International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

STUDENT STRESS DETECTION THROUGH EXPLORATORY DATA ANALYSIS AND MACHINE LEARNING

MS. P.Sindhuri [a] ***, G.Obuliraju*** [b] ***,T.Prakash*** [c] ***,B.Rithika*** [d] ***,S.Thiru Praveen*** [e]

[a] Assistant Professor Faculty of Engineering and Computer Science, Dhirajlal Gandhi College of Technology, India. Students in the Department of Computer Science and Engineering, b, c, d, e ,Dhirajlal Gandhi College of Technology, India.

ABSTRACT:

Understanding the stress levels that can impair our social and personal wellbeing requires effective stress management. The World Health Organization estimates that one in four people suffer from psychological disorders brought on by stress, which can result in mental and financial difficulties, strained relationships at work, and in extreme situations, suicide. One essential tool for helping people manage stress is counselling. Although it is impossible to completely avoid stress, stress can be managed with the help of preventive measures. At the moment, the only people who can tell if someone is under stress are medical and physiological professionals. However, the conventional approach to stress detection, which relies on people's self-reported responses, is not accurate. A more accurate method is to use physiological cues to automatically detect stress levels.

Keywords: Stress Detection, Natural Language Processing(NLP),Heart Rate Variability.

1. INTRODUCTION

Stress is an unavoidable part of life that leads to negative emotional states, particularly for people who spend a lot of time in front of computers. Therefore, it is essential for people's safety to keep an eye on their mental state in such circumstances. In order to make the man-machine interface more adaptable and user-friendly, a camera is placed to take a close-up view of the person working in front of the computer. Automated age assessments lack the unique information that human experts have about visual traits that signify aging, such as smoothness, face structure, wrinkles, inflammation, and under-eye bags. Asymmetric data can be used to improve the trained model's generalizability in order to solve this problem. Predicting mood levels is the goal of the suggested model.

The system is able to provide targets for any new input by the users after sufficient training. It also compares the output with the correct, intended output and find errors in order to modify the model accordingly.

2.1 THE FEASIBILITY OF WEARABLE AND SELF-REPORT STRESS DETECTION MEASURES IN A SEMI-CONTROLLED LAB ENVIRONMENT

The viability of self-report and wearable stress detection techniques in a semi-controlled laboratory setting In order to evaluate the viability and dependability of employing wearable sensors in conjunction with self-report measures for stress detection in a semi-controlled laboratory setting, Aristizabal et al. (2021) carried out an extensive investigation. Twenty-five healthy adults participated in the study, and they were given both non-stressful baseline settings and a variety of stress-inducing tasks, including the Trier Social Stress Test and the Stroop Color-Word Test. Two wearable devices—the Polar H10 chest strap and the Empatica E4 wristband—were used to gather the physiological data. Indicators like skin temperature, movement, heart rate, heart rate variability, and electrodermal activity (EDA) were measured by these devices.

2.2 EARLY LIFE STRESS DETECTION USING PHYSIOLOGICAL SIGNALS AND MACHINE LEARNING PIPELINES

Early Life Stress Identification Using Machine Learning Pipelines and Physiological Signs .The goal of the study "Early Life Stress Detection Using Physiological Signals and Machine Learning Pipelines" is to determine how early life stress (ELS) affects people by using physiological biosignals and cutting-edge machine learning techniques. Early detection is essential because early life stress, which is frequently brought on by negative childhood events, has long-term repercussions on both physical and mental health. This study highlights the use of objective biomarkers for stress detection that are non-invasive, such as respiration rate, electrodermal activity (EDA), and heart rate variability (HRV). To model stress patterns and distinguish between people with and without a history of ELS, machine learning pipelines that include preprocessing, feature extraction, selection, and classification are employed.

2.3 HEART RATE VARIABILITY-BASED MENTAL STRESS DETECTION: AN EXPLAINABLE MACHINE LEARNING APPROACH

Detecting Mental Stress Based on Heart Rate Variability: An Explainable Machine Learning Method The paper "Heart Rate Variability-Based Mental Stress Detection: An Explainable Machine Learning Approach" looks into how heart rate variability (HRV) features can be used to identify mental stress, with an emphasis on using explainable machine learning models. The change in the time intervals between successive heartbeats, or HRV, is a commonly used physiological indicator of stress response and autonomic nervous system activity. In this study, subjects' HRV signals are recorded under both stress-inducing and baseline (calm) settings. To record stress-related alterations, a variety of time-, frequency-, and non-linear HRV characteristics are retrieved.

2.4 STRESS DETECTION SYSTEM FOR SOCIAL MEDIA USERS

The use of machine learning, behavioral analysis, and natural language processing (NLP) techniques to detect psychological stress in online activity is the main focus of recent research on stress detection for social media users. People frequently share their feelings and mental states on social media sites like Facebook, Reddit, and Twitter, which are rich sources of user-generated material. In order to identify stress indicators, these studies usually analyze text posts, posting frequency, engagement patterns, and even active time. In order to train machine learning models, lexical features including the use of negative words, first-person pronouns, and emotional tone are retrieved together with information.

2.5 A STRESS DETECTION METHOD FOR METAL COMPONENTS BASED ON EDDY CURRENT THERMOGRAPHY

A METAL COMPONENT STRESS DETECTION METHOD BASED ON EDDY CURRENT THERMOGRAPHY The research with the ID: 15 Quotes found: 0.36 percent The study "A Stress Detection Method for Metal Components Based on Eddy Current Thermography" examines a non-destructive testing (NDT) method for locating faults and stress concentrations in metallic compounds. In order to identify surface and subsurface irregularities, Eddy Current Thermography (ECT) integrates the concepts of infrared thermography and eddy current induction. This technique produces localized heating by using alternating current to create eddy currents on the conductive metal surface. Thermal imaging uses variations in heat distribution to infer material fatigue, microcracks, or stress concentrations.

3 SYSTEM STUDY

3.1. EXISTING SYSTEM

CURRENT SYSTEM The current technology relies on digital signal processing for stress detection, which considers skin temperature, pupil dilation, blood volume, and galvanic skin reactivity. Other studies on this subject use a range of physiological cues and visual cues (such as head movement and eye closure) to gauge an individual's level of stress at work. However, in reality, these measures are painful and intrusive. The amount of stress is calculated by comparing each sensor reading to a stress index, which is a numerical value.

3.1.1 DISADVANTAGES

Physiological signals are frequently pigeonholed for analysis by non-stationary temporal performance, and the stress index of the physiological signals is directly revealed by the extracted characteristics. Finding a consistent pattern to describe the stress emotion is challenging since various people may react or express themselves differently under stress. The widely utilized peak j48 approach is used to analyze the ECG signal right away.

3.2 PROPOSED SYSTEM

On the dataset, we applied the deep neural network technique. Instead of requiring manually created features, deep neural networks use the layers of the neural networks to extract features from raw data. Neural network testing and training were conducted using the datasets. To determine the machine's correctness, the neutral networks algorithm is applied.

3.2.1 ADVANTAGES

Advanced deep learning characteristics make it simple to assess stress. It reduces workplace stress, and businesses don't need to spend money on stress management. Higher productivity and a more harmonious workplace are the outcomes of properly identifying and managing stress among employees.

4 SYSTEM SPECIFICATIONS

4.1 HARDWARE REQUIREMENT:

- RAM : 4 GB RAM
- Hard disk : 80 GB Hard Disk
- Processor : Above 2GHz Processor

4.2 SOFTWARE REQUIREMENTS

- Language - Python

- OS - Windows 10
- Tool - Anaconda navigator

5 MODULES DESCRIPTION

5.1 LIST OF MODULES

- Collection of Dataset
- Balancing of Data
- Prediction of Stress

5.2 MODULES DESCRIPTION

5.2.1 COLLECTION OF DATASET

Initially, we collect a dataset for our human stress prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing.

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Pre-processing of data is required for improving the accuracy of the model.

5.2.2 BALANCING OF DATA

There are two methods for balancing datasets. They are both over- and under-sampled.

(a) Under Sampling: In Under Sampling, the size of the abundant class is decreased in order to achieve dataset balance. When there is sufficient data, this procedure is taken into consideration.

(b) Over Sampling: In Over Sampling, the size of the limited samples is increased in order to achieve dataset balance. When there is insufficient data, this procedure is taken into consideration.

5.2.3 PREDICTION OF STRESS

In this article, classification is accomplished using five distinct machine learning algorithms. The EDA algorithms have been compared and contrasted in this investigation. Finally, a EDA algorithm with the highest rate of accuracy for stress with each parameters of dataset and hence shown in which parameter the stress is high and slow.

6. SYSTEM ARCHITECTURE

The architecture for human stress prediction using mobile and wearable device data is composed of a structured, layered pipeline that begins with data collection and culminates in stress prediction using machine learning. The process starts at the **data acquisition layer**, where data is collected from smartphones and wearable devices. These data sources provide detailed information such as call logs, SMS records, Bluetooth activity, app usage statistics, physical activity patterns, sleep behavior, and heart rate. In addition to sensor-generated data, survey-based inputs related to perceived stress levels are also gathered to serve as ground truth or labeling information for supervised learning tasks.

Once the raw data is collected, it is forwarded to the **data preprocessing layer**, which is a critical stage for ensuring data quality and consistency. Here, several preprocessing operations are performed, including the elimination of missing values, removal of redundant data, normalization of feature scales, and necessary data transformations to make the data suitable for machine learning algorithms.

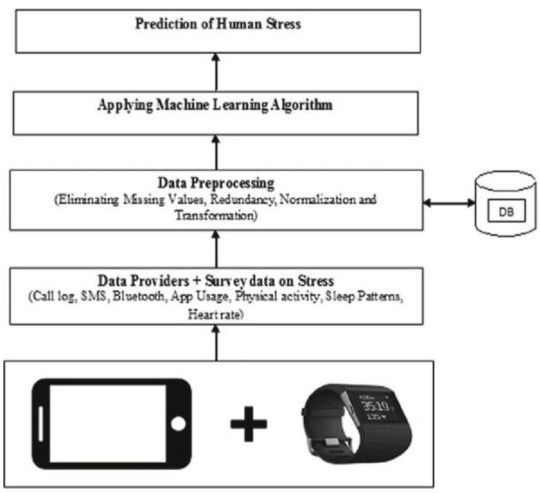


Figure 6.1.1 System Architecture

7. SYSTEM IMPLEMENTATION

The procedure that actually produces the lowest-level system components in the hierarchy is called implementation. These components can be manufactured, purchased, or repurposed. Forming, removing, connecting, and finishing hardware fabrication operations, coding and testing software realization processes, and operational procedure creation processes for operator roles are all included in production. A manufacturing system that makes use of the existing technical and management procedures can be necessary if implementation entails a production process. Designing and building (or fabricating) a system element that complies with its design specifications and/or attributes is the aim of the implementation phase. The component is built using industry standards and relevant technologies. This procedure connects the integration and system definition procedures.

The results of the implementation process may produce constraints that are applied to the architecture of the higher-level system, depending on the technologies and systems selected when a decision is made to produce a system element. These constraints are typically identified as derived system requirements and added to the set of system requirements applicable to this higher-level system. These limitations must be considered in the architectural design.

7.1 Algorithm Explanation

7.1.1 Exploratory Data Analysis

Data scientists utilize exploratory data analysis (EDA), which frequently uses data visualization techniques, to examine and analyze data sets and highlight their key features. It makes it easy for data scientists to find patterns, identify anomalies, test a hypothesis, or verify assumptions by assisting in figuring out how to best manipulate data sources to obtain the answers you require. EDA gives a better knowledge of the variables in the data collection and their relationships, and is mostly used to investigate what data might reveal beyond the formal modeling or hypothesis testing assignment. It can also assist you in assessing the suitability of the statistical methods you are thinking of using for data analysis.

7.1.2 Exploratory data analysis tools

With EDA tools, you may carry out specific statistical operations and methods like dimension reduction and clustering, which aid in producing graphical representations of high-dimensional data with several variables. summary statistics and a univariate depiction of every field in the raw dataset. You can evaluate the link between each variable in the dataset and the target variable you're examining by using bivariate visualizations and summary statistics.

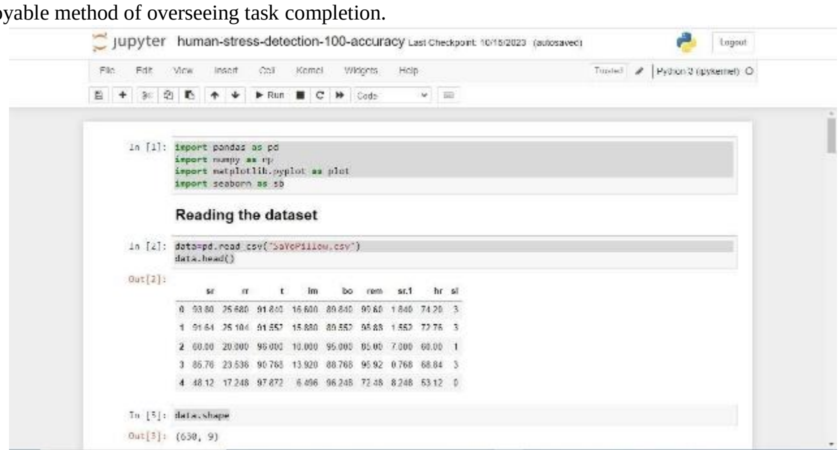
8.CONCLUSION AND FUTURE ENHANCEMENTS

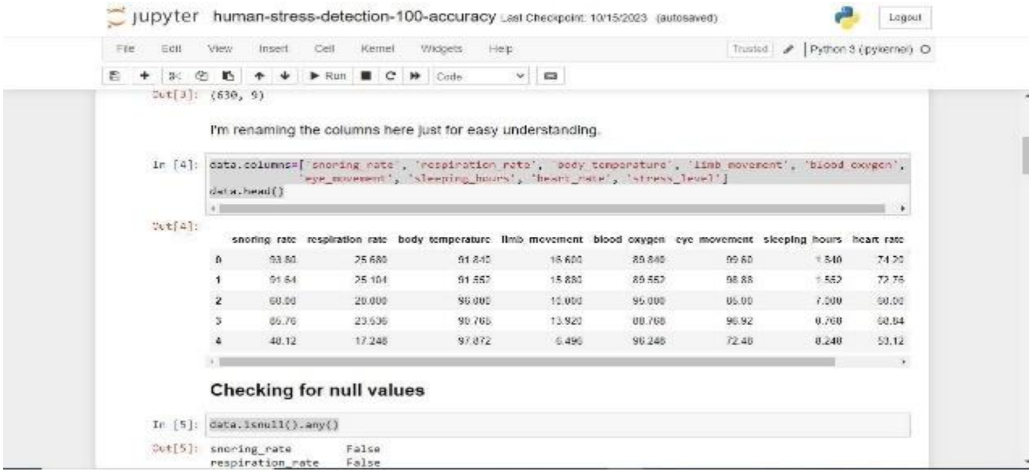
8.1 Conclusion

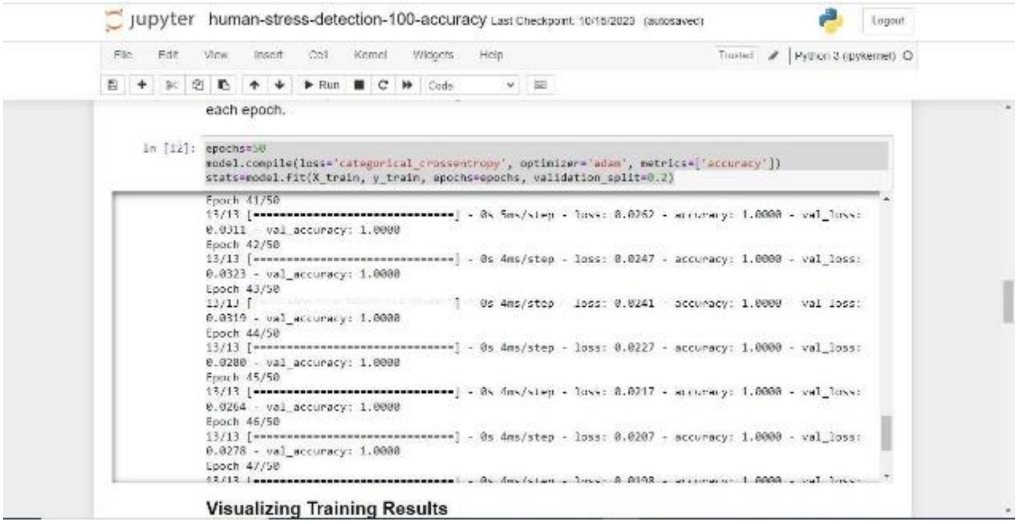
Human stress detection is a crucial tool for preventing stress and protecting the emotional and physical well-being of those who experience it. The use of stress-reduction and stress-detection strategies benefits corporate sector workers. The organization has the capacity to motivate workers to work effectively, reap the greatest rewards from company expansion, and entice new hires to stay for a longer period of time. Deep learning technology can predict stress with high efficiency and accuracy. A Network Model for Prediction is produced by utilizing a number of open-source tools, including TensorFlow, Google Colab, and a few Python modules. To produce precise predictions, the convolutional algorithm is trained using the visual data and labels.

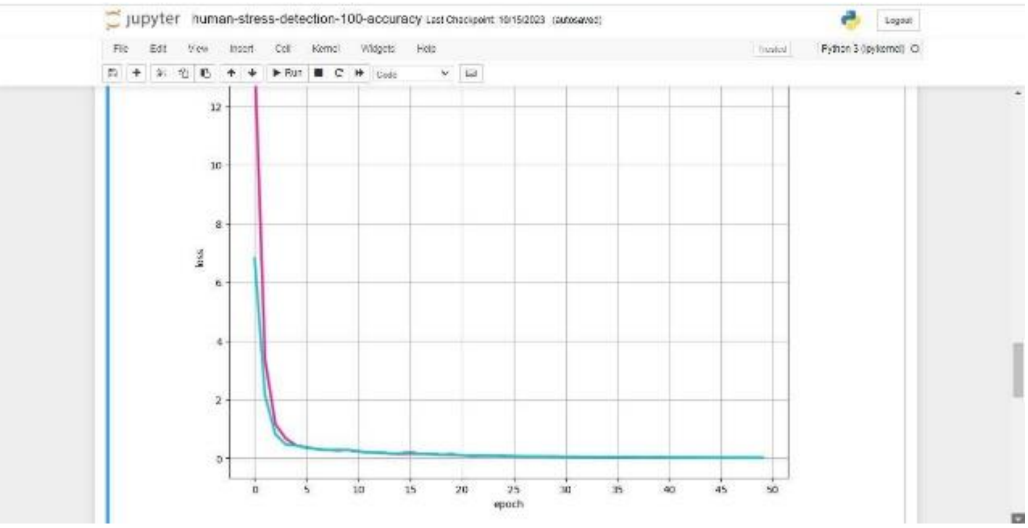
8.2 Future Work

Since we use modular implementation, updating algorithms is simple, and future work on changing the model's implementation may be undertaken. By introducing the new future that uses audio processing techniques to detect tension, we can also make the project better. Playing calming music can also help to prolong the use of stress detection by lowering employee stress and alerting the management to take the necessary actions to lessen workload and have a profitable and enjoyable method of overseeing task completion.

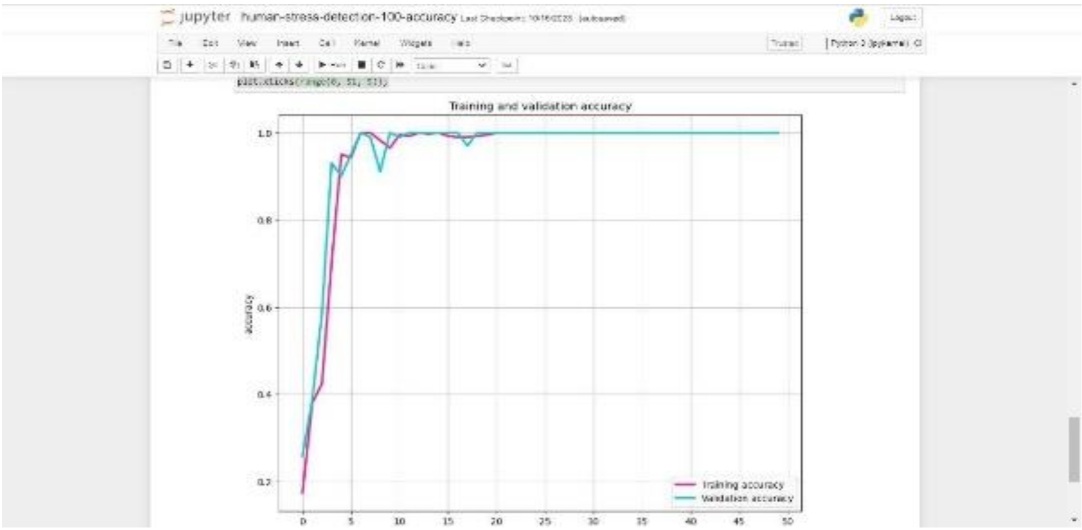


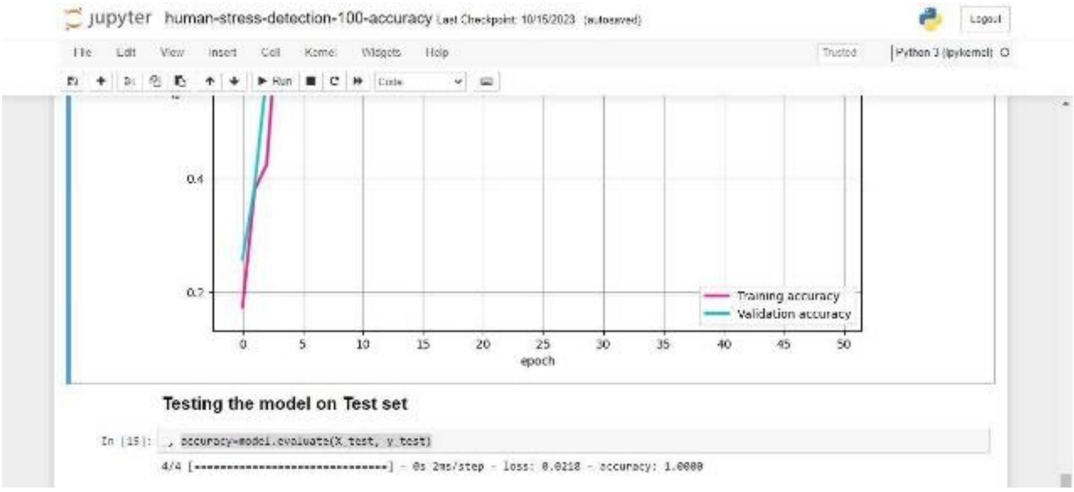

jupyter human-stress-detection-100-accuracy Last Checkpoint: 10/15/2023 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Out[3]: (630, 9)
I'm renaming the columns here just for easy understanding.
In [4]: data.columns=['snoring rate', 'respiration rate', 'body temperature', 'limb movement', 'blood oxygen', 'eye_movement', 'sleeping_hours', 'heart_rate', 'stress_level']
data.head()
Out[4]:
snoring rate respiration rate body temperature limb movement blood oxygen eye movement sleeping hours heart rate
0 93.80 25.680 91.840 16.600 89.840 99.60 1.840 74.20
1 91.64 25.104 91.552 15.880 89.552 98.88 1.552 72.76
2 60.00 20.000 96.000 10.000 95.000 85.00 7.000 60.00
3 85.76 23.536 90.768 13.920 88.768 96.92 0.768 68.84
4 48.12 17.248 97.872 6.496 96.248 72.48 8.248 53.12
Checking for null values
In [5]: data.isnull().any()
Out[5]: snoring_rate False
respiration_rate False


jupyter human-stress-detection-100-accuracy Last Checkpoint: 10/15/2023 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
each epoch.
In [12]: epochs=50
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
stats=model.fit(X_train, y_train, epochs=epochs, validation_split=0.2)
Epoch 41/50
13/13 [==============================] - 0s 5ms/step - loss: 0.0262 - accuracy: 1.0000 - val_loss: 0.0311 - val_accuracy: 1.0000
Epoch 42/50
13/13 [==============================] - 0s 4ms/step - loss: 0.0247 - accuracy: 1.0000 - val_loss: 0.0323 - val_accuracy: 1.0000
Epoch 43/50
13/13 [==============================] - 0s 4ms/step - loss: 0.0241 - accuracy: 1.0000 - val_loss: 0.0319 - val_accuracy: 1.0000
Epoch 44/50
13/13 [==============================] - 0s 4ms/step - loss: 0.0227 - accuracy: 1.0000 - val_loss: 0.0280 - val_accuracy: 1.0000
Epoch 45/50
13/13 [==============================] - 0s 4ms/step - loss: 0.0217 - accuracy: 1.0000 - val_loss: 0.0264 - val_accuracy: 1.0000
Epoch 46/50
13/13 [==============================] - 0s 4ms/step - loss: 0.0207 - accuracy: 1.0000 - val_loss: 0.0278 - val_accuracy: 1.0000
Epoch 47/50
Visualizing Training Results


jupyter human-stress-detection-100-accuracy Last Checkpoint: 10/15/2023 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Python 3 (ipykernel)
Run
Code
12
10
8
6
4
2
0
loss
0
5
10
15
20
25
30
35
40
45
50
epoch

+





REFERENCE

[1] Aristizabal, Sara, et al. "The feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment." IEEE Access 9 (2021): 102053-102068.

[2] Shahbazi, Zeinab, and Yung-Cheol Byun. "Early Life Stress Detection Using Physiological Signals and Machine Learning Pipelines." Biology 12.1 (2023): 91.

[3] Banerjee, Jyoti Sekhar, Mufti Mahmud, and David Brown. "Heart Rate Variability-Based Mental Stress Detection: An Explainable Machine Learning Approach." SN Computer Science 4.2 (2023): 176.

[4] Febriansyah, Mochamad Rizky, Rezki Yunanda, and Derwin Suhartono. " stress detection system for social media users." Procedia Computer Science 216 (2023): 672-681.

[5] Zu, Ruili, et al. "A stress detection method for metal components based on eddy current thermography." NDT & E International 133 (2023): 102762.

[6] Kuttala, Radhika, Ramanathan Subramanian, and Venkata Ramana Murthy Oruganti. "Multimodal Hierarchical CNN Feature Fusion for Stress Detection." IEEE Access (2023)

[7] Phukan, Orchid Chetia, et al. "An Automated Stress Recognition for Digital Healthcare: Towards EGovernance." Electronic Governance with Emerging Technologies: First International Conference, EGETC 2022, Tampico, Mexico, September 12–14, 2022, Revised Selected Papers. Cham: Springer Nature Switzerland, 2023.

[8] Nijhawan, Tanya, Girija Attigeri, and T. Ananthakrishna. "Stress detection using natural language processing and machine learning over social interactions." Journal of Big Data 9.1 (2022): 1-24

[9] AlShorman, Omar, et al. "Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection." Journal of Integrative Neuroscience 21.1 (2022): 20.

[10] Bin Heyat, Md Belal, et al. "Wearable flexible electronics based cardiac electrode for researcher mental stress detection system using machine learning models on single lead electrocardiogram signal." Biosensors 12.6 (2022): 427.