



# IT EMPLOYEE STRESS PREDICTION BY USING MACHINE LEARNING AND COMPUTER VISION TECHNIQUE

Reeja S R

Associate Professor-Computer Science Department  
Dayanand Sagar University  
Bangalore, Karnataka  
[reeja-cse@dsu.edu.in](mailto:reeja-cse@dsu.edu.in)

Lavanya S

Student-Computer Science Department  
Dayanand Sagar University  
Bangalore, Karnataka  
[lavanyasreddy214@gmail.com](mailto:lavanyasreddy214@gmail.com)

Monish L

Student-Computer Science Department  
Dayanand Sagar University  
Bangalore, Karnataka  
[monishlreddy@gmail.com](mailto:monishlreddy@gmail.com)

Abdul Basit Lone

Student-Computer Science Department  
Dayanand Sagar University  
Bangalore, Karnataka  
[mail.basitlone@gmail.com](mailto:mail.basitlone@gmail.com)

**Abstract:** Stress assessment must be performed in early stages because stress related issues tend to rise in depression, heart attacks and strokes. Stress leads to have a greater impact on the thoughts sometimes provokes suicidal attempts within employees. The machine learning techniques has proved to be extensive for medical analysis and prediction. This approach can further be used with neurological tools. The 2017 OSMI mental health survey dataset represents working professionals within the IT company and this dataset is used for classifying the stressed or unstressed employees. After data cleaning and pre-processing further training of model is performed by using machine learning techniques. The accuracy is predicted for both the models. Among those models boosting has given highest accuracy.

**Keywords:** Stress, SVM, Random forest classifier, Mental disorder

## I. INTRODUCTION

Stress is one of mental health disorder among the working professional. Several surveys are considered to find out the people who are suffering mostly from this health issue. According to the surveys conducted by tech company association, ASSOCHAM, forty-three percentage of employees suffer from anxiety and depression due to long working hours and deadline boundary. As recorded in the survey by Optum [1] the mental stress related issues are increasing in a greater proportion in India which is published in the Economic Times article. The survey is performed for 8 Lakh employees from over 69 companies with 4000 or above workforce. For increased productivity and wellbeing of the employees must be given a prime importance by creating stress free environment. Numerous steps can be taken to help employees cope up with stress for mental well-being like stress management sessions and counselling for career guidance. Late detection of such employees will lead to poor management of stress.

Now-a-days system turns to learn by itself and improve through trial and error from self-experience without being programmed using machine learning. Such an application can be done through artificial intelligence (AI). Programs which was developed using machine learning requires huge amount

of data to train for proper classification and to train up model. By using image mining the hidden data, additional pattern which are unclearly visible can be easily extracted. It is indirectly related to Machine Learning, Artificial intelligence, Data mining and Image Processing fields.

We hope this process can be implemented in easier way by using machine learning methods and by developing a model which perform proper classification of employees who is depressed and predicting them at early stage, giving them proper guidance to balance stress by considering all professional and personal parameters. This technique helps HR to take a preventive measure about their employees and reduce the chance of employee's resignation from company. Early prediction for early diagnosis can be performed to treat a serious mental health condition.

## II. LITERATURE SURVEY

**U. Reddy et al. [1]** Stress is major drawback for the individuals working in IT industries. The change in their lifestyles and the working styles of the individuals will exaggerate the stress to higher levels. As listed in OSMI dataset 2017 from tech industry [1], different Machine learning strategies like Decision Tree, Random Forest, boosting, etc. used different characteristics like gender type, age, medical records, family ancestry, etc. In this paper U.

Reddy concluded that 75% of the IT employees were marginally at danger [1].

**G. Harrison et al. [2]** Globalization steps towards innovation leading to advancement in everyday life with the most fundamental technology is cell phone, playing a vital role in everyone's life. As discussed in the paper in UK 94% of grownups have a cell phone; and more than three fourth of those are advanced mobile phones [2].

the individual faces. Life without the cell phones cannot be imagined.

**Samrat Qaraqe et al. [3]** As indicated by this paper Sensor signs can be used to distinguish feeling of the person. It involves physiological detection of gadgets are using them to gather these signs. Signal pre-processing and pre handling must be considered to decide the highlight for choosing the sign. When the highlights are chosen then the calculations are made by the AI grouping model. The Pulse variations, skin temperatures and reactions will represent the self-sufficient sensory system movement.

**Huijie Lin et al. [4]**. As discussed in the paper youngsters in the range of 8 and 11 years, are stressed these lower monetary statuses, who clashed among themselves and this had influence on their understudies. Stress will influence the understudies, wellbeing and leads to lack of sleep, weakness and will lead to sorrow.

Sandhi and Asrabadi (1994). The primary worry during the understudies such as language difficulties, racial segregation and visit the family. Understudies quotes that the person stays away from home and settle in other workplace are susceptible to lot of pressure. This study helps in outstating the stress experienced by a person. The understudies are performed from different social foundations states that the person feels like they have their own way of life. The correspondence issue was found in their investigation which lead to worry in the studies.

**N. C. F. Codella et al. [5]** By using dermoscopy images, deep ensemble learning is used in recognition of melanoma. The classification of melanoma is done from dermoscopic images. This classification is done on very scale images availability. The proposed work is only classifying the linear images but need improvement of non-linear images as well and requires improvement in pre-processing stage.

**Chigerwe, Boudreaux et al. [6]** Studies on Maslach Burnout Inventory-Educator Survey (MBI-ES) on veterinary therapeutic. The MBI-ES assesses burnout, including avid weariness, depersonalization, and low close to household success. There is poof to prove that there will be a passionate pressure in veterinary understudies by Chigerwe et al. (2014). In this way, scrutiny of the supreme device to gage the degree of stress will support the distribution and assurance of understudy strengthen the possessions".

**W.Smit et al. [7]**. Using already existing sensors in phones the plans of pressure is observed by using Biobeats. There were lot of precautions or measures taken against the mental shakiness of every individuals by biobeats. Like AI applications wearable sensors are used to predict, check and cure the issues related to mental health. The determination of there is to observe how cerebrum will respond to weight [6].

### III. PROPOSED SYSTEM

Most of the people working as IT professional are suffering from stress disorder. There is the rise in stress of employee with varying work cultures. Though the company offers various schemes for mental health related issues it is far away from the control. In this paper, we applied machine learning techniques to predict the stress patterns in IT employees and analyses the factors that causing in increase in stress levels by using random forest and support vector machine. The OSMI mental health dataset from Kaggle of working employees in IT field was considered. Both the algorithms from machine learning such as SVM and random forest were applied after data cleaning and pre-processing for model training. Once the model building and training is done the results of both the algorithm is obtained and accuracy is compared accordingly. The precision of the above models was obtained and studied comparatively. The prominent features that contribute for stress is identified as gender, family history by using decision trees. And further this application is implemented to extract facial features such as mouth and eyebrows. Eyes and mouth majorly contribute for prediction of stress. The tiredness in mouth, dullness in eyes and furrowed eyebrow all this leads to the indication of stress of an employee. This mechanism encompasses convolving the image with Haar wavelets for distinguishing features in face well and choose two best features for the eyes and mouth. Morlet wavelet is applied to detect the vertical gradient after mouth from the image is identified. The curvature of the lips is calculated, the negative curvature indicates the smile and positive indicates frown. With these findings the industries could slim down the stress on the employees and create a playful environment and workspace for the employees to work.

#### Advantages:

1. Enables to increase the performance of employees by keep motivating them.
2. Even there is a stressful situation this model helps HR to keep track of employee's situation and cheer up them to work.
3. Leads to work even in tougher time.
4. Conflicts in workplace can be reduced.

### IV. SYSTEM ARCHITECTURE

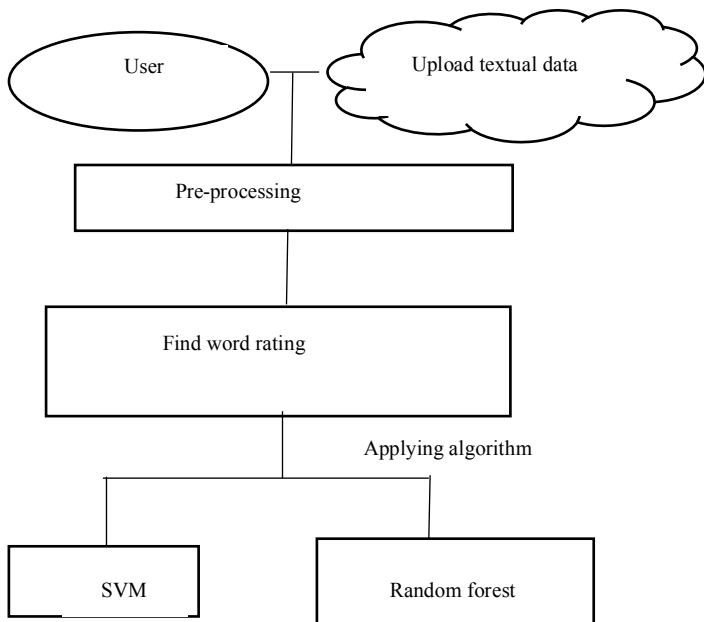


Figure 1: SYSTEM ARCHITECTURE

Figure 1 shows the architecture of the system. The textual data is given to pre-processing stage. Once preprocessing is done then further train the model with SVM and random forest. Further, the results of both the algorithms is compared and predict the best result. Results are discussed in the section V

## V. SYSTEM DESIGN AND IMPLEMENTATION

### Stage I: Upload the Input Data: (Textual data)

Conversations, chats, tweets, sentences or blogs are uploaded as inputs in the textual form. Stress can be analyzed using this textual data. A .txt file with a single sentence or a blog with a paragraph will be loaded into the system as input. Pre-processing step is succeeded.

### Stage II: Pre-Processing Phase:

This phase involves several sub processes to eliminate the noise and the unwanted data which is not required for sentiment analysis. The steps are as follows

**I) Remove Stop Words:** Stops words include the words like I, she, am etc. These stop words are not necessary for sentiment analysis. The common stop words used in English can be given less priority because these stop words will not change the meaning of the sentences or will not indicate any stress related emotions, but will follow along the sentences

- **Replace Slang Words:** In a sentence there could be slang words to indicate the negation, example: wouldn't is a slang word and this should be replaced by would not. These words indicate opposition or negation. So, in the pre-processing stage, all these shortcuts will be replaced by their respective full forms.

- **Remove URL:** In Pre-Processing stage all the URLs and the hyperlinks are removed.

**II) Remove Special Symbol:** Most of the social media data set will have unwanted symbols and emojis, these emojis are converted to special characters, the algorithmic rule will not work on these special symbols. Though the special symbol gives lot of information in short, but it is difficult to implement a algorithm to analyze these special symbols. This phase removes the special symbols like “! @, #, \$, %, &”.

### Stage III: Word Sense Disambiguation:

The Pre-Processed data is the input to this phase. Based on the Disambiguation concept there are two different categories to proceed further.

### Stage IV: Algorithm

Apply the Algorithm. The extracted information is passed to this stage and the algorithm will classify into different classes like Happy, Depressed etc. The different algorithms used are SVM and Random Forest.

**SVM:** SVC classification model is used to learn the dataset. Based on this further classification and prediction was carried out. This algorithm gives good accuracy compared to the other machine learning algorithms. This final categorized result, i.e. Stress, Happy, etc. is given as output by the system.

**Random Forest Classifier:** Cluster of smaller Decision trees are called as Random Forest. RFC is more effective than a single decision tree. One of the major advantages of Random Forest is it works even without hyper tuning the data and RFC is more effective than a single decision tree.

### Stage V:

#### Output

This stage marks the end of the architecture. This stage outputs the Results of the given input sentences, tweet, or blogs in category like Stressed, Depressed, Happy and Non-identified.

## VI. EXPERIMENTAL RESULTS

The two Machine learning algorithms used are SVM and random forest classifier. These algorithms classify the tweets as Stress, Depression, etc. The results identified by the system is expressed with the help of Test cases.

### Measuring Performance:

- **Confusion matrix:**

A confusion matrix is a tabular representation of the effectiveness of the predictions made by the classifier. Based on the confusion matrix, True Positive and the False Positive rates are computed.

Performance of the model is evaluated using the model. Performance metrics like accuracy, precision, recall, and F1-score are computed for the model.

- **Precision and Recall:**

**Precision** is a fraction of instances that are correct, recall is the fraction of instances that are relevant and retrieved. Precision is the measure of quality or completeness. Recall represents the true positive rate for the class. In analysing the information Precision and Recall are used to measure the performance of the system. Precision and Recall are also used in the evaluation matrices. Where, Recall is the measure of sensitivity and Precision is positive predictive value. It will give the percentile of results, and the Recall is the measure of total pertinent results which are classified correctly. Precision (P) is given by the number of True Positives (TP) by the sum of number of False Positives (FP) and True Positives (TP). The formula is as follow:  $Precision = TP / (FP + TP)$

**Recall (R)** is given by number of True Positives (TP) by the sum of total number of False negatives (FN) and True Positive (TP). The formula is as follow:  
 $Recall = TP / (TP + FN)$

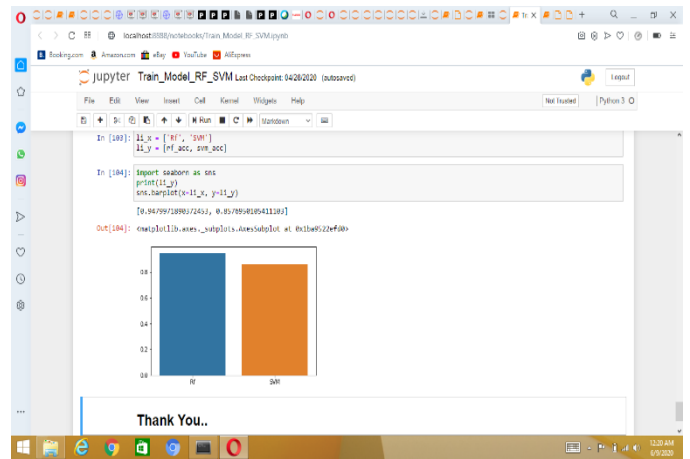


Figure 3: Bar graph to show the prediction through SVM and Random forest

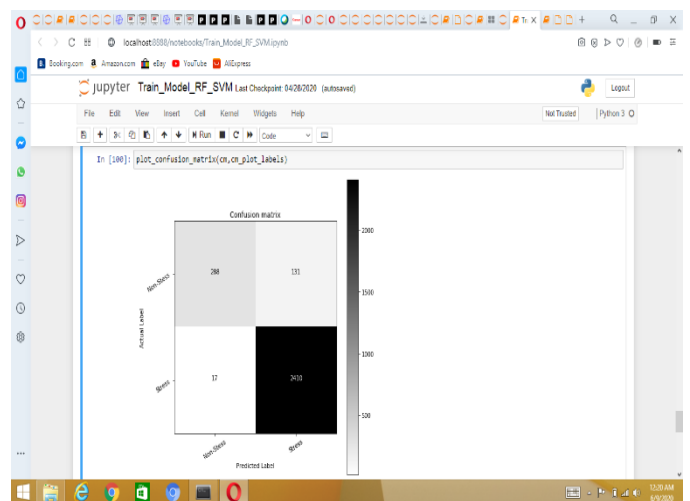


Figure 4: Confusion matrix

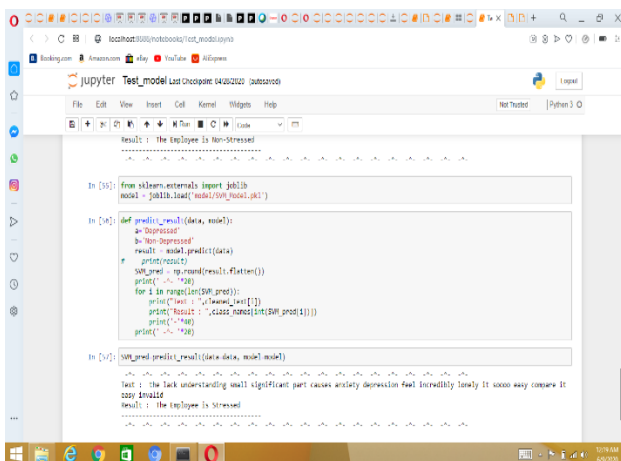


Figure 2: Output of stressed employee

## VII. CONCLUSION

Even though various mental health schemes were provided to their employees by employer the stress levels controlling is too far away. There are various factors contributing to stress such as gender, family history of illness. From our survey, it was found that employees working in tech industry were tilted more towards stress even if their role was not IT based work. These survey helps industry to frame better policies for their employees. The highest accuracy was obtained by ensemble methods like boosting followed by random forest. An 84.13% accuracy indicates the substantial results of the Machine Learning algorithms used for prediction of the Stress levels and the different mental health conditions.

## VIII. REFERENCES

[1] U Srinivasulu Reddy, Aditya vivek Thota, Adharun,” Machine learning Techniques for Stress prediction in working

employee” 2018 IEEE International Conference on computational intelligence and computing research.

[2] G. Mikelsons, M. Smith, A. Mehrotra, M. Musolesi, “Towards deep learning models for psychological state prediction using smartphone data: Challenges and opportunities, 2017.

[3] Samrat Bisai, Richa Chaudhary, “Stress Among the Students of an Engineering Institution in India: An Empirical Analysis”, Jindal Journal of Business Research, Volume: 6 issue: 2, page(s): 186-198 Article first published online: October 30, 2017; Issue published: December 1, 2017

[4] Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, “User-level psychological stress detection from social media using deep neural network”, 2014, ACM, New York, NY, USA, 507-516

[5] C. Liao, R. Chen, and S. Tai, Emotion stress detection using EEG signal and deep learning technologies - IEEE Conference Publication, 2018 IEEE Int. Conf. Appl. Syst. Invent., no. 2, pp. 9093, 2018

[6] Chigerwe et al. BMC Medical Education 2014, Published on: 28 November 2014

[7] Gender and Stress. (n.d.). Retrieved from APA press release 2010.

[8] A.P. Singh and H.C. Singh, “Occupational Stress, Security and Insecurity with Job Involvement of First Level Industrial Supervisors”, Indian Journal of Industrial Relations, Vol. 20, No. 2, 1984, pp. 177-183

[9] T.R. Rajeswari, “Employee Stress: A Study with Reference to Bank Employees”, Indian Journal of Industrial Relations, Vol. 27, No. 4, 1992, pp. 419-425

[10] Ashwinkumar.U.M and Dr. Anandakumar K.R, "Predicting Early Detection of cardiac and Diabetes symptoms using Data mining techniques", International conference on computer Design and Engineering, vol.49, 2012

[11] Takashi Shimizu and Shoji Nagata, “Relationship between Job Stress and Self Rated Health among Japanese Full-Time Occupational Physicians”, Environmental Health and Preventive Medicine, Vol. 10, 2005, pp. 227-232.

[12] Jay P. Mulki, Jorge F. Jaramillo and William B. Locander, “Effect of Ethical Climate on Turnover Intention: Linking Attitudinal and Stress Theory”, Journal of Business Ethics, Vol. 78, 2008, pp. 559-574.