iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35177

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Predictive Analysis for Real Time Stress Detection

Srivathsa Bharadwaj K S[1], Shrihari Bhat[2], Siddesh S[3], Dr. Ganashree T S[4]
[1, 2, 3, 4]*Department of Telecommunication Engineering, Dayananda Sagar College of Engineering*

Abstract: ***Stress can be characterized as a feeling of either emotional or physical tension. Due to the biology of the human body, it releases some hormones when under stress. These hormones might cause tensed muscles, increase pulse rate or the heart rate, increase brain activity to make the brain more alert to the surrounding. Stress can be predicted well before it happens by constantly measuring the Heart Rate Variability (HRV) parameters obtained using the pulse sensor. In this project a supervised machine learning model is created using the data acquired from Physionet, once the data is acquired it is cleaned and the missing data is filled. This data set is later used to create a random forest classifier and is saved using pickle library. Once the model is created it is used to detect stress in real time. Pulse sensor amped is used to get the required pulse data in the form of a CSV file and a numpy array is created using inter beat interval information got from pulse sensor. Once a numpy array is created neurokit2 library is used to extract the HRV information of the R-R interval. Later these parameters are compared with the created model and checked to see if the subject is stressed, if the model detects the subject as stressed an alerting message is sent to the subject's smartphone using Twilio.***

Keywords: Supervised Machine Learning, Heart Rate Variability, Pulse sensor, Bio Signals, R-R interval.

I. INTRODUCTION

Stress is generally described as a feeling of pressure and strain. In our culture, stress is a rising issue. It is part of our daily life, and it is endured by many people. We spend most of our time in the office, often with heavy workloads and time pressure, which helps raise our levels of stress. Major causes for stress are being under a lot of pressure, faced with major changes or/and worrying about something. Stress is the secondary most common health issue in Europe related to work, followed by musculoskeletal disorders, which in some cases can also be a symptom of stress. Some causes of stress are cataclysms/crisis, big life occurrences, micro stressors/daily annoyances and stressors in the environment. Bio signals: Any signal in living creatures that can be continuously measured and monitored is referred to as a bio signal. Bio signal is a term that is frequently used to refer to bioelectrical signals, although it can also apply to non-electrical signals.

A. Electrocardiogram

This is a technique used to measure the heart activity of a subject by connecting electrodes to skin. The electrocardiography measures the electrical activity of the heart.

B. Photoplethysmogram (PPG)

A PPG is a technique used to measure the blood volume changes and, also used to measure the heart rate. A pulse sensor uses optical techniques to measure the heart rate and/or blood volume changes.

C. Predictive analysis

Predictive analysis is a technique of predicting the future outcome of a system by utilizing the currently available data.

D. Machine Learning

Machine learning (ML) is a means of computers figuring out how to do things without being categorically computed to do so. Machine learning algorithms create a model based on specimen, referred to as "training data," in order to make judgments without being explicitly programmed. Machine leaning models can be classified as supervised machine learning models and unsupervised machine learning models.

1) *Supervised Machine Learning Models:* Supervised machine leaning models use the labelled data while being trained. A supervised models examine the sample data and generates an inferred operation that can be applied to fresh data. The algorithms will be able to accurately regulate the class labels for imaginary examples in the best-case scenario. Few of the examples include Random Forest Classifier, Support Vector Machine, and Gradient Boosting Classifier etc.
2) Unsupervised Learning Models: Unsupervised learning (UL) is a kind of algorithm that uses untagged data to learn patterns. Unsupervised models are superior to supervised models in terms of processing capabilities, but it can become unpredictable compared to the latter.

II. LITERATURE SURVEY

A. Heart Rate Variability as an Index of Differential Brain Dynamics at Rest and After Acute Stress Induction: Tara Chand et al [1]
This paper talks about how heart rate varies before and after stress induction. The brain and the body are interconnected by the Dynamic network, structural network, and functional network. As the name says brain dynamics is being monitored when the person is stressed and when he is not. For this they selected few test subjects and experimented for two days by giving them placebo treatment. On both the days the participants were made to carry out arithmetic calculations and mental rotations under time pressure with a jury panel watching. The anxiety and stress levels were measured and there was substantial increase in nervousness after stress induction.

B. A Novel Wearable EEG and ECG Recording System for Stress Assessment: Joong Woo Ahnet al [2]
This paper tells us about a wearable device used to measure ECG and EEG. This device has active electrodes which are mounted on left and right sides at the bottom of mastoids so that EEG both hemispheres of brain can be measured. The participants were made to take Stroop color word test and arithmetic test as tools of stress induction. The findings of this experiment show that in stress scenario the right alpha power diminished to a greater degree than the left alpha power. The brain releases alpha waves which in normal state is 8-12Hz while creative thinking it is more depending on how much the person is trying to utilize his/her brain. This is alpha power.

C. Stress Detection in daily life Scenarios using smart phones and wearable sensors: A survey: Can et al [3]
This paper basically talks about various devices used to measure heart rate and brain activity using ECG and EEG (electro encephalography). Stress also impacts muscles. By placing electrodes on selected muscles, the muscle activity during stress is measured. This is called electromyogram (EMG). EEG is measured using devise whose electrodes are placed on the scalp. And ECG by placing electrodes on arms, wrists, shoulders, and hips. Today we have electronic watches and smartphones which can detect heart rate.

D. Heart Rate and High Frequency Heart rate Variability During Stress as Biomarker for Clinical Depression. A Systematic Review: Schiweck et al. [4]
This paper talks about Traumatic stress and how major depressive disorder affects a person's heart rate. The heart rate of the MDD patients in laboratory-controlled environment is measured to see the variability in heart rate and to see how they respond to stress. ECG was used to monitor their heart rate. The study shows hypo reactivity in majority of the stress cases due to depression and fluctuation in heart rate at high frequency. Hypo reactivity is people having hard time reacting to stimulus resulting in slower reactions.

E. Beat-to-beat ECG features for time resolution improvements in stress detection: Dustin Axman et al [5]
In this paper they are using heartbeat morphology for stress sensing. Heartbeat morphology is classifying heartbeats into 5 different classes. The detected heartbeat comes under one of these. They were using Tier Social stress test to induce stress. In this method the participants are made to give an interview style presentation where the panel does not give feedback or encouragement. Throughout the experiment, ECG signals were monitored using a vital jacket.

III.METHODOLOGY

A. Software used

1) *Python:* Python is an open source high level language which can be used for web development, data analytics, data visualization, machine learning etc.
2) *Arduino IDE:* Arduino can be coded using an open source software called Arduino IDE, either C or C++ can be used to code the Arduino microcontroller.

B. Hardware used

1) *Arduino Uno:* Arduino Uno is a type of micro controller which has Harvard architecture, it is based on ATmega328P Microchip. It can be coded using Arduino IDE which is an open source software.
2) *Pulse sensor:* Pulse sensor is a type of sensor which can be used to detect pulse signal and can be used to measure the heart rate of a subject.

C. Packages used to build the stress detector

1) *Neurokit2:* NeuroKit is a Python module that offers high level integrative functions with good and configurable defaults, allowing users to concentrate on what matters. It is a Python tool for neurophysiological signal processing that is open-source, community-driven, and user-centered. It has many processing procedures for a wide range of body signals.
2) *BioSPPy:* BioSPPy is a Python-based toolkit for bio signal processing. The toolkit combines a variety of signal processing and pattern recognition approaches aimed towards bio signal analysis. It supports various bio signals like BVP, ECG, EDA, EEG, EMG, Respiration and can filter and perform frequency analysis on these signals.
3) *pyHRV:* This is a package in Python for bio signal processing. This package has functions for the computation of time domain and frequency domain parameters of a bio signal. It also has functions to compute nonlinear HRV parameters which can be extremely helpful during researching on HRV.
4) *NumPy:* Python has a module called NumPy. It is a multidimensional array object library with a set of array processing routines.
5) *Pandas:* Pandas is an open-source, BSD-approved Python library that provides high-performance, easy-to-use Data Structures and Data Analysis tools for Python.
6) *Scikit-Learn:* Scikit-learn offers a standard Python interface for a variety of supervised and unsupervised learning techniques. Scikit-learn is based on SciPy (Scientific Python).

IV. IMPLEMENTATION

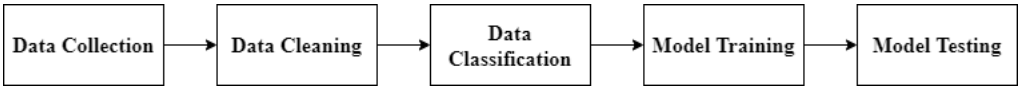


Fig. 1 Model Creation

First step in any machine learning process is the data collection part. In this we have used data from Physionet which is a repository of freely available medical research data, which is managed by MIT Laboratory. The data comes from another github repository as well, which is created by Christopher Andre and Ottensen. The data has ECG, EMG, GSR values as well as other heart related parameters like heart rate, HRV, RR interval etc. After the data is collected it has to be cleaned to required format, some manual cleaning of the data is required, and the missing data is filled by the mean of the available data. Christopher Andre and Ottensen has used the GSR values from leg to classify the available data as stressed or not stressed. This is an important step in the process as we will be using supervised machine learning and for supervised machine learning labelling of the available data is an important step. Once data cleaning and data labelling is done, we can move on to the next step that is model training. Different types of supervised ML models exist, depending on the data each model gives different label of accuracy. For the current data Gradient Boosting Classifier (GBC) and/or Random Forest Classifier(RFC) gives higher level of accuracy for the test data which comes around 85%.
Sklearn library is used to create GBC or RFC model, the data is divided into 80% train data and 20% test data. The particular columns which are used from the dataset as X_train and x_test are mean NN interval, RMSSD, pNN50, ULF, VLF, LF, HF and LF/HF. Y label is stress label. This is run using python with GBC values set to learning rate of 0.01 and max depth of 90.
The RFC values are set with criterion as entropy and max features of 5 and max depth of 100. The model hence created gives an accuracy of around 85%. Using Sklearn's classification report we can check the metrics with accuracy of 86% and F1 score of 86%. The model which is got after training and testing is saved using pickle library, this will later be used to analyze in real time.

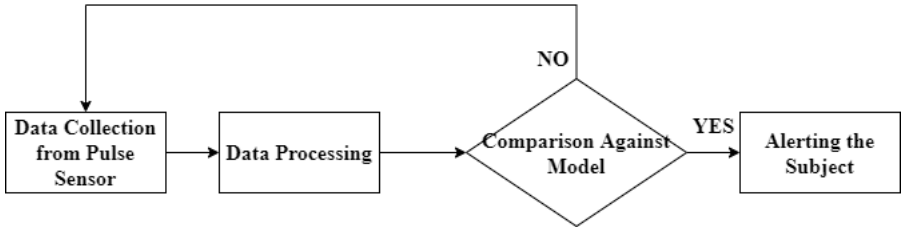


Fig. 2 Real Time Analysis using Created Model.

Pulse sensor and Arduino Uno are used to collect the real time data. Pulse sensor library available in the Arduino IDE is used to interface the pulse sensor. The heart rate and inter beat intervals are calculated using this library and is sent to a serial port using Arduino IDE, this is connected to a python program using the Serial library. This program creates a CSV file in a particular location and appends heart rate and IBI coming through the serial port.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

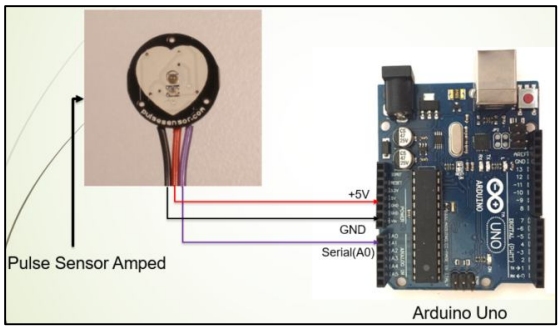


Fig. 3 Connection of pulse sensor to Arduino

For using this information to compare against the model HRV parameters must be extracted. To do this, first the CSV file has to be extracted into a numpy array and this can be done using np.genfromtxt function. To skip the header of the file a parameter called skip_header can be used to set the number of rows to be skipped.

Once the IBI information is extracted to a numpy array, we have to get the RR interval from this information, to do this we can use np.cumsum to calculate the cumulative sum of the IBI. This information gives the RR interval which can later be used to calculate the HRV parameters using the neurokit2 library. We must make sure that the cumulative sum is calculated since the first beat is detected, as there can be fault in the sensor.

Once the RR interval is saved in a numpy array, we can use an especially useful bio signal processing python library neurokit2 to extract the HRV parameters by calling the hrv function from the neurokit2 library and passing the RR interval numpy array as the signal for this function.

After doing this step we will get the HRV parameters in the form of pandas data frame. This data frame is passed as X parameter and is tested against the created model. Since the data frame as all the columns used in training steps, the code runs without any error and gives if the value passed is stressed (1) or not stressed (0). If the model detects that the subject is stressed, then an alerting message is sent to the subject using an online application called Twilio.

V. RESULT

```
[10]: def do_gbc():
        X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                            train_size=0.80, test_size=0.20)
        gbc = GradientBoostingClassifier(n_estimators=150, learning_rate=0.01,max_depth=90, random_state=0)
        gbc.fit(X_train, y_train)
        print(gbc.score(X_test, y_test))

        return gbc, X_test, y_test
      gbc_classifer, X_test, y_test = do_gbc()

      0.8571428571428571

[11]: def do_rfc():
        X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                            train_size=0.80, test_size=0.20)
        rfc = RandomForestClassifier(criterion='entropy',max_features=5, n_estimators=200,max_depth=100, random_state=10)
        rfc.fit(X_train, y_train)
        print(rfc.score(X_test, y_test))

        return rfc, X_test, y_test
      rfc_classifer, X_test, y_test = do_rfc()

      0.8619854721549637
```

Fig. 4 Model Creation

Fig. 4 shows the model creation using GBC and RFC. As the accuracy is more for these two models, these are used during real time analysis. The metrics seen after the model creation shows the accuracy of the respective models.

```
[13]: from sklearn.metrics import classification_report

      report = classification_report(pred,y_test)

      print(report)
```

	precision	recall	f1-score	support
0	0.85	0.86	0.86	394
1	0.87	0.87	0.87	432
accuracy			0.86	826
macro avg	0.86	0.86	0.86	826
weighted avg	0.86	0.86	0.86	826

Fig. 5 Classification report of the model

Fig. 5 shows the classification report of the model create. Classification report includes the accuracy, recall, precision and F1 score of the model. Fig. 5 shows that the model created gives around 85% for all the metrics.

```
[6]: filename = "Final_Model.pkl"
     model = pickle.load(open(filename, "rb"))
     y_pred_test = model.predict(df)
     y_pred_test

[6]: array([0])

[7]: if(y_pred_test == 1):
         account_sid = "ACf31035bccdeb2bfb79ebf90dff601232"
         auth_token = "9d23ae0b1b7a14c2894b45d8578647b9"

         client = Client(account_sid, auth_token)

         message = client.api.account.messages.create(
                         to='+919900559465',
                         from_='+13184371013',
                         body="You are stressed, please take rest")

         print("You are stressed, please take rest")
     else:
         print("Not stressed")

     Not stressed
```

Fig. 6 Model Detecting Signal as Not Stressed

A resting signal was downloaded from neurokit2 library and HRV function was run on this signal to get the HRV parameters, after getting these parameters the data frame was compared with the trained model. The trained model detected this signal as not stressed, as this was the resting data set.

If the model had detected this signal as stressed, then an alerting message would have been sent to the subject using an online API called Twilio.

```
[10]: filename = "Final_Model.pkl"
      model = pickle.load(open(filename, "rb"))
      y_pred_test = model.predict(df)
      y_pred_test

[10]: array([1])

[11]: if(y_pred_test == 1):
          account_sid = "ACf31035bccdeb2bfb79ebf90dff601232"
          auth_token = "9d23ae0b1b7a14c2894b45d8578647b9"

          client = Client(account_sid, auth_token)

          message = client.api.account.messages.create(
                          to='+919900559465',
                          from_='+13184371013',
                          body="You are stressed, please take rest")

          print("You are stressed, please take rest")
      else:
          print("Not stressed")

      You are stressed, please take rest
```

Fig. 7 Model Detecting Signal as Stressed

The CSV file got using python and serial library is loaded into a numpy array and necessary calculations are done to this array and HRV parameters are found out. Once these parameters are found out, it is run against the model to check if the subject is stressed. The above figure shows that the subject is stressed as y prediction is giving the value as 1.

Once the model detects the signal is stressed signal, it initiates Twilio to send a message to the user using personal authentication ID and authentication token.

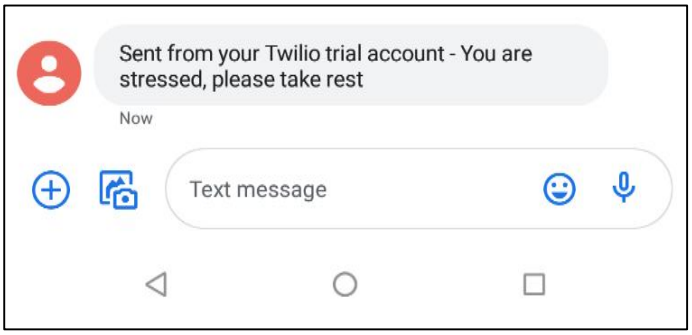


Fig. 8 Message Received

Fig. 8 shows the message the user receives when they are stressed. Once the model detects that subject is stressed, it sends a message to the user.

VI. CONCLUSION AND FUTURE SCOPE

The dataset was successfully collected, and cleaning of this data was successfully achieved. The dataset was successfully labelled as stressed or not using the value of GSR sensor. Supervised models like GBC and RFC were successfully created with an accuracy of over 85% using the cleaned dataset. Once the model was created by dividing the available dataset into 80% training data and 20% testing data, 20% testing data was run against the created model to get the accurate classification report of the model. Model was successfully saved using pickle, which will be used later during testing.

The ECG and pulse sensors signals were successfully compared to check if they yielded similar HRV parameters using pyHRV library. The resting signal was successfully downloaded from the neurokit2 library and the HRV parameters of this was calculated using neurokit2 hrv function. The calculated parameters were successfully compared with the previously trained model and checked to see if the subject was stressed. The pulse sensor was successfully interfaced using Arduino and the IBI data was successfully saved as a CSV file. The CSV file was used to get the IBI numpy array which was later used to get the HRV parameters using the neurokit2 library. These HRV parameters from the pulse sensor was used to compare between the model and data and was found that the data is stressed data. A message was successfully sent to the user smartphone using Twilio application once the model detected that the subject is stressed.

The data downloaded does not have speed data. This data can be added later for getting more accurate results. The dataset has some of the missing data which makes the model vulnerable to error, this can be avoided by having more accurate data. The dataset does not have enough data points to perform deep learning, this can be added in the future. The accuracy of the model can be increased by increasing the data points available. The data got during real time can be improved by having more expensive and accurate sensors.

REFERENCES

[1] Chand T, Li M, Jamalabadi H, Wagner G, Lord A, Alizadeh S, Danyeli LV, Herrmann L, Walter M, Sen ZD. Heart Rate Variability as an Index of Differential Brain Dynamics at Rest and After Acute Stress Induction. Front Neurosci. 2020 Jul 2;14:645. doi: 10.3389/fnins.2020.00645. PMID: 32714132; PMCID: PMC7344021

[2] Ahn, J.W.; Ku, Y.; Kim, H.C. A Novel Wearable EEG and ECG Recording System for Stress Assessment. Sensors 2019, 19,1991.https://doi.org/10.3390/s19091991S.

[3] Can YS, Arnrich B, Ersoy C. Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. J Biomed Inform. 2019 Apr;92:103139. doi: 10.1016/j.jbi.2019.103139. Epub 2019 Feb 27. PMID: 30825538.

[4] Schiweck C, Piette D, Berckmans D, Claes S, Vrieze E (2018). Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. A systematic review. Psychological Medicine 1–12. https://doi.org/10.1017/ S0033291718001988.

[5] D. Axman, J. S. Paiva, F. de La Torre and J. P. S. Cunha, "Beat-to-beat ECG features for time resolution improvements in stress detection," 2017 25th European Signal Processing Conference (EUSIPCO), 2017, pp. 1290-1294, doi: 10.23919/EUSIPCO.2017.8081416.

[6] A. Singh, N. Thakur and A. Sharma, "A review of supervised machine learning algorithms," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp. 1310- 1315.

[7] Forte, Giuseppe, Francesca Favieri, and Maria Casagrande. "Heart rate variability and cognitive function: a systematic review." Frontiers in neuroscience 13 (2019): 710.

[8] Shaffer, Fred, and J. P. Ginsberg. "An overview of heart rate variability metrics and norms." Frontiers in public health 5 (2017): 258.

[9] Burkart, Nadia, and Marco F. Huber. "A survey on the explainability of supervised machine learning." Journal of Artificial Intelligence Research 70 (2021): 245-317.

[10] Elgendi, Mohamed, and Carlo Menon. "Machine learning ranks ECG as an optimal wearable biosignal for assessing driving stress." IEEE Access 8 (2020): 34362- 34374.

[11] Taelman, Joachim, Steven Vandeput, Arthur Spaepen, and Sabine Van Huffel. "Influence of mental stress on heart rate and heart rate variability." In 4th European conference of the international federation for medical and biological engineering, pp. 1366-1369. Springer, Berlin, Heidelberg, 2009.

[12] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

[13] Villarejo, María Viqueira, Begoña García Zapirain, and Amaia Méndez Zorrilla. "A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee." Sensors 12, no. 5 (2012): 6075-6101

[14] OTTESEN, CHRISTOPHER. "Tutorial: Stress Detection With Wearable Devices And Machine Learning".Dataespresso,2019.https://dataespresso.com/en/2019/01/30/Stress-detection-with-wearable-devices and-Machine-Learning/.

[15] J. Bakker, M. Pechenizkiy and N. Sidorova, "What's Your Current Stress Level? Detection of Stress Patterns from GSR Sensor Data," 2011 IEEE 11th International Conference on Data Mining Workshops, 2011, pp. 573- 580, doi: 10.1109/ICDMW.2011.178.

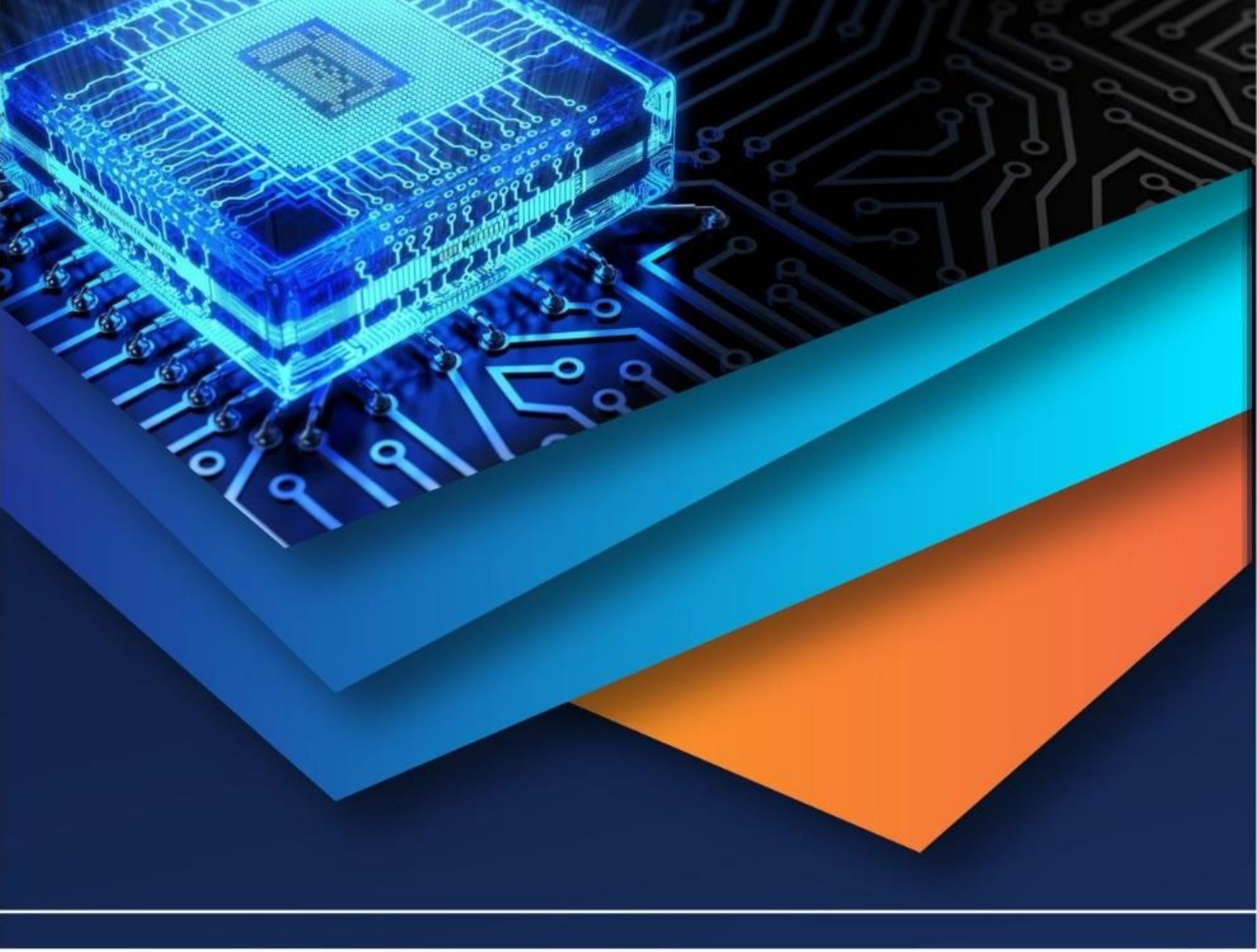
[16] Ernst, Gernot. "Heart-rate variability—More than heart beats?." Frontiers in public health 5 (2017): 240.

[17] P. Gomes, P. Margaritoff, and H. P. da Silva, “pyHRV: Development and evaluation of an open-source python toolbox for heart rate variability (HRV),” in Proc. Int’l Conf. on Electrical, Electronic and Computing Engineering (IcETRAN), pp. 822-828, 2019

[18] Lu G, Yang F, Taylor JA, Stein JF. A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. J Med Eng Technol. 2009;33(8):634-41. doi: 10.3109/03091900903150998. PMID: 19848857.

[19] Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lesspinasse, F., Pham, H., Schölzel, C., & S H Chen, A. (2020). NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing. Retrieved March 03, 2021.

[20] R. K. Nath, H. Thapliyal, A. Caban-Holt and S. P. Mohanty, "Machine Learning Based Solutions for Real Time Stress Monitoring," in IEEE Consumer Electronics Magazine, vol. 9, no. 5, pp. 34-41, 1 Sept. 2020, doi: 10.1109/MCE.2020.2993427.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY