Authors recorded data samples of the test subjects in motion as well as their physiological data values using wearable sensors device attached to wrist and chest part. Total of 15 subjects participated in this experiment and was performed in a lab, in a controlled environment. The database was consist of three affective states, which were neutral, stress, and amusement state. Through Random Forest classifier, for three class classification (neutral, stress, and amusement), the authors achieved average accuracy of 58.24% while in the case of binary classification (stress vs non-stress), Random Forest gave average accuracy of 77.21%.

Oti et. al. [51] proposed an algorithm that estimates stress level of pregnant women based on Heart Rate Variation (HRV) and simple heart rate signals. Authors distributed their algorithm in edge-enabled Internet of Things (IoT) system. They used different supervised and unsupervised learning algorithms through an unlabelled dataset gathered after 7 month monitoring. For monitoring, wearable smart wristbands were given to twenty pregnant women. Authors achieved an accuracy of 97.9% by cross validating the proposed algorithm using Random Forest classifier.

R. One vs All Multiclass Model

Akbulut et. al. [48], as mentioned in V-O, demonstrated a smart wearable device that provided continuous medical health monitoring and could also create a health chart showing risk of disease over time, called as Cardiovascular Disease Monitoring (CVDiMo) device. Total of 30 participants, who were than divided into two groups, were selected for experimentation. Different routines of test were applied on six discrete bio-signals received through six different sensors including ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure and Glucometer. Beside above bio-signals, physical activity and stress levels deduced from emotional state analysis was also taken into account to increase the performance of risk estimation. By using One vs All Multi-class Model, the authors achieved average accuracy of 72%.

S. Ada-boost

Mozos et. al. [27], as described in section V-E, presented a machine learning model for automatic detection of social stress through combining data captured from physiological and social responses of the subject. They measured the performance of their model in terms of accuracies achieved from SVM, AdaBoost and kNN classifiers. Authors concluded that combining the readings of both the responses gives an accurate classification of neutral and stressful states, by conducting Trier Social Stress Test (TSST). Eighteen subjects were used for this experimentation and accuracy achieved with AdaBoost classifier was 94.3%. Authors also claimed that their system can detect stress in real-time and is a wearable device that allows user to determine his/her stress state at any instance.

Schmidt et. al. [19], as illustrated in section V-A, introduced a new dataset for stress. They recorded data of the subjects

physiological and in motion data using sensors attached to wrist and chest part. They experimented on 15 subjects in a lab. Their database is consist of three affective states i.e. neutral, stress, and amusement. AdaBoost classifier as three class classifier i.e. neutral, stress, and amusement state, achieved average accuracy of 60.43% while, in the case of, AdaBoost as binary classifier (stress vs non-stress), they achieved average accuracy of 77.2%.

T. Hidden Markov Model

Dinges et. al. [55] developed an approach that tracks changes in facial expressions of human during a long distance space-flight. Authors applied OCR algorithms to detect low and high grade stress through facial changes during critical operations. Social and workload feedback were taken from 60 healthy peoples (29 men, 31 women), to create aa database. Stress states were tracked using salivary cortisol, heart rate and self-report ratings. Authors used dataset of face videos of the subjects during different conditions for training an OCR algorithm. For classification, Hidden Markov Model was used and accuracy of classification of high grade stress from low stress state was 75 to 88%.

U. Support Vector Machine Classifier

Verma et. al. [25], as described in section V-B, presented a five-layered fog cloud IoT-based student stress monitoring system to predict stress state of the student in a certain context. The model calculated four parameters of stress, which were leaf node evidence, context, workload, and student health trait. Authors have calculated stress indexes of 55 students. Main classification technique used in the paper was Bayesian Belief Network (BNN) but authors used SVM classifier along with some other classifiers to calculate overall performance of the algorithm. The accuracy achieved through SVM classifier was 90.8%.

Barreto et. al. [21], as explained in section V-A, proposed the use of some physiological signals for example Blood Volume Pulse (BVP), GSR, Skin Temperature (ST) and Pupil Diameter (PD) from subjects to classify subject's affective states by implementing few pattern learning algorithms. Authors designed a computer based test, called "Paced Stroop Test", that acted as stimulus to any emotional stress in subject. Naive Bayes, Decision tree and SVM were used as a learning algorithms. Total of 32 subjects (which were students) were recruited from Florida International University and feature vectors of 192 features (samples) were generated (three stress; In-congruent Stroop and three non-stress; Congruent Stroop). Eleven (11) features were determined for each segmentation method. Only 20 samples were used as test samples while all other samples were used to train the classifiers. Accuracy achieved with SVM classifier when all the 11 features were used was 90.10%. Accuracy decreased to 58.85% when authors removed feature of Pupil Diameter and got 90.1% inputs, correctly predicted, when only Skin Temperature was removed from feature vector.

Calvo et. al. [26], as described in section V-C, investigated the relative effects of different specifications of classification of mental states. These specifications included number of sessions recorded, number of subjects, sampling rate, and compared with eight different classification methods which are ZeroR, OneR,FT, Naive Bayes, Bayes Net, MLP, LLR and SVM. The signals chosen for this study were ECG, EMG and GSR. There were only three subjects (2 Male and one Female) who participated in the experiment. The readings were taken in three different sessions having sampling rate of $40Hz$ and $1KHz$. Length of each session was not more than 24 minutes. The accuracies achieved by using SVM classifier was 94.6% for sampling rate of $40Hz$ and for $1KHz$ sampling rate, accuracy was 95.8%.

Jebelli et. al. [28], as explained in section V-E, presented a method for automatic detection of worker's stress using EEG signals. Authors collected EEG signals data of 11 workers, working in construction field and pre-processed it to get high quality signals. Worker's salivary cortisol, which is a stress hormone, was also collected during their work and low or high stress levels were tagged (labelled). Frequency and time domain features were intended by sliding as well as fixed window approach. Authors applied kNN and verity of SVM techniques to recognise worker's stress during their work. They achieved average accuracy of 75.9% with fixed window SVM classifier while accuracy of 67.54% with sliding window SVM.

Dobbins et. al. in [39], as illustrated in section V-J, explored some alternative procedures for negative emotions detection which can provide better fidelity using psycho-physiological data to tag data dynamically, derived from task of driving like speed or road type. Existing systems relays on tags acquired from individual self-report but authors used machine learning techniques to generate tags. Tags generation was done in two steps. First, they derived tags from individual self-reports and then they tagged the data using psycho-physiological activities (like Heart Rate (HR) and Pulse Transit Time (PTT), etc.) for generating dynamic tags of low vs high stress for every participant. Tagging techniques were evaluated using Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). Authors did two types of studies. In study 1, there were thirteen participants (seven females and six males) while study 2 included only eight participants (six females and two males). ECG, ear PPG and Accelerometer were used to gather data from drivers. Total of 525,697,711 instances of data was gathered for both studies. The average accuracy (in terms of Area under the curve) of study 1 while using SVM classifier was 58.8%. Average accuracy (in terms of Area under the curve) of study 2 was 65.8%.

Mokhayeri et. al. [46], as illustrated in section V-N, demonstrated a novel procedure for stress detection. They utilized Pupil diameter (PD), ECG and PPG signals using soft-computation and extracted most relevant features. These features were then optimized using Gentic Algorithm (GA)and passed to Fuzzy SVM for classification of stress and relaxed states of the subject. Data was collected from sixty healthy subjects (25 female and 35 male). The detection of eye region was done by constructing a template by using average image of 100 images. After the segmentation (done by GA algorithm), SVM classifier was applied to classify between stressed and

relaxation states. The achieved accuracy of SVM classifier was 90.1%.

Salafi et. al. [57] presented a study on an unobtrusive wearable stress monitoring device which had three physiological sensors namely; heart rate monitoring sensor (ECG, PPG), skin conductance (GSR) and skin temperature monitoring sensor. Total of 50 participants (23 male, 27 female) form National University of Singapore were selected to gather training data. Support Vector Machine (SVM) learning algorithm was applied to the data and accuracy of 91.26% was achieved.

Betti et. al. [58] developed and tested the ability of a wearable sensor devices to detect human stress using ECG, EEG and EDA signals. The authors also assessed co-relation of detected stress signals and salivary cortisol level, which is considered as a reliable bio-marker for stress. For experiment, 15 healthy subjects wore a set of three commercially available sensors to record readings of physiological signals when a Maastricht Acute Stress Test (MAST) was conducted. Salivary samples were also collected, at random, during the test. Statistical analysis were performed using Support Vector Machine (SVM) classifier. Classification algorithm, provided an accuracy of 86% on selected (significant) features.

VI. DISCUSSION AND CONCLUSION

In mental health research, psychological and sociological stress are considered as one of the most important as well as most complex areas of the current century. High stress does becomes thread to health of a person. The reasons behind this extensive stress are complex personal, social and ecological environment, diversity in expressing stress as well as multiple transactions of human due to his/her surroundings. Eventhough stress is a routine characteristic of life, nowadays, but if it becomes continuous and increasing, individual might show problematic symptoms which threatens their health as well as peoples in their surroundings.

This study focused on classification of stress states. Total 21 different machine learning algorithms are discussed in this review paper that uses different parameters for training and prediction of stress, see Table I. All the parameters considered, are correlated with the stress. Few of them are distinct parameters, for example GSR and HR, while some are in conjunction with other parameters to monitor and recognize stress, for example Skin temperature (ST) and EMG. But only using parameters shown in Table I, stress can not be defined. We also require information about the context to interpret the data collected from the sensors and also to understand what was going on at the time of reading collection. This context information can be gathered using mobile phones (IoT based) or from computers as we devote a lot of time using them for our daily life work.

In this study we gave a better in look of state-of-the-art machine learning algorithms and their use as stress level classifier. Table I show 21 algorithms tested on the task of stress level monitoring. We can see that only 6 out of 21, namely; SVM, Random Forest, kNN, MLP, Ada-Boost and Decision Forest machine learning algorithm were able to recognize and classify stress state of a subject with the accuracy of more than 90%. In terms of consistent higher classification rate, SVM classifier can be considered as best classifier for stress monitoring procedures but one have to see number of subjects and type of sensors (signals) required for training and testing of the classifier. The question that which classification algorithm is best for utilization in stress monitoring device remains in place as different people have different perception about the usage of machine learning algorithms. For some people, simpler algorithms are the best while some people thinks no matter how complex an algorithm is but the accuracy should be higher. For us, it is a kind of trade-off between computation time vs accuracy vs price of the device.

REFERENCES

[1] L. Richard, T. Hurst, and J. Lee, "Lifetime exposure to abuse, current stressors, and health in federally qualified health center patients," *Journal of Human Behavior in the Social Environment*, pp. 1–15, 2019.

[2] G. S. Everly and J. M. Lating., "The anatomy and physiology of the human stress response," *A clinical guide to the treatment of the human stress response*, pp. 19–56, 2019.

[3] M. C. Staff, "Stress symptoms: Effects on your body and behavior," 2016. [Available Online] https://www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/stress-symptoms/art-20050987.

[4] E. Hemmingsson, "Early childhood obesity risk factors: socioeconomic adversity, family dysfunction, offspring distress, and junk food self-medication," *Current obesity reports*, vol. 7, pp. 204–209, 2018.

[5] M. Kivimäki and A. Steptoe, "Effects of stress on the development and progression of cardiovascular disease," *Nature Reviews Cardiology*, vol. 15, no. 4, p. 215, 2018.

[6] A. Reisner, P. A. Shaltis, D. McCombie, and H. H. Asada, "Utility of the photoplethysmogram in circulatory monitoring," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 108, no. 5, pp. 950–958, 2008.

[7] Z. Villines and S. Flack, "Why do we use pulse oximetry?," 2017. [Available Online] https://www.medicalnewstoday.com/articles/318489.php.

[8] O. Crew, "How and why to track respiratory rate trends with the oura wellness ring," 2017. [Available Online] https://ouraring.com/track-respiration-rate-trends/.

[9] M. Laudat, S. Cerdas, C. Fournier, D. Guiban, B. Guilhaume, and J. Luton, "Salivary cortisol measurement: a practical approach to assess pituitary-adrenal function," *The Journal of Clinical Endocrinology & Metabolism*, vol. 66, no. 2, pp. 343–348, 1988.

[10] P. Braveman and L. Gottlieb, "The social determinants of health: it's time to consider the causes of the causes," *Public health reports*, vol. 129, no. 1_suppl2, pp. 19–31, 2014.

[11] S. Salimetrics USA, "Salivary cortisol," 2019. [Available Online] https://www.salimetrics.com/analyte/salivary-cortisol/.

[12] M. S. Nagananda *et al.*, "Quantization of mental stress using various physiological markers," *arXiv preprint arXiv:1504.03343*, 2015.

[13] K. W. Kallus and M. Kellmann, *The recovery-stress questionnaires: user manual*. Pearson Frankfurt/Main, 2016.

[14] P. M. Boynton and T. Greenhalgh, "Selecting, designing, and developing your questionnaire," *Bmj*, vol. 328, no. 7451, pp. 1312–1315, 2004.

[15] E. G. Pintelas, T. Kotsilieris, I. E. Livieris, and P. Pintelas, "A review of machine learning prediction methods for anxiety disorders," in *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pp. 8–15, ACM, 2018.

[16] R. R. Bouckaert, "Choosing between two learning algorithms based on calibrated tests," in *ICML*, vol. 3, pp. 51–58, 2003.

[17] S. K. Murthy, "Automatic construction of decision trees from data: A multi-disciplinary survey," *Data mining and knowledge discovery*, vol. 2, no. 4, pp. 345–389, 1998.

[18] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, no. 3, pp. 159–190, 2006.

[19] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pp. 400–408, ACM, 2018.

TABLE I
COMPARISON TABLE OF MACHINE LEARNING APPROACH FOR STRESS MONITORING AND IDENTIFICATION

S.No	Method/Technique	Ref	No. of subjects	Signals/Devices Used	Avg_Accuracy (%)
1	AdaBoost	[19]	15	Motion as well as physiological data using sensors	77.2
2		[27]	18	Physiological and social responses	94.3
3	ANN	[23]	10	ECG and PPG	82.5
4		[24]	20	ECG and PPG	98.8
5		[25]	55	Leaf node evidence, context, workload, and student health trait.	85.3
6	Decision Forest	[48]	30	ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure, and Glucometer	96
7		[48]	30	ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure, and Glucometer	66
8	Decision Tree	[19]	15	Motion as well as physiological data using sensors	74
9		[21]	34	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	88.02
10		[20]	42	Heart Rate Variability (HRV)	79
11	Genetic Algorithm	[46]	60	Pupil diameter (PD), ECG and PPG signals	N/A
12	HMM	[55]	60	Salivary cortisol, heart rate, and self-report ratings + moment of eyebrows and mouth	75-88
13	kNN	[19]	15	Motion as well as physiological data using sensors	70.5
14		[24]	20	ECG and PPG	97.5
15		[27]	18	Physiological and social responses	87.3
16		[28]	11	EEG signals + salivary cortisol	63.46
17	KSOM	[35]	20	GSR, PPG (Respiration rate), ECG and HRV.	81.6
18	LVQ	[33]	50	Skin conductance, electrocardiogram, and respiration rate	86.8
19	LDA	[19]	15	Motion as well as physiological data using sensors	78.05
20		[39]	21	Heart Rate (HR) and Pulse Transit Time (PTT)	60.8 approx
21		[31]	33	Electrodermal activity (EDA)	78.35 approx
22	Logistic Regression	[42]	42	Electroencephalogram (EEG) signals	89 approx
23	MLP	[45]	40	ECG and electrical bio-impedance (TEB) signals	67.7
24		[26]	3	ECG, EMG, and GSR	95.2 approx
25	Naïve Bayes	[21]	33	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	78.65
26		[26]	3	ECG, EMG, and GSR	64 approx
27	Nearest Class Center (NCC)	[31]	33	Electrodermal activity (EDA)	74.52 approx
28	OneR	[26]	3	ECG, EMG, and GSR	49.35 approx
29	One-v-all	[48]	30	ECG, Body Temperature, Pulse Oximeter, GSR, Blood Pressure and Glucometer	72
30	PCA	[37]	25	Heart rate monitor, respiration rate monitor and electrodermal activity sensors	60
31	Random Forest	[19]	15	Motion as well as physiological data using sensors	77.21
32		[51]	20	Heart Rate Variation (HRV) and simple heart rate signals.	97.2
33	SVM	[21]	32	Blood Volume Pulse, GSR, Skin Temperature and Pupil Diameter	90.1
34		[39]	21	Heart Rate (HR) and Pulse Transit Time (PTT)	62.3 approx
35		[46]	60	Pupil diameter (PD), ECG and PPG signals	90.1
36		[57]	50	Heart rate monitoring sensor (ECG, PPG), skin conductance (GSR) and skin temperature monitoring sensor	91.26
37		[58]	15	ECG, EEG and EDA signals + Salivary samples	86
38		[25]	55	Leaf node evidence, context, workload, and student health trait.	90.8
39		[26]	3	ECG, EMG, and GSR	95.2 approx
40		[28]	11	EEG signals + salivary cortisol	71.72
41	ZeroR	[26]	3	ECG, EMG, and GSR	12.5 approx

Note: approx means it is average of all the accuracies achieved by author/s in different scenarios. OneR and ZeroR have accuracies less than 50% (selection criteria) but are baseline algorithms that is why they are included here.

[20] R. Castaldo, W. Xu, P. Melillo, L. Pecchia, L. Santamaria, and C. James, "Detection of mental stress due to oral academic examination via ultra-short-term hrv analysis," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3805–3808, IEEE, 2016.

[21] A. Barreto, J. Zhai, and M. Adjouadi, "Non-intrusive physiological monitoring for automated stress detection in human-computer interaction," in *International Workshop on Human-Computer Interaction*, pp. 29–38, Springer, 2007.

[22] G. P. Zhang, "Neural networks for classification: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 30, no. 4, pp. 451–462, 2000.

[23] M. Singh and A. B. Queyam, "A novel method of stress detection using physiological measurements of automobile drivers," *International Journal of Electronics Engineering*, vol. 5, no. 2, pp. 13–20, 2013.

[24] D. Cogan, M. B. Pouyan, M. Nourani, and J. Harvey, "A wrist-worn biosensor system for assessment of neurological status," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5748–5751, IEEE, 2014.

[25] P. Verma and S. K. Sood, "A comprehensive framework for student stress monitoring in fog-cloud iot environment: m-health perspective," *Medical & biological engineering & computing*, vol. 57, no. 1, pp. 231–244, 2019.

[26] R. A. Calvo, I. Brown, and S. Scheding, "Effect of experimental factors on the recognition of affective mental states through physiological measures," in *Australasian Joint Conference on Artificial Intelligence*, pp. 62–70, Springer, 2009.

[27] O. M. Mozos, V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, R. Dobrescu, and J. M. Ferrandez, "Stress detection using wearable physiological and sociometric sensors," *International journal of neural systems*, vol. 27, no. 02, p. 1650041, 2017.

[28] H. Jebelli, S. Hwang, and S. Lee, "Eeg-based workers' stress recognition at construction sites," *Automation in Construction*, vol. 93, pp. 315–324, 2018.

[29] H. Parvin, M. Mohamadi, S. Parvin, Z. Rezaei, and B. Minaei, "Nearest cluster classifier," in *International Conference on Hybrid Artificial Intelligence Systems*, pp. 267–275, Springer, 2012.

[30] H. Parvin, M. Mohamadi, S. Parvin, Z. Rezaei, and B. Minaei, "Nearest cluster classifier," in *International Conference on Hybrid Artificial Intelligence Systems*, pp. 267–275, Springer, 2012.

[31] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 410–417, 2010.

[32] J. Brownlee, "Learning vector quantization for machine learning," 2016. [Available Online] https://machinelearningmastery.com/learning-vector-quantization-for-machine-learning/.

[33] S. C. Pauws, M. Biehl, *et al.*, "Insightful stress detection from physiology modalities using learning vector quantization," *Neurocomputing*, vol. 151, pp. 873–882, 2015.

[34] A. Ahmad and R. Yusof, "A modified kohonen self-organizing map (ksom) clustering for four categorical data," *Jurnal Teknologi*, vol. 78, no. 6-13, 2016.

[35] R. R. Singh, S. Conjeti, and R. Banerjee, "Biosignal based on-road stress monitoring for automotive drivers," in *2012 National Conference on Communications (NCC)*, pp. 1–5, IEEE, 2012.

[36] L. D. Hamilton, "Introduction to principal component analysis (pca)," 2014. [Available Online] http://www.lauradhamilton.com/introduction-to-principal-component-analysis-pca.

[37] H. M. Khan, B. Ahmed, J. Choi, and R. Gutierrez-Osuna, "Using an ambulatory stress monitoring device to identify relaxation due to untrained deep breathing," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1744–1747, IEEE, 2013.

[38] S. Raschka, "Linear discriminant analysis– bit by bit," 2014. [Available Online] https://sebastianraschka.com/Articles/2014_python_lda.html.

[39] C. Dobbins and S. Fairclough, "Detecting negative emotions during real-life driving via dynamically labelled physiological data," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pp. 830–835, IEEE, 2018.

[40] A. Navlani, "Understanding logistic regression in python," 2018. [Available Online] https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python.

[41] M. Nandu, "Machine learning using logistic regression in python with code," *Becoming Human: Artificial Intelligence Magazine*, 2019.

[42] A. R. Subhani, W. Mumtaz, M. N. B. M. Saad, N. Kamel, and A. S. Malik, "Machine learning framework for the detection of mental stress at multiple levels," *IEEE Access*, vol. 5, pp. 13545–13556, 2017.

[43] J. Brownlee, "How to estimate a baseline performance for your machine learning models in weka," 2016. [Available Online] https://machinelearningmastery.com/estimate-baseline-performance-machine-learning-models-weka/.

[44] s. Sayad, "Oner," 2019. [Available Online] http://www.saedsayad.com/oner.htm.

[45] I. Mohino-Herranz, R. Gil-Pita, J. Ferreira, M. Rosa-Zurera, and F. Seoane, "Assessment of mental, emotional and physical stress through analysis of physiological signals using smartphones," *Sensors*, vol. 15, no. 10, pp. 25607–25627, 2015.

[46] F. Mokhayeri, M. Akbarzadeh-T, and S. Toosizadeh, "Mental stress detection using physiological signals based on soft computing techniques," in *2011 18th Iranian Conference of Biomedical Engineering (ICBME)*, pp. 232–237, IEEE, 2011.

[47] G. Ericson, T. Petersen, J. Takaki, K. Sharkey, J. Martens, and Jeffry, "Two-class decision forest," 2018. [Available Online] https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-forest.

[48] F. P. Akbulut and A. Akan, "A smart wearable system for short-term cardiovascular risk assessment with emotional dynamics," *Measurement*, vol. 128, pp. 237–246, 2018.

[49] G. Ericson, T. Petersen, K. Sharkey, J. Martens, and J. Takaki, "Two-class decision jungle," 2018. [Available Online] https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-jungle.

[50] N. Donges, "The random forest algorithm," 2018. [Available Online] https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd.

[51] O. Oti, I. Azimi, A. Anzanpour, A. M. Rahmani, A. Axelin, and P. Liljeberg, "Iot-based healthcare system for real-time maternal stress monitoring," in *2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 57–62, IEEE, 2018.

[52] X. Zhang, G. Ericson, K. Sharkey, J. Martens, and J. Takaki, "One-vs-all multiclass," 2018. [Available Online] https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/one-vs-all-multiclass.

[53] Z. Lai, "A short introduction: Boosting and adaboost," 2016. [Available Online] https://helloacm.com/a-short-introduction-boosting-and-adaboost/.

[54] D. S. Batista, "Hidden markov model and naive bayes relationship," 2017. [Available Online] http://www.davidsbatista.net/blog/2017/11/11/HHM_and_Naive_Bayes/.

[55] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation, space, and environmental medicine*, vol. 76, no. 6, pp. B172–B182, 2005.

[56] A. Bilski, "A review of artificial intelligence algorithms in document classification," *International Journal of Electronics and Telecommunications*, vol. 57, no. 3, pp. 263–270, 2011.

[57] T. Salafi and J. Kah, "Design of unobtrusive wearable mental stress monitoring device using physiological sensor," in *7th WACBE World Congress on Bioengineering 2015*, pp. 11–14, Springer, 2015.

[58] S. Betti, R. M. Lova, E. Rovini, G. Acerbi, L. Santarelli, M. Cabiati, S. Del Ry, and F. Cavallo, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, pp. 1748–1758, 2018.