

MLSCD – Stress Detection Analysis model using Stress Symptoms

Dr. M. V. Vijaya Saradhi,

Professor, Dept. Of CSE, ACE Engineering College, Ghatkesar, Hyderabad, India.

Email: meduri.vsd@gmail.com

Dr. Kalli Srinivasa Nageswara Prasad

Professor, Department CSE, Ramachandra College of Engineering, Eluru, Andhra Pradesh.

Email: kallisinprasad@gmail.com.

Article Info

Volume 81

Page Number: 2655- 2665

Publication Issue:

November-December 2019

Abstract

Fundamental analysis of the stress and its related elements signify that the stress could be attributed to various factors like the professional demands, too much of physical or mental exertion, unrealistic goals and ambitions, non-conducive working or family or social environment and various such conditions. The statistics revealed in many of the healthcare and the social science studies reflect on the alarming rates of the stress conditions. Though there were many effective models that were developed in the past, there are a very limited set of self-analysis-oriented models in stress assessment. This manuscript proposes a model MLSCD which relies on 15 features classified from three categories of symptoms (mind related, body-related, and behavior related). Considering behavior related as dependent with the other two as independent variable groups, the model is trained with SVM classifier for over 900 records and is tested for 300 records. The accuracy rate of 97% indicates that the model is potential and if iterative improvements can be made, it can be a pragmatic solution for real-time applications

Keywords: *Stress, Machine learning model for stress, Stress Analysis, Stress index.*

Article History

Article Received: 5 March 2019

Revised: 18 May 2019

Accepted: 24 September 2019

Publication: 12 December 2019

Introduction

1.1 Outlook

The facet called stress has become an epidemic and today, irrespective of the age group, the number of cases detected with the stress-related health and mental problems is growing exponentially. Technically, the concept of

stress can be defined as a condition wherein a human is prone to severe duress conditions resulting in mental unrest, physical fatigue, depressive symptoms and many more other health complications[1].

The root cause analysis studies of stress signify that the stress could be attributed to various factors like the professional demands,

too much of physical or mental exertion, unrealistic goals and ambitions, non-conducive working or family or social environment and various such conditions. The statistics revealed in many of the healthcare and the social science studies reflect on the alarming rates of the stress conditions[2]. Today, the medical institutions are witnessing hundreds of cases every day where the health complications reported are vehemently the outcome of physical and mental stress that is encountered by the individuals.

Numerous studies are focusing on preventive care and the treatment conditions for addressing the stress-related suffering in the patients. While there are many organic and natural treatments, self-healing models (meditation, yoga, cross-fit exercises, music therapy and many more such) are presented as preventive or curing methods, there are many physiological and psychological treatments too were advocated and counseled to the patients in the medical institutions[3].

But the crux that remains as concerned in the whole scenario is about thousands of people who either do not realize the internal levels of stress that they are encountering in their lives or in the denial to such stress conditions. Hundreds of studies that have focused on the conditions of both physical and mental stress conditions indicate the conditions which can lead to more complexities and conditions of health implications for people[4].

In lines of healthcare developments, one of the significant developments is the usage of the information systems and computer-aided solutions in the diagnostics, treatment, and care for the patients, continuous monitoring of the patient conditions, etc. There are many wearable devices too that are available in the market, which is enabling the users to have tracking of the stress levels. Predominantly, a few of the factors that signify the conditions of stress among the people are the fluctuating

blood pressure levels, higher pulse rates and enigmatic behaviors of the people.

In medical parlance, there are a distinct set of gradings for the vivid kind of stress conditions, and it is very important that the early diagnosis of the stress conditions in the people can do a lot in terms of preventive measures, treatment to ensure that the patient is not affected deep. There are many computer-aided models, devices, and solutions too that are prevalent in terms of stress detection measures[1].

However, there are certain limitations that are being encountered in the real-time environment, wherein as discussed in the earlier lines, that the denial from people in accepting the stress conditions and seeking treatment. In the other dimension, the social stigma attached in terms of people getting treatment for stress and depression too is high in many societies which are leading to poor numbers of patients seeking help from the medical professionals or the other alternative treatment models.

The disparity in terms of a number of people being reported with stress-related impacts, the people suspected to be prone to the stress are alarming, while very few percentages of the suffering or prone are taking treatments. Such widespread gap in the system is leading to conditions wherein the practical measures that can be used for in-depth analysis of the conditions and enabling counseling sessions can help in addressing the epidemic[2], [3].

1.2 Purpose of the Research

There are numerous models of stress levels indication models that were proposed in the past. Many such models are pragmatic and are even used in a distinct set of medical and non-medical counseling and personality assessment conditions. The parameters or features that are used in the case of analysis is varying between one set of tests to the other. Metrics that are related directly or indirectly towards reflecting on the stress conditions in the human body and

mind are taken in to account for understanding the stress conditions, which could provide considerable insights into the system functioning[4].

Technically, there are some tests that are considered as significant in medically relative terms, and certain levels of computer-aided detection models too were proposed in the earlier studies. Even certain levels of wearable device-based metrics like the pulse rate, sleep quality and other such factors too were taken into account and some models of stress indicators were proposed.

Despite all such developments, there are very few models that were developed to have minimal human intervention activity and towards improving the process outcome which can lead to sustainable ways of providing the stress impact analysis conditions. If there is a system wherein a questionnaire is answered by a person, the system can detect the stress level classification[5].

If the system can be developed, which can provide more accuracy and consistency in its detection process, it shall help the people in having even personal level analysis of the conditions and reflection for further ways to address to the issue[6].

1.3 Proposed Solution

The proposed solution towards the model is about developing a comprehensive and self-analyzing system for stress score detection; the inputs for the model depends on certain metrics that shall be used for garnering inputs which can be used for detailed analysis using machine learning solutions. The model can be developed into a weblink that can be accessed by the users, provide input to the key inputs gathered in the questionnaire which will be evaluated and the stress score indication shall be provided to the user[7].

1.4 Organizing of Paper

Setting the background context and purpose of this solution manuscript, in the section-2 the emphasis is on the conditions and metrics that are chosen for the proposed model, based on the earlier models of stress analysis systems. Followed by in section 3, the framework of the model is discussed along with the insights pertaining to the proposed algorithm, datasets used, and the test reports are discussed. Section-4 concludes the insights from the paper.

2 Stress management Solutions

Numerous AI-based stress detection models were proposed in the past. While some of them has taken in to account ECG based solutions [number], there are a certain set of studies that have focused on image-based detection models, that can support in understanding the stress levels. Few of the other models that have focused on the domain has provided solutions in the form of using the deep learning network models and the other parameters that can support in an in-depth analysis of the condition[8].

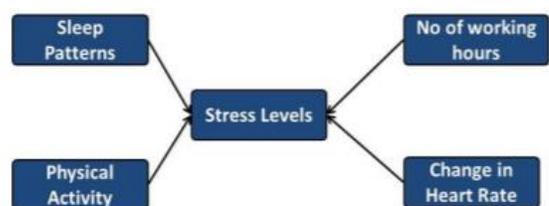


Figure 1: The Conceptual model of STRESS.

The Figure 1 above reflects a comprehensive system that is used in the case of the stress detection model, wherein the key dimensions in which the parameters are chosen are sleep pattern-based inputs, working hours and working conditions, physical activity-related conditions and the body vitals-based metrics.

Predominantly, the studies have focused either on one or more features from all the above or specifically from some of the above,

towards developing a system for analysis. However, the models have been effective in terms of detecting the stress scores in alignment with the common conditions[9].

The medical treatment models either focus on the symptoms-based understanding of the problem or focusing deep on the root cause conditions for addressing the health ailment. In lines of homeopathy treatment, profoundly for the majority of the health reasons, one or the other levels of stress factors over the body or mind could trigger dis-ease in the body and if the root-cause conditions can be addressed, the health issues encountered by the member can be significantly addressed[10].

The other dimension, in which the facets of stress management analysis can be considered is to focus on the distinct dimensions like a combination of symptoms, causes and behavioral actions that are attributed to certain kinds of stress conditions.

Indeed, the stress symptoms have a significant effect on the body, thoughts and feelings and the behavior conditions. If the common stress symptoms are recognized it will help the people take pertinent actions to address the stress conditions. If the stress issues are ignored, the studies emphasize the potential chances of more health complications like the increased blood pressure, cardiovascular problems, obesity, diabetes (type-II) the ones that can lead to more challenges for the system[2].

If the medical terms of the system can be evaluated for detailed analysis, there are increased chances of stress detection. In the case of the proposed models, it is imperative that if the key parameters of medical and non-medical features are used in pragmatic metrics, it can lead to significant insights into the stress detection analysis[3].

There are many online tools that enable the stress detection patterns, however, if there is a comprehensive system that can support with the overall parameters-based detection system, it can lead to the significant outcome for the process.

3 MLSCD Model

3.1 Framework

The proposed solution is about developing an integrated stress scoring system that takes in to account certain key features that are integral to the performance conditions of the case and accordingly use the process of weighted average based analysis in terms of detecting whether the person is undergoing stress and if so the grading in terms of what level of stress an individual is encountering.

The model chosen for the process is to select around 19 symptoms that are more pragmatic in the case of the stress conditions. In the absence of imbalanced scoring rates for the respective features or the criteria that decide the intensity, a nominal solution in the form of simple average scoring analysis models can be used as the solution. But in terms of in-depth analysis, the weightage for each of the parameters might change, which will eventually change the metrics by which the scoring patterns and the implications are evaluated[11].

3.2 Metrics /Features

For the proposed model, the classification of the features that are taken into account is from three major classification segments like the ones that are symptoms from the body, mood, and the behavior respectively.

The following set of Table 1 information provides insights about conditions or metrics that are chosen for detailed analysis from the system.

Table 1: Stress Features Grouping

| On the Body | On the Mood | On the Behavior |
|--------------------------|-----------------------------|---------------------------|
| Headache | Anxiety | Overeating / under eating |
| Muscle tension /pain | Restlessness | Angry outbursts |
| Chest pain | Lack of motivation or focus | Drug or alcohol misuse |
| Fatigue / Sleep problems | Feeling overwhelmed | Tobacco use |
| Change in sex drive | Irritability or anger | Social withdrawal |
| Stomach upset | Sadness or depression | Exercising less often |

The approach for the process is about developing a questionnaire wherein the questions are designed in simple and precise manner for the user to comprehend and respond. The scaling of the response is linked to the range of 0-10 wherein the data must be rated for analysis.

Classification of the stress category is one of the critical factors that are to be considered, wherein the stress levels are graded as High, Moderate, Low or No stress conditions. The classification and rating are on varying parameters, wherein if the behavioral symptoms score for certain parameters are low, and the body related parameter score averages are high, in that case, the stress shall be considered as moderate or low depending on the score range. Whereas if the behavior score is extremely high and the body and mind scores are average, in such conditions the classification of stress shall be marked as High.

3.3 Modeling the System

The system proposed for development is the usage of machine learning model MLSCD, which shall be trained with labeling classification for various features considered and the correlation co-efficiency for the system. The process for the proposed model is There are various conditions that are discussed as symptoms in the conditions identified above,

Step 1. Provide Questionnaire for response sheet

Step 2. All the parameter related questions are reviewed for absence of null and void

Step 3. Calculate correlation among the parameters

Step 4. Compare assigned and defined correlation values to estimate weightage

Step 5. Execute the data analysis to classify the grade of stress

which provides comprehensive inputs into various features that can be described in simple terms by the user.

3.4 Correlation Co-efficiency

Correlation is the statistical inference reflecting on whether and how strongly pair of variables are related. For the proposed model, vividly used solution Pearson correlation model shall be used, wherein it is possible to estimate the relationship among the two variables whilst removing the effect of one of two of the other variables. The actual definition of correlation coefficient could be attributed to the **correlation coefficient** (or "r"). It ranges from -1.0 to +1.0. The closer r is to +1 or -1, the more closely the two variables are related.

The formula that shall be used for the training sets to understand the correlation among the parameters are

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2] [n\sum y^2 - (\sum y)^2]}}$$

P= Pearson's correlation Coefficient value

Even in the case of the testing data, the algorithm shall consider the P-value and the correlation among various attributes for further levels of analysis.

3.5 Data Classifiers

In the training of the machine learning model, the role of classifiers used for handling the training process has significant importance. For the proposed model, the SVM classifiers are used to train the algorithm to understand the data and the classification.

The SVM (Support Vector Machines) are highly dependent on the decision planes which define the set of decision boundaries that impact the decision conditions. In the case of a decision plane, the process is about segregation among the set of objects to different class membership. For representative purposes, in the following Figure 2, the objects are classified into two categories as Red or Green and the separating lines in the image below reflects on the boundary to a specific classification.

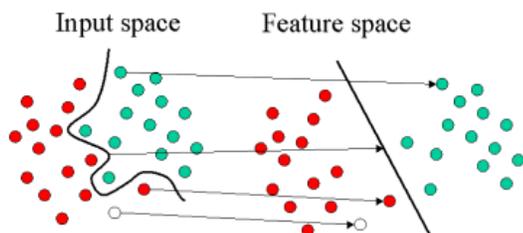


Figure 2: SVM Classifier Hyperplane

As SVM enables the algorithm in handling both regression and classification tasks and can handle multiple continuous and categorical variables, the model is chosen for the proposed machine learning model. To identify the categorical variables, a dummy variable is created with case values as either 0 or 1. Thus, a categorical dependent variable consisting of three levels, say (A, B, C, D), is represented by a set of four dummy variables:

$$A: \{1\ 0\ 0\ 0\}, B: \{0\ 1\ 0\ 0\}, C: \{0\ 0\ 1\ 0\}, D: \{0\ 0\ 0\ 1\}$$

In terms of developing an optimal hyperplane, SVM works on iterative training algorithm model which shall be used for minimizing any kind of errors in the function.

The classification training model for the project is trained using the following set equation

$$\frac{1}{2} w^T w - \nu \rho + \frac{1}{N} \sum_{i=1}^N \xi_i$$

subject to the constraints:

$$y_i (w^T \phi(x_i) + b) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, \dots, N \text{ and } \rho \geq 0$$

{
w be vector of coefficients,
b is a constant, and

ξ_i represents parameters for handling non-separable data (inputs).

The index i labels the N training cases.

Let, $y \in \pm 1$ denote the class labels and x_i denote the independent variables.

The kernel ϕ is embraced to alter data from the input (independent) to the feature space.

}
In the usage of the above classifier model for training the machine learning system, the functional dependency of the dependent variable y over the set of independent variables x shall be considered. It is presumed in the model that the relationship among the independent and dependent variables are provided using the deterministic function f plus the additive factors.

3.6 Algorithm

The algorithm used for the proposed model is about

Let S be the stress score used for rating and classification with three subcategories for evaluation as

Bp – Features of Body parameters

Mp – Features of Mind Parameters

Fp – Features of Behavioral Parameters

{

P is Correlation Co-efficiency P-value of all the parameters together

$$P = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2] [n\sum y^2 - (\sum y)^2]}}$$

The correlation values are estimated to understand the significance and relativity among the patterns.

Whereas, individually all the sub-score (Fb_{1-5}) are significant for the system and the value range is

$$Fb = \geq 1 \leq 50$$

All the five parameters with scoring range of 1-10

If the $Fb \geq 36$, the case is considered for "Highly Stress" denoted as HS

If the $Fb \geq 27 \leq 35$, the case is considered for "Moderate Stress" denoted as MS

If the $Fb \geq 19 \leq 26$, the case is considered for "Low Stress" denoted as LS

If the $Fb \geq 1 \leq 18$, the case is considered for "No Stress" denoted as NS}

{
The Mind score category is considered as Mind parameters Mp and Body parameters Bp are considered for the stream of scoring range 1-10

$$Bp = \geq 1 \leq 50$$

$$Mp = \geq 1 \leq 50$$

$$R = \text{average of } \left(\sum (Bp + Mp) \right)$$

If the $R \geq 36$, the case is considered for "Highly Stress" denoted as HS

If the $R \geq 27 \leq 35$, the case is considered for "Moderate Stress" denoted as MS

If the $R \geq 19 \leq 26$, the case is considered for "Low Stress" denoted as LS

If the $R \geq 1 \leq 18$, the case is considered for "No Stress" denoted as NS

$$R = \text{either } W \text{ or } X \text{ or } Y \text{ or } Z$$

{
When the input "S" is tested for S-value allocation, the process followed is about

{
Eligibility score

If the Fp has the label of HS.... then only Rw must be used for analysis

If the Fp has the label of MS ...then Rx or Rw or Ry can be considered for allocation accuracy prediction

If the Fp has a score of LS, only Rx or Ry shall be considered

If the Fp has a score of NS, Rx or Ry shall be considered

}

{

If the S score validation is effective, then the scoring output for task allocation accuracy is estimated as

$$S = 2 * ((Fp * R) / (Fp + R))$$

}

If the $S \geq 27$, the Task is rated as Highly Stress

If the $S \geq 20 \leq 26$, the Task is rated as Moderate Stress

If the $S \geq 15 \leq 19$, the task is rated as Low Stress

If the $S \geq 1 \leq 18$, the task is rated as No-Stress

}

3.7 Data Sets

The proposed model is tested based on the data scores generated using the random sampling online systems, wherein the datasets are validated for completion of the inputs from all the key parameters sought. The privacy information if any supplemented in the questionnaire is segregated and the key scores for all the metrics are considered.

For the purpose of classification, the Body and the Mind parameters are independent variables which have a significant impact on the behavior as the critical variable which is dependent for both the conditions.

Based on the inputs evaluated, the model is tested for random classification for the developed datasets. However, the data set that is considered is self-rated scoring, wherein it is expected that the user-centric scores used for training and testing are relatively accurate to the personal data.

The following Table 2 provides the inputs for the distinct set of data used for the training and testing of the proposed model. The major

classifications that the model is trained for are High stress, moderate stress, low stress, and no-stress. The first three categories are technically

considered as positive results, and the no-stress class is considered for a negative result.

Table 2 Dataset Classification

| Classification | Volume of Records | Additional Inputs |
|---|-------------------|---|
| Total Data used for analysis | 1200 | |
| 3 sets used for training | 300 | Every set constituting 300 records for training, wherein the weightage p-value of each of the set has different higher correlations. For instance, in the first set, body parameters correlation is high, and in the second set the mind parameters were high and the third set had the average for all three parameters body, mind, and behavior. |
| Positive records in the training set | 650 | The records that can be classified for high stress or moderate stress or Low Stress |
| Records of No-effective, in the training set. | 250 | Records that constitute No-stress conditions training records. |
| Total records used for testing the system | 300 | Test data |
| Positive records in the test data | 250 | Records that are to be classified for a positive category of stress |
| Negative records in the test data | 50 | Records to be traced by the system as no-stress |

3.8 Test Result Analysis

The test result analysis that could be attributed to the model is discussed in this section. Based on the inputs classified in the dataset's tables, the proposed model is trained and tested for accuracy in the output predictions. The

following are some of the key metrics that are integral to the proposed conditions.

The set of outcomes that is predicted as rightly by the system are depicted in the following Table 3

Table 3: Performance metrics

| Metrics | Detection from Test Records | Proportion Accuracy |
|----------------|-----------------------------|---------------------|
| True positive | 244 | 97.60 |
| False Positive | 6 | 2.40 |
| False Negative | 3 | 6.00 |
| True Negative | 47 | 94.00 |

The total records tested in the model are about 300 and the model has the true positive rate of performance could be rated to 97.60 which is highly significant in terms of the overall performance of the system. Similarly, in the context of the number of negatives that are depicted wrongly by the system, it is imperative that the rate of performance is a little lower in terms of showing only around 94%. Hence, there is a need for improving the rate of detecting the [12]false-positive factor conditions in Table 4, Figure 3.

Table 4: Accuracy Classification

| Metrics | Performance |
|-------------|-------------|
| Precision | 97.60 |
| Recall | 98.79 |
| Accuracy | 97.00 |
| F1 Score | 98.19 |
| Specificity | 88.68 |



Figure 3: The Performance Accuracy Clasifications.

Towards detecting the conditions of efficacy in a system, it is very important to know about the performance metrics like the precision, recall, accuracy, F1 score, which leads to indications over the specificity and sensitivity of the model. Taking stock of the performance conditions depicted for the model, it is evident that the model is pragmatic and if can be used more in lines with trained dataset conditions, still the accuracy rates of the systems can be improved.

In the other integral classification that is considered in the evaluation metrics, is the accuracy of detection for various classes of positive categorization. This is also very important in terms of knowing how the class classification is accurate in terms of rating the stress conditions. For instance, in factual conditions, if a case is suffering from high stress, reflecting the case in moderate or low-stress category shall lead to more implications. Hence, the class-based accuracy of the model is also evaluated using descriptive inference[12].

Table 5 Group Level Classification

| Classification | Actual Segregation | Positive Tracing | Accuracy % |
|-----------------|--------------------|------------------|------------|
| Highly Stress | 74 | 72 | 97.3 |
| Moderate Stress | 107 | 104 | 97.2 |
| Low Stress | 69 | 68 | 98.6 |
| No-Stress | 50 | 47 | 94.0 |

The key inputs evaluated in the model indicates that out of the total 250 records that are considered for the positive tracing, with segregation to different stages as depicted in the Table 5, at the high-stress levels, the parameters of evaluation indicates effectively at 97.3% accuracy, while the low-stress level detection accuracy is higher at the value of 98.6%.

The issue of specificity in the system needs certain levels of improvement as the model strikes at 88%, which needs iterative learning to be enforced on the model towards detection of the parameters. If such iterative development is developed, there could be a more significant outcome in the performance.

4 Conclusion

There are numerous models of stress levels indication models that were proposed in the past. Many such models are pragmatic and are even used in a distinct set of medical and non-medical counseling and personality assessment conditions. The parameters or features that are used in the case of analysis is varying between one set of tests to the other. Metrics that are related directly or indirectly towards reflecting on the stress conditions in the human body and mind are taken in to account for understanding the stress conditions, which could provide considerable insights into the system functioning. Considering the gap in the existing forms, a three-class symptom-based model proposed in this study is trained using the SVM classifier for the datasets. The records of performance imperative for the model reflects that the model is more significant in terms of values that are attained for respective

categories, and if the specificity factor is improved more with iterations, the model can be more sustainable for practical application of the system in a real-time environment. As a future scope of extension to the model, the weighted average for respective categories to decide the dependent and independent variable category could be explored, which can make the system more dynamic.

References

- [1] D. Shon, K. Im, J. H. Park, D. S. Lim, B. Jang, and J. M. Kim, "Emotional Stress State Detection Using Genetic Algorithm-Based Feature Selection on EEG Signals," *Int. J. Environ. Res. Public Health*, 2018.
- [2] V. Corcoba-Magaña, M. Muñoz-Organero, and X. G. Pañeda, "Prediction of motorcyclist stress using a heartrate strap, the vehicle telemetry androadinformation," *Journal of Ambient Intelligence and Smart Environments*. 2017.
- [3] A. Liapis, C. Katsanos, D. Sotiropoulos, M. Xenos, and N. Karousos, "Stress recognition in human-computer interaction using physiological and self-reported data: A study of gender differences," in *ACM International Conference Proceeding Series*, 2015.
- [4] 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*, 2017.
- [5] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybernetics and Biomedical Engineering*. 2019.
- [6] P. Gastaldo, L. Pinna, L. Seminara, M. Valle, and R. Zunino, "A tensor-based approach to touch modality classification by using machine

- learning,” *Rob. Auton. Syst.*, 2015.
- [7] H. Jebelli, M. M. Khalili, S. Hwang, and S. H. Lee, “A supervised learning-based construction workers’ stress recognition using a wearable electroencephalography (EEG) device,” in *Construction Research Congress 2018: Safety and Disaster Management - Selected Papers from the Construction Research Congress 2018*, 2018.
- [8] B. Shiner *et al.*, “Measuring use of evidence based psychotherapy for posttraumatic stress disorder,” *Adm. Policy Ment. Heal. Ment. Heal. Serv. Res.*, 2013.
- [9] R. Gopalakrishna Pillai, M. Thelwall, and C. Orasan, “Detection of Stress and Relaxation Magnitudes for Tweets,” 2018.
- [10] S. Tuarob *et al.*, “How are you feeling?: A personalized methodology for predicting mental states from temporally observable physical and behavioral information,” *J. Biomed. Inform.*, 2017.
- [11] V. J. Madhuri, M. R. Mohan, and R. Kaavya, “Stress management using artificial intelligence,” in *Proceedings - 2013 3rd International Conference on Advances in Computing and Communications, ICACC 2013*, 2013.
- [12] D. N. Carvajal and P. C. Rowe, “Sensitivity, specificity, predictive values, and likelihood ratios,” *Pediatr. Rev.*, vol. 31, no. 12, pp. 511–513, Dec. 2010.