DEVELOPMENT OF A MODEL FOR STRESS PREDICTION USING

SUPPORT VECTOR MACHINE AND GRID SEARCH-CROSS VALIDATION TECHNIQUE

BY

JIBRIN, Aisha

MTech/SICT/2019/9849

DEPARTMENT OF COMPUTER SCIENCE

FEDERAL UNIVERSITY OF TECHNOLOGY

MINNA

JULY, 2023

DEVELOPMENT OF A MODEL FOR STRESS PREDICTION USING

SUPPORT VECTOR MACHINE AND GRID SEARCH-CROSS VALIDATION TECHNIQUE

BY

JIBRIN, Aisha

MTech/SICT/2019/9849

A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA, NIGERIA IN PARTIAL

FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE

DEGREE OF MASTER OF TECHNOLOGY IN COMPUTER SCIENCE

JULY, 2023

ABSTRACT

Stress is a state of worry or mental strain caused by an unpleasant situation. Stress is one of the most common causes of mental health issues. A little worry does not affect mental health, but when the worry festers it can be harmful to mental health. Bad mental health leads to decline in physical health. Models that predict stress require high prediction accuracy rate to be able to determine if an individual is stressed or not. Stress is measured to get data values that can be input into a system. Stress prediction models require high prediction accuracy rate to be able to determine if the data that is input into the system is stressed or not. A stress management system must first be able to predict stress before it can provide a way to relieve it. The models developed for stress prediction tend not to yield results with high accuracy. This study proposed the development of a model to predict stress. Grid search-cross validation (CV) technique is used to finetune Support Vector Machine (SVM). Grid Search-CV tries different parametric combination to fine-tune SVM using Cross Validation method to choose the combination with the highest accuracy percentage. The results obtained from the optimised SVM model used in this study yielded an accuracy of 99.9% compared to the existing model which gave an accuracy of 98.2%. It is recommended that any organization that requires a lot of thinking in their operations that could cause stress like knowledge workers can adopt the model to have workers or employees that will perform more efficiently.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Tables **Page**

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the Study

When the human body is confronted with a demand or obstacle, it experiences stress. The term stress is often used, yet it is difficult to define because stress is subjective, and defining a situation that signals stress is unclear. According to research carried out by Park *et al* (2021), humans are by nature, becoming ambitious nowadays and seek every possible opportunity to grow professionally. Anxiety, depression, stress frustration and dissatisfaction have become commonplace that people now believe them to be part and parcel of professional life (Priya *et al.*, 2019).

Stress detection can aid the mitigation of stress only after the subject has become stressed. A method of predicting stress in advance could improve stress mitigation strategies by allowing stress mitigation to begin before the subject enter a stressed state (Clark *et al.*, 2021) . Generally, stress may be triggered by the discrepancy between the situational demands and the individual's inability to leverage complicated situations. Extreme stress minimizes the efficiency of work and leads to numerous illnesses and negative emotions. Continuous stress damages several internal organs, and this results in various disorders. Simple stress management techniques can help an individual feel less stressed while also reducing health risks. These are some of the most significant advantages of stress reduction; it aids in the elimination of wasted energy, stored resiliency is maintained, anger, irritation, and impatience are all reduced. Increased access to intuition and improved physical and mental wellness. According to research carried out by Park *et al.* (2021), one of the most important factors in mental health is stress.

According to Figueroa and Aguilera (2020), worry and distress can be harmful to anyone suffering from anxiety disorders. Anxiety disorder is a type of mental health disorder characterized by worry or nervousness. Adults find it challenging to avoid feeling anxious and worried all the time when the pressure of succeeding is on, which results in poor choices like, lack of attention, and a host of other unfavourable problems. Intense fear and panic symptoms that are immediately brought on by a physical health issue are signs of anxiety disorders that are caused by medical issues. Gender is widely acknowledged as one of the most important factors to consider in stress studies, but this is not always the case.

Activities that could cause bad stress are; Lack of funds, not having a goal to work towards, lack of sleep, poor eating habits, or too much free time. Extreme stress can make it difficult to work effectively, as well as cause poor professional achievement and reduced focus. Individuals who have gone through stressful life events experience poor health and a lower quality of life. Individuals are faced with situations that cause frustration, nervousness or any strong negative emotion and that makes the body react to it with stress. The stress can be good when it helps in working under pressure and meeting deadlines, but when it causes a negative reaction in a persons' mental health it needs to be managed. Mental health is an example of a critical health area in dire need for technological solutions to enable timely and effective interventions (Wang *et al.,* 2020).

Some educational sectors have minimal educational resources, thus no effort is put into developing mental health education programs (Shen, 2021). There are people who do not exhibit any mental illness, some might have and do not know, and those who have and do not want to be stigmatized, so mental health awareness or aid put in place is hardly used. Accurate prediction of stress aids in relieving stress. Physical, mental and network mental health are all elements in the health of college individuals in vocational colleges (Shen, 2021). When there is a problem with physical health, help is easy to seek for as opposed to when the mental health declines seeking for help is hard.

Importance is placed on physical health and less on any type mental health.

A person's mental health is related to their cognitive, behavioural, and emotional well-being. Manav *et al.* (2020) worked on how to help determine a person's mental state. The paper analyses the different techniques and algorithms used in the project to derive the results. The World Health Organization reports that mental health issues are one of the World Health Organization's priorities. In particular, preventing these problems is clearly an effective solution. Mental illnesses are vast and more are being discovered. Mental health is an example of a critical health area in desperate need of technological solutions that will allow for timely, effective, and scalable interventions (Mahsa *et al.*,

2020).

1.2 Statement of the Problem

The demand that comes with the desire to fulfil familial and personal obligations results in mental stress. A lack of motivation, energy, and time are the key barriers to improved stress management. The complexity of modern society is increasing the importance of stress management. This leaves room for different kinds of mental disorders. Wang *et al.*

(2020) believe that stress management is essential for individuals. Therefore, this study aims to develop a model that can accurately predict stress in individuals. In the study carried out by Pankajavalli *et al.* (2021), numerous lifestyle-related illnesses are brought on by today's sedentary lifestyle. This has prompted the mission to foresee illnesses before they happen. According to the research, stress prediction has traditionally been conducted in a laboratory setting. However, the development of non-invasive strategies for stress prediction using wearable devices has been the focus of recent research. Due to the highly subjective nature of stress patterns and the fact that they vary from person to person, the models developed for stress prediction do not typically yield accurate results.

1.3 Aim and Objectives

The aim of this research is to develop a machine learning model for predicting stress, in order to actualize this aim the following objectives are to be achieved;

i) To preprocess the data downloaded.

ii) To finetune the data using SVM classifier with grid search CV.

iii) To develop a model using the best parametric combination.

iv) To evaluate the performance of the developed model in (iii) with metrics like accuracy, precision and recall.

1.4 Scope of the Study

A stress management system must first be able to predict stress before it can provide a way to relieve it. The models developed for stress prediction tend not to yield results with high accuracy. This research work focuses on the development of a stress prediction model with a machine learning algorithm and an exhaustive search to obtain the optimal prediction.

1.5 Significance of the Study

Bad stress with no knowledge of it could lead to physical health problem and could be bad enough that a person may develop depression which is not good for mental health growth (Shen, 2021). This research will provide a new addition to already exiting stress prediction models using different combination of methodologies to achieve a higher accuracy.

In research or study, determining the problem is always the first step. Predicting accurately if a subject is stressed or not will determine the next step in stress management. Although individuals can use the Internet to gain learning tools and engage in thought-provoking discussions with others, information, both good and bad, can have a harmful impact on individuals' mental health (Shen, 2021). Not being able to filter information accessed from the internet, both negative and positive information might have an unknown impact on a person's psyche.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Stress

Stress is characterized as an emotional or physical condition of a person. The body reacts with stress when faced with a task or demand. Stress is a term used to describe a person's physical or psychological response to the demands of daily life. Common sources of stress include work responsibilities, financial independence, social obligations, and health. Because almost all activities can cause stress, it is impossible to avoid stress entirely (Park *et al.,* 2021). Being distressed affects a person both mentally and physically. According to Figueroa and Aguilera (2020), a sudden change in the environment and mental health can result in a significant increase in the need for psychological support. Which festers and manifests as mental illness that will require more time, energy and money if it not treated correctly. Ignoring the level of stress a person is under does not make it go away. Sometimes it represents as a headache to signal an issue with the body.

Recognizing the huge potential of stress in disease prevention and treatment enhances quality of life. Moreover, low-cost wearable gadgets for vital sign monitoring and stress detection are readily available. Heart rate variability (HRV), a particular vital sign, is also used to gauge stress levels; several wearable devices gather this information. The article written by (Jin *et al.,* 2021) suggested a framework for real-time measurement of stress levels in particular female athletes.

Stress, can affect individuals' subjective well-being, leading to poor grades and dropouts. Given the negative effects of stress on academic and psychological performance in college individuals, it is important to study predictors of stress. In the study conducted by Karaman *et al.* (2019), a conceptual framework was used consisting of factors theoretically and empirically related to academic stress to examine differences between male and female individuals.

Skin conductance that is sweat forms the basis of many physiologically based emotion and stress detection systems. However, such systems typically do not recognize the biomarkers present in sweat and thus do not utilize the biological information of sweat. Similarly, such systems do not detect volatile organic compounds (VOCs) produced under stress conditions. Zamkah *et al.* (2020) provides an overview of the current state of bio markers of human emotional stress and proposes important potential bio markers for future wearable sensors of the emotional system. Emotional stress is considered one of the main causes of several social problems related to crime, health, economy and even quality of life.

Knowledge workers are a group of people who suffer from chronic stress. Due to the recent deterioration of the domestic economy, stress in the workplace is on the rise. For this reason, efforts are needed and methods to reduce stress. Park *et al.* (2021) describes how to effectively manage workplace stress for knowledge workers to ensure productivity and efficiency, it also introduces a mental health management system that specifically identifies if they are stressed or not, and collects physiological sensor data like e electroencephalogram (EEG), heart rate variability (HRV), blood volume pulse (BVP), electromyogram (EMG), and galvanic skin response (GSR) to recommending the best stress relief solution for any worker that might need it.

Can *et al.* (2020) believes an automatic stress detection system that employs unobtrusive smart bands will benefit human health and well-being by mitigating the effects of high stress levels. However, there are a number of challenges for detecting stress in unrestricted daily life, which results in lower performances of such systems when compared to semi-restricted and laboratory environment studies.

According to Figueroa and Aguilera (2020) worry and distress can be harmful to people suffering from anxiety disorders. Anxiety disorder is a type of mental health disorder characterized by worry or nervousness. Adults find it challenging to avoid feeling anxious and worried all the time when the pressure of fulfilling their goals is on, which results in poor choices, lack of attention, and a host of other unfavourable problems. These symptoms may be present in those who have anxiety disorders. Anxiety, restlessness, or agitation, fatigue easily, difficulties focusing, quick temper, headache, muscle discomfort, abdominal pain, or any inexplicable pain. Having trouble managing worry and having trouble falling or staying asleep.

According to the research done by Gupta *et al.* (2021) on World Economy Skills and Agro Development (WESAD) by various researchers, each person acts differently even under similar circumstances, and their bodies show different physiological changes for stress, amusement, depression, and neutral state. A person's behaviour in society depends on upbringing, status, and genetics. Genetic drives activate different people in different ways.

Suppose a grumpy person reacts angrily to a situation and finds himself in an awkward position. Behaviour also provides insight into an individual's psyche and is driven in part by thoughts and feelings that reveal attitudes and values. Activation of the stress response generates a cascade of potentially harmful physiological responses when repetitive or chronic and when triggered after exposure to psychological/emotional stressors.

Kim *et al.,* (2021) worked on Brain to Music: Musical representation from stress-induced EEG. Four participants aged 19 to 24, all of whom were right-handed, took part in an experiment to collect EEG data under stressful and non-stressful circumstances. Many measurement studies employing stress evaluation techniques have been carried out since stress has a detrimental effect on the mind and body. While self-stigma makes anxiety of interacting with specialists visible, public stigma towards stress-related illness makes people afraid to seek therapy. As a result, individuals suffering from mental illnesses may choose not to seek treatment from mental health professionals. Stress research has increased due to a rise in the need for mental health services. Self-stigma is the negative beliefs that people with mental illness have regarding their illness and can be caused by lack of understanding from friends, family, and coworkers, lack of opportunities to work, attend school, participate in social events, or find housing, and lack of health insurance coverage (Kaga & Kato, 2019).

The quality of the family environment has a direct impact on children's physical and mental health. There is no doubt that the main body of family education plays an important role in individuals' mental health education. An adult with a strong mental and emotional upbringing from their family is able to navigate and understand the causes of stress and is better equipped to manage it on their own, whereas an adult with a weak mental state will be easily overwhelmed and find it difficult to cope. Having a correct way to relieve their stress will stop any negative thoughts and coping mechanism (Jing, 2021).

For Nigerians in particular, mental health issue becomes apparent only when a person starts behaving insane. Mental disorders are widespread and have a significant negative impact on functioning and quality of life (Choudhury *et al.,* 2021). Because mental disorders are common, it appears acceptable and normal for a person to be unaware of what is going on with them. People may respond to stress by breathing more quickly, and ethical behavior can be more likely to occur under stressful conditions. Mental diseases like persistent stress and depression can also be comprehended (Thomas *et al.,* 2020).

2.1.1 Causes of Stress

Stress can be caused by any event that causes an individual to react in any way or act against what makes them feel comfortable. The development of big data, cloud computing, and other technologies in the information age has an effect on individuals' mental health. The network environment has a significant impact on individuals in the

Internet age due to the rapid development of Internet information technology. Individuals can obtain learning resources from the Internet and brainstorm with others, but both good and bad information can have a negative impact on an adult's mental health. Because individuals are relatively self-confident, have weak discernment, and lack social experience, network information can change the inner mental world of individuals (Shen, 2021).

Although individuals can obtain learning resources from the internet and engage in thought-sparking with others, information that is both good and bad can have a negative impact on individuals' mental health. The lifestyle of 21st century youths has changed dramatically, particularly since the development of computer networks. Their values and ways of thinking are very different from those born in previous eras. From these aspects, the introduction of networks can help individuals learn more effectively and help individuals learn more knowledge better. individuals can use technology in most areas of the curriculum to improve their overall engagement and content comprehension (Shen, 2021).

A person's mental health is related to cognitive, behavioural, and emotional well-being. Simply put, it can be used to describe how people feel, act, think and how they deal with situations. Mental health is an important aspect of a person's well-being. Several factors are detrimental to mental health, including stress and anxiety. If people suffering from mental illness are not treated properly, the consequences can be disastrous. The culmination of today's technology has allowed the use of the power of computers to predict the onset of such mental illnesses. The advancement of technology has made many previously impossible things possible, but with its benefits also comes some drawbacks, such as screen addiction and poor time management. Social media plays a complicated role in mental health management (Figueroa & Aguilera, 2020). The allure that social media presents and the dopamine the users get every time they use it presents a bit of challenge when they try to get help for themselves. There are links between social media use and mental illnesses such as: Increased incidence of anxiety, stress, depression, body image issues and loneliness in teens and young adults.

Internet can be used to communicate between people and obtain massive amounts of information. As a result, the personalities of college individuals are frequently more lively, confident, and unrestrained. Having massive and unlimited amounts of information at their fingertips drastically reduces the time it would take them to search for information. Their values and ways of thinking are very different from those of previous generations. Though the Internet, is a two-edged sword, when used the right way it can make life easier and faster. Helping the users achieve their goals faster. While it is convenient and abundant in resources, it has a negative impact on the mental health of college individuals in higher vocational colleges. The pressure of social competition is extremely high in the new era (Shen, 2021).

Relying on network technology to complete social behaviours does not necessitate direct contact between two or more people in real life. This type of social networking is somewhat unstable. Undergraduates who have not been immersed in the world and have little social experience are easily duped. This type of illusion has the potential to easily immerse unconscious college students in negative thoughts, expand the negative psychological influence, and cause negative consequences (Zhaorong & Yiwen, 2019).

Youths frequently dislike communicating with teachers and parents who are not of their generation (Shen, 2021). Teachers and parents are in positions of authority, so discussing what is causing stress may add to the stress. individuals may be scared to go to a teacher because they get sent to the hospital and most doctors just give them different types of anti depressants that will not solve the underlying problem but treat the symptoms. It also causes issues like addiction. Some family cannot afford the cost of an hospital bill.

Going to college is a turning point in a young person's life. However, Trigueros *et al.,* (2020) found that previous research focused on adult behaviour in the classroom rather than on adult lifestyle impacts. The purpose of the study was to analyse the impact of college students' emotional intelligence on resilience, study stress, test anxiety, and diet in the context of a college-level Mediterranean diet.

2.2 Stress Management

Lin and Yang (2020) studied research on the value of physical education in school to the physical & mental health development of teenagers. Teenagers are the country's future, and their development is linked to the fate and future of the Chinese people. As a result, the government, society, schools, and families have paid close attention to the physical and mental health of teenagers. There are many challenges when it comes to data management, especially in mental health care organizations, due to the relative complexity of the data. The purpose of the paper presented by Tumpa *et al.* (2020) is to focus on decision making using BI to improve patient services, explore different definitions of BI, and explore them using a framework to help businesses to better understand the role of implementing intelligence.

The cost of mental health care professionals' services from psychotherapists to clinical psychologists and psychiatrists varies by country and is further influenced by country regulations and subsidies. However, the cost to the patient is primarily determined by the number of practising professionals available in a given country. When communicating with artificial systems, people show lower evaluation in fears, sadness, and objectively rated disclosure. This means that by introducing therapeutic options that they perceive to be safer and stigma-free, the number of people who do not seek treatment can be reduced (Kolenik & Gams, 2021). However, there are potential risks associated with such technologies that must be recognized and addressed in order for persuasive technology to reach its full potential. Mental health care is in dire straits given that more than 60 million Nigerians suffer from various mental illnesses and only about 10% of them receive adequate care. This leaves more than 90% without access to medical care, and this group is known as the mental health treatment gap. There are various factors that cause gaps, such as knowledge gaps where people are under-informed about the causes and treatments of mental illness. There are several factors, such as myths and traditional beliefs, that hinder the way mental illness is dealt with in Nigeria. Inadequate mental health facilities and large numbers of mental health professionals.

A computer network refers to interconnected computing devices that can exchange data and share resources. These connected devices use a set of rules called communication protocols to transfer information through physical or wireless technology. Computer network technology has become an integral part of people's daily lives. As an early adopter group, computer network technology has greatly impacted their daily lives. Similarly, the advent of computer networking technology has changed the way people express good and bad feelings. Therefore, it is necessary to combine the progress and development of the times to apply computer network technology to the work of college adult mental health counselling. Yan (2020) presented a research that explains the issues that should be addressed and make reasonable suggestions.

Hasanbasic *et al.* (2019) studied Recognition of stress levels among individuals with wearable sensors. Stress is an inevitable part of a adult’s life. The intense pressure of achieving high grades, maintaining a social life, and dealing with financial obligations while away from family home, present daily challenges for college individuals. This places a great deal of emphasis on passing every exam, whether in written form or during the presentation. Individuals can benefit from early recognition of stress by providing them with various stress coping mechanisms. Wearable sensors enable real-time monitoring of individuals' physiological signals, which can be linked to stress.

Can *et al.* (2020) investigated Real-Life Stress Level Monitoring using Smart Bands in the Light of Contextual Information. An automatic stress detection system that employs unobtrusive smart bands will benefit human health and well-being by mitigating the effects of high stress levels. However, there are a number of challenges for detecting stress in unrestricted daily life, which results in lower performance of such systems when compared to semi-restricted and laboratory environment studies. Schools are constrained by their own circumstances. Some colleges and universities have limited educational funds and make little investment in the development of mental health education courses

(Shen, 2021).

Aristizabal *et al.* (2021) wrote on the feasibility of wearable and self-report stress detection measures in a semi-controlled lab environment. Workplace-related stressors, economic strain, and lack of access to educational and basic needs have exacerbated feelings of stress in the United States. Continuous stress can increase the risk of cardiovascular, musculoskeletal, and mental health disorders. Similarly, workplace stress can lead to lower employee productivity and increased costs associated with employee absenteeism in an organization. Detecting stress and the events associated with stress during the course of a workday is the first step toward addressing its negative effects on health and well-being.

Using mental health applications help individuals think in the right direction and inspire them to adopt correct behaviours, but it also provides one-time assistance to individuals who are experiencing similar psychological confusion and improves the efficiency of mental health education. It also allows individuals to self-test through on-line psychological testing, which greatly reduces counselling workload and solves the problem of insufficient coordination between the number of psychological counsellors and the number of individuals. The use of the internet to carry out mental health education can overcome time and space constraints and provide individuals with assistance at any time and from any location (Yijun, 2020).

While health is one of the fundamental needs underlying various quality of life indicators, many health care organizations continue to lag behind in terms of effective data management and handling when compared to other sectors, such as the financial sector. For example, in the treatment of mental health, predicting a patient's needs is critical.

Data management can generate information that can be used to make better decisions (Tumpa *et al.*, 2020). Using large amounts of information allows researchers and medical professionals to spot patterns that are normally hard to spot. The types of data collected for mental health can range from moods, communication logs, social activity to GPS data.

Collected data for a particular patient helps us know the patient's hospitalization status. Researchers can collect large amounts of information relatively quickly and easily.

Mental health applications are used to unwind, keep track of their moods, practice mindfulness, self-care, or develop healthy habits. The applications are also used in conjunction with professional treatment. The application was used by users because it makes them more positive, happy, self-conscious, calm, fun, focused, relaxed, motivated, mindful, self-controlled, or sleep better. Importantly, they value the ability to Figure out what is wrong with their health, to keep track of their progress, to see a link between the causes and effects of their health problems, to conduct self-evaluation and self-reflection, to learn good things, to form good habits, and to provoke, reframe, and organize their thoughts. A person's mental health reveals their emotional, psychological, and social well-being. It has an impact on how a person thinks, feels, and reacts to a situation. Individuals with good mental health are more productive at work and reach their full potential. It is necessary for living a healthy, balanced life. One's mental health influences their thoughts, behaviour, and emotions. It can have an impact on an individual's productivity and effectiveness (Laijawala *et al.,* 2020).

2.3 Prediction Models

Priya *et al.* (2020) collected data from episodes of depression, anxiety, and stress were predicted at five levels of severity by five different machine learning algorithms. The authors claim that algorithms are very accurate and are especially good at predicting mental health issues.

A person's emotional, psychological and social well-being are reflected in their mental health. It affects how a person thinks, feels, and handles situations. Positive mental health helps people work productively and reach their full potential. Mental health is important at every stage of life, from childhood to adulthood. Many factors contribute to mental health problems that lead to mental illness, including stress, social anxiety, depression, obsessive-compulsive disorder, substance addiction, work problems, and personality disorders. Laijawala *et al.* (2020) planned to implement classification algorithms such as Decision Trees, Random Forests and Naive Bayes. The target group was the working class or anyone over the age of 18. Once the model was created, it was integrated into the website so that the results can be predicted according to the details provided by the user.

Sulistiana and Muslim (2020) performed validation and evaluation to measure the performance of the proposed method. The dataset was split into training and testing data. Training data used to train the model. Then, the model is tested using testing data. The results obtained from this process used to measure model performance.

Clark *et al.,* (2021) used a data set that consisted of physiological data from 17 subjects as they completed an approximately 20-mile driving route through Boston. The data analysis of the stress prediction model was divided into three phases; data preprocessing, feature extraction, and feature expansion. The model was trained by fitting a random forest classifier with 100 estimators using the Gini function and a maximum depth of 30 to the training set. The train and test sets were selected using the leave-one-subject-out testing approach, meaning that a drive was selected for the test set with the remaining drives composing the train set. After the model was evaluated, a different drive was selected for the test set such that each drive was selected once. The approach was repeated with each of the four expanded feature sets. The model was then evaluated based on accuracy classifying the training and test data, and F1-score on the test data. The precision, recall, and F1-score of the model on the test data were also measured for the high and low-stress categories.

A variety of independent variables has been used in generating prediction models. To simplify presentation, we have chosen to classify these variables as presage variables or in-progress variables. Presage variables refer to those variables that are available or determined before modeling is initiated; for example, high school Grade Point Average (GPA) or math background. However, in-progress variables are measures gathered in the context of the outcome being assessed; for example, midterm grade or timings of assignment submissions. Model evaluation techniques vary in the literature as well.

Some researchers use descriptive modeling, where relationships between independent and dependent variables are studied on the training set itself. This is useful for studying relationships in past data, but is not intended for making predictions about future occurrences. For such predictions, researchers use predictive modeling, where part of the data (the test set) is held back and used to evaluate the accuracy of the model (Clark *et al.*, 2021) .

2.4 Review of Machine Learning Algorithms

In order to reach judgements at various levels, machine learning decision tree approaches use a tree data structure. Because it is simple to understand and has a steady structure, it is appropriate for prediction challenges. Using a randomly chosen subset of the training dataset, a random forest classifier generates several decision trees. Due to its superior classification capabilities and presentation quality, SVM is now widely utilized in applications where the data is split into two different classes (also known as hyperplanes) and the distance between the two classes is maximized. Artificial neural network is a type of supervised learning algorithm. The network basically learns the relationship between input and output by adjusting each neuron's weight according to the data set. Artificial neural network is a type of supervised learning. Basically, the network learns the relationship between input and output by the adaptation of the weight of each neuron depending on the data set (Priya *et al.*, 2020) .

2.5 Support Vector Machines (SVM)

Support vector machine (SVM) is a supervised machine learning method that can be used to categorize data points in terms of classification and regression. Given a set of training data that is clearly classified to its own category. SVM looks for a hyperplane to divide the data points into two classes and assign them a classification based on whether they are located on the positive or negative side. Sulistiana and Muslim (2020) describes how SVM operates, Grid Search is used to find the hyperplane with the greatest margin between two classes during an experiment. A probabilistic model is used to categorize data during the testing phase. Cross validation is used to adjust the SVM parameter. Grid search seeks to locate the ideal set of hyperparameters so that the classifier can accurately predict the unknown data.

By transforming the data into a high-dimensional feature space, SVM may categorize data points even when they are not otherwise linearly separable. The data are changed to allow for the hyperplane representation of the separator once a dividing line between the categories has been found. It performs poorly when the target classes overlap and there is greater noise in the data set. When there are more attributes for each data point than there are training data samples, the support vector machine will not perform well. According to Jin *et al.* (2021), use of Support Vector Machine (SVM) yields great accuracy by previous studies for stress identification from a methodological perspective. Kavitha *et al.* (2019) investigated an improved feature selection and classification of gene expression profile using SVM. Support Vector Machine (SVM) is the most widely used classifier for performing the classification of a massive dataset.

2.5.1 Application areas of SVM

Klyueva (2019) worked on enhancing the multi-class SVM classification's quality through feature engineering. Binary classification issues can be successfully resolved with the SVM classifier. Yet, there are frequently instances in real classification issues where the initial data set contains more than two classifications of items. With the aid of tools from well-known multiclass classification algorithms like the Decision Tree, Random Forest, and AdaBoost, the work tried to investigate methods for enhancing the quality of SVM classification based on the engineering of additional characteristics of objects in the original data set.

Kaga and Kato (2019) investigated Extraction of useful features for stress detection by performing mental arithmetic on various bio signals. A mental arithmetic task stress loading, simultaneous measurements of ECG, NIRS, and nasal skin temperature (NST), random forest and stepwise methods, and feature extraction and selection from bio signals were used. The model was used to investigate useful features, and it was discovered that the NST's slope was the most helpful feature for identifying stress. Sulistiana (2020) Grid search is an exhaustive search based on a defined subset of the hyper-parameter space.

Liu *et al.* (2019) investigated the Effective Data Classification via Combining Neural Networks and Support Vector Machine (NN-SVM), Iris dataset and Lead isotope dataset were used to evaluate the performance of NN-SVM. A stress loading caused by a mental arithmetic task, simultaneous measurements of bio-signals by ECG, NIRS, and nasal skin temperature (NST), and random forest and stepwise methods were used to extract and select features from bio signals. The model was used to investigate effective features and found that the slope of the NST was the most useful feature to detect stress.

2.6 Grid Search Cross Validation

Sulistiana and Muslim (2020) describe how SVM operates, Grid Search is used to find the hyperplane with the greatest margin between two classes during an experiment. A probabilistic model is used to categorize data during the testing phase. Cross validation is used to adjust the SVM parameter. Grid search seeks to locate the ideal set of hyperparameters so that the classifier can accurately predict the unknown data.

A novel grid searching method in spherical coordinate system is proposed to further strengthen it in raise angle location, which is theoretically analysed to locate angle of PD with higher resolution than distance, and circumvent its weakness in high computational cost. Simulation training demonstrates that a similar trajectory method is used to classify the mental health of college individuals, and that its classification accuracy is high. As a result, this method has good predictive value in predicting the mental health of college individuals, and the classification rules obtained can be used as schools to develop mental health (Gao & Shi, 2021).

2.7 Related Studies

A supervised machine learning chatbot for perinatal mental health care. Perinatal mental health problems (PMH) are a type of mood disorder that occurs during pregnancy and within the first 24 months after childbirth and affects pregnant women, newborns, and family relationships. These problems can appear at any stage of the mother's woman. PMH is primarily diagnosed by behavioural observation, self-report, and behavioural scale testing. Chatbots are an effective technology (Wang *et al.*, 2020).

Bu *et al.* (2021) researched pattern recognition of mental stress levels from differential RRI time series using LSTM networks. The paper reported some experimental results of a preliminary study on development of a mental stress recognition method using a deep learning-based model. Stress level recognition experiments, in which HRV data were obtained with mental stress levels induced in a virtual reality environment, have been discussed in terms of training and classification performance. Test configurations with varying numbers of classes, segmentation lengths, and datasets were investigated. When compared to the experimental results of the authors' previous studies, it was discovered that using differential RRI can significantly reduce the difficulty of convergence in the

LSTM training process.

Mental health assessment of college students based on similar trajectory clustering algorithm. In order to improve the accuracy of the method to assess the mental health of college individuals, Gao and Shi (2021) proposes a method to assess the mental health of college individuals based on the clustering algorithm of similar trajectories. A similar trajectory clustering algorithm is used to discover college adult psychology data, build a college adult mental health assessment model, and extract rules for college adult mental health assessment. Experimental results show that a similar trajectory clustering algorithm effectively classifies college individuals' mental health status and improves the accuracy of college individuals' mental health assessment.

Laijawala *et al.* (2020) implemented classification algorithms such as Decision Trees, Random Forests and Naive Bayes. The target group is the working class or anyone over the age of 18. Once the model was created, it was integrated into the website so that the results can be predicted according to the details provided by the user. Thomas *et al.* (2020) during their research found growing research has shown that using medical imaging to better understand mental illnesses such as depression and chronic stress can help. These findings, however, are not applicable in real-world clinical settings. While there are many determinants of mental health, such as socio-economic, biological, and environmental factors, assessing an individual's mental health can be difficult.

Clark *et al.* (2021) presented a model to predict a driver's stress level up to one minute in advance. If the research could better predict future stress, it might be possible to start reducing stress before the subject felt it, reducing or avoiding stress-induced performance decline. Features were extracted from the galvanic skin response (GSR) signal of the driver and respiratory and electrocardiogram (ECG) signals from the driver's chest.

The theory, technology and resources of big data are being used to improve the mental health of a community. Today, big data shows an increasingly rapid development trend and offers bright prospects. From the perspective of technical architecture, the current big data technology is mature and has important implications for the reform and development of many fields. Community is the basic unit of people's social life, so starting with community is effective in improving the mental health of Chinese people

(Haowei & Ting, 2020).

The determinants of a user's mental health are multiple, interrelated, and often multifaceted. Choudhury *et al.* (2021) uses a structural equation model to examine how internet use, perceived quality of care, patient education, and patient-centered communication affect mental health. This result suggests that increased Internet use, including for health purposes, has a negative impact on mental health. The study also found that internet use, patient-centered communication, patient education, and perceived quality of care can affect mental health. As society increasingly seeks health information from online sources, our research recommends making online health information sources more user-friendly and trustworthy.

Kolenik and Gams (2021) investigated how persuasive technology, which tries to influence people's behaviour or attitudes for their own goals without coercion, can be used in the domain of mental health to increase equality to mental health care access as well as equality to mental health care access as well as equality in general.

Recent reports indicate that in one quarter of families, at least one family member has a mental disorder. In current practice, most diagnostic methods in psychiatry are based on clinical interviews and questionnaires, which are subjective and can lead to recall and interviewer bias. In the context of healthcare, virtual reality (VR) has powerful potential to improve decision-making and help patients become more connected to reality, manage pain, and overcome psychiatric disorders such as anxiety and depression. is showing. Bryant *et al.* (2020) integrates sensor technology (such as eye tracking) with VR simulations of healthcare environments to improve clinical decision support systems for the diagnosis and assessment of psychiatric disorders. Traditional scenario-based patient simulations serve as the basis for the development of virtual reality modules.

Comprehensive healthcare services have evolved significantly in recent years, and IoMT is rapidly changing the pace and scope of healthcare delivery. A promising application of IoMT is to capture patterns of psychobehavioural symptoms based on biosignals and transmit them to appropriate hospitals or psychologists for remote monitoring. However, data volume and performance, device versatility and interoperability, hacking and abuse, and barriers to acceptance and adoption still limit the practical and appropriate use of these devices. Gupta *et al.* (2021) presents a plausible solution to overcome data overload and processing delays in real-time sensory data collected by wearable mental health monitoring devices.

Mental health represents a huge disease and social burden, and a significant part of ubiquitous computing research has been directed to developing technologies for continuous monitoring, diagnosis, and care of mental illness. The paper reviews his ten years of research on mental health technology, with a focus on the use of mobile and wearable technology. Bardram and Matic (2019) found 46 systems that were analysed in a historical context, then discussed the psychiatric disorders they covered, the type of technology, and the type and size of clinical trials in which they were used. In the paper, they present input from nine leading researchers in the field and discuss key technical and clinical challenges in the development of ubiquitous computing technologies over the next decade.

Wang *et al.* (2020) proposed a novel Mental Health Problem Prevention System (MHPPS) for the prevention and early detection of mental health problems MHPPS can collect and analyse multi-modality data (e.g., audio, video) from users in their daily lives. Liang *et al.* (2020) examined Creating a mHealth application to track academic stress and physiological reactions to stress. The paper described the creation of a mobile health application called NokoriMe, which includes the creation of an original academic stress questionnaire as well as the implementation of the application. The NokoriMe application allows individuals to measure and track stress over time, as well as visualize trends and correlations in stress and physiological responses to stress (that is sleep quality and physical activity patterns). A pilot usability study reveals that the developed application is simple to use, but there is room for improvement in questionnaire design and data visualization.

Bryant *et al.* (2020) combines sensing technology (e.g., eye tracking) with a virtual reality simulation of healthcare environments to improve the clinical decision-support system for the diagnosis and assessment of mental disorders. Nilufar *et al.* (2019) tried increasing Self-Compassion in Young People through Virtual Reality to control mental health

issues.

Virtual technologies have the potential to aid personal and clinical change. Real environments can be replaced with virtual ones, thereby allowing a transformation of our external experience. Their hypothesis is that immersive virtual reality (VR) will have significant impact on increasing participants' self-compassion and decreasing

self-criticism, paving the way for clinical use of it in the future. Since the VR technology is personally being monitored, it raises a lot of questions (Nilufar *et al.*, 2020).

Numerous studies have shown that adult procrastination is a major factor influencing adult performance in on line learning. Therefore, it is important for educators to be aware of the existence of such behavioural tendencies, as individuals with low procrastination tendencies typically perform better than those with high procrastination tendencies. Hooshyar *et al.* (2019) proposes a new algorithm that uses adult submission behaviour to predict the performance of individuals with learning disabilities due to procrastination (referred to as PPP).

2.8 Summary of Review

Use of smartphone applications, virtual reality solutions, and digital devices can reduce stress and achieve a digital detox with these practices. Social well-being, which includes mental well-being, is a valuable social goal in and of itself (Charith, 2021). Social and emotional well-being are essential to overall health and well-being. Being socially and emotionally strong means being able to handle life's normal pressures, work productively, and demonstrate an ability to contribute to community. Determinants of user mental health are diverse, interconnected, and frequently multifaceted. Although internet use has been linked to poor mental health. Determinants of mental health include factors such as socio-economic status, education, physical environment, employment and social support networks, and access to health care (Choudhury *et al.,* 2021). Early detection, accurate diagnosis, and effective mental health treatment can alleviate the suffering of individuals suffering from behavioural health problems and their families. The research of Figueroa and Aguilera (2020) discovered that the proper infrastructure is required to provide digital interventions securely, without violating personal privacy, and with the risk of data breaches minimized. Mental illnesses are vast and more are being discovered. Mental health is an example of a critical health area in desperate need of technology solutions that will allow for timely, effective, and scalable interventions (Mahsa, 2020).

In the literature evaluated, the stress prediction models developed were for people in the professions that require a lot of physical exertion. This research work focuses on data collected while carrying out mundane tasks like everyday office work and interruptions. The benchmark literature for this work is titled Internet of Things (IoT) based psychological and physical stress evaluation in sportsmen using heart rate variability.

The methodology the research used is a combination of Deep convolutional neural network (DCNN) and Multi-output regression with an accuracy of 98.2%. this work is using Support Vector Machine with Grid Search CV. This study makes use of Support

Vector Machine and grid search-cv to achieve the highest accuracy possible.

CHAPTER THREE

3.0 RESEARCH METHODOLOGY

3.1 MATERIALS AND METHODS FOR THE RESEARCH

This work used Jupyter lab which is an integrated development environment tool for data science. It provides the convenience and easy installation of necessary tools or dependency to perform operations such as; data importation, preprocessing, training and development of machine learning models. Jupyter lab version 3.4.3 and python version

3.9.0 was selected as the choice of programming language.

The Heart Rate Variability dataset downloaded contains 41033 data sample, 35 independent variables with 1 dependent variable.The dataset was downloaded from kaggle repository. It is a result of experiments conducted on 25 subjects doing typical office work (for example writing reports, making presentations, reading e-mail and searching for information). The subjects went through typical working stressors such as receiving unexpected emails interruptions and pressure to complete their work on time.

The proposed approach is illustrated in a simple step below as follows;

1. Importing of HRV dataset.
2. Data preprocessing and cleaning.
3. Dataset is transformed and standardized for efficient prediction.

4. Splitting of dataset into training and testing set.
5. Training of model using Support Vector Machine with varieties of SVM parameter using Grid-Search-CV.
6. Train the model with the best parameter turning.
7. Testing of dataset using standard evaluation metrics (Accuracy, recall, precision and F1-score).

3.2 Mental Health Architecture

The framework designed for this research which is depicted in Figure 3.1, shows the architectural design for this research. It illustrates the conceptual design process of the proposed stress prediction model using Grid-Search-CV for parameter tuning and support vector machine (SVM) learning algorithm (this identifies the best classification accuracy using cross fold validation and reveal the best SVM parameter combination for training). Heart Rate Variability (HRV) data sample is firstly preprocessed and transformed into a suitable format for machine learning prediction. The pre-processed data sample are fed into the Grid-Search-CV to explore the dataset further. This is done by performing cross-fold validation and parameter tuning to identify the best parameter combination that yields the best accuracy. Then the model is trained using SVM classifier with the best parameter identified by Grid-Search-CV. Finally, the classification model is made available for update and prediction of adult stress level.

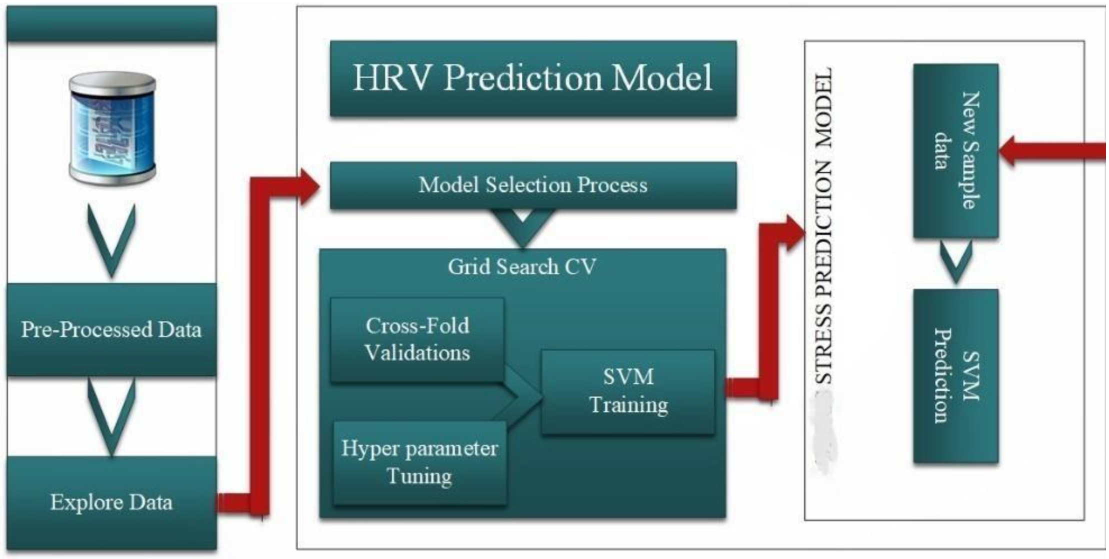


Figure 3.1: Prediction Model Process.

3.2.1 Data collection

The dataset available for developing the proposed model is downloaded from Kaggle repository. Kaggle is data repository which offers a wealthy amount of dataset. Kaggle is an on line community platform for data scientists and machine learning enthusiasts. Kaggle allows users to collaborate with other users, find and publish datasets, use GPU integrated notebooks, and compete with other data scientists to solve data science challenges. The HRV dataset downloaded contains 41033 data sample, 35 independent variables with 1 dependent variable.

3.2.2 Preprocessing

Immediately after data importation into the jupyter lab environment, considering this study, the data preprocessing stage will include encoding the independent variable into numerical forms and standardization of data value using min max scaler techniques, this ensures that the data point are uniformly distributed. Finally, the dataset will be split into training (model development) and testing (model evaluation) set.

3.2.3 Data exploration

Data exploration is the initial step in data analysis, where users explore a large data set in an unstructured way to uncover initial patterns, characteristics, and points of interest. Exploratory Data Analysis is first performed, then the data is loaded, basic information about data is learned, duplicate values are searched for, find the unique values in the data, visualize the unique counts, find if there are null values, finally the null values are replaced. It is essential to explore the dataset in other to get meaningful insight about the

HRV data samples.

3.2.4 HRV prediction model

A good heart rate variability (HRV) score is relative for each person. HRV is a highly sensitive metric and responds uniquely for everyone. As a rule of thumb, values below

50 ms are classified as unhealthy, 50–100 ms signal compromised health, and above 100 ms are healthy. This work will use heart rate variability dataset for the prediction model as it is one of the different ways to measure stress. When the nervous system is responding to a stressful situation, the difference between heartbeats tends to be low compared to when the system is in a relaxed mode where the difference between the heartbeats is high. If an individual is suffering from anxiety disorder or any form of stress induced mental illness, they are determined by chronically low heart rate variability (HRV) compared to healthy individuals during resting state conditions or non stressful situations.

3.2.5 Model selection process

This is the process used to select the best model among a set of models to be used to predict stress. Model selection is a process used to compare the relative value of different statistical models and determine which one is the best fit for the observed data.

3.2.5.1 Grid search - cross validation

Grid Search divides the range of parameters to be optimized into a grid and crosses all points to get the optimal parameters. Grid Search optimizes SVM parameter using cross validation technique as a performance metric. Grid search is a hyper parameter tuning method used in finding the optimal parameter values from a given set of parameters in a given set of parameters in a grid. The technique is cross validation.

3.2.5.2 Cross fold validation

Cross-validation is a technique for evaluating Machine Learning models by training several machine learning models on subsets of the available input data and evaluating them on the complementary subset of the data. Cross-validation is used to detect overfitting, which means failing to generalize a pattern. Cross-validation (CV) is also used to select one of the multiple models.

3.2.5.3 Hyperparameter tuning

Hyper parameter tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm (support vector machine). Hyper parameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning.

3.2.5.4 SVM training

Support vector machine also called SVM is a supervised learning model with associated learning algorithms that analyses data for classification and regression analysis. SVM works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. A separator between the categories is found, then the data are transformed in such a way that the separator could be drawn as a hyperplane. The study will be using this approach to fine tune the support vector machine hyper parameters.

Algorithm 3.1: Grid Search for parameter C on SVM

Input: A list of C candidates
Output: Optimal hyperparameter, MaxC
START
Initialize list of C candidates
FOR every c in list of C candidates Train
    SVM with c on TrainingSet
    Evaluate SVM classification on ValidationSet
    IF accuracy > MaxAccuracy **THEN**
        save MaxC = c
    ENDIF ENDFOR
RETURN MaxC END

3.2.6 Stress prediction model

The prediction model will predict a person's stress level determining if the person is stress using the given sample data. Developing a stress prediction model is the process of predicting future events or outcomes by analysing patterns in a given set of input data.

3.2.6.1 New sample data

Input new data to analyse if the Heart Rate Variability levels are high or low. This will determine if the subject is stressed or not.

3.2.6.2 SVM prediction

The support vector machine classifier will accurately perform the prediction to come out with the optimal result. The internal working is based on the principle of structural risk management to identify the appropriate or best hyperplane distinguishing two or more classes in an n-dimensional space. Furthermore, for multiclass problem the SVM can be trained to differentiate a class from the rest of the classes. The support vector machine can fine tune the value of different kernel in other to efficiently distinguish each class.

3.3 Stress Model Formulation

The Grid-Search-CV technique is adopted to fine-tune the SVM classifier with different parametric combination, the best parameter configuration is selected for training the model

Mathematical definition of an SVM multiclass problem with Mc_i classes, with and input vector *x*,

$$\sum_{i=1}^{M} P(c_i|x) = 1 \tag{3.3}$$

The probability of correctly classifying class =

$$P_c = \sum_{i=1}^{M} P(x \in R_i, c_i) = \sum_{i=1}^{M} P(c_i) \int_{R_i} p(x|c_i)dx, \tag{3.4}$$

R*i* represent the region of the *N* feature space, by definition region R is represented as

$$P_c = \sum_{i=1}^{M} \int_{R_i} P(x|c_i)p(x)dx \geq \frac{1}{M}\sum_{i=1}^{M}\int_{R_i} p(x)dx,$$

$$\Rightarrow P_c \geq \frac{1}{M}; \tag{3.5}$$

Finally, the probability of the multiclassification error =

$$P_e = 1 - P_c \leq 1 - \frac{1}{M} = \frac{M-1}{M}. \tag{3.6}$$

3.4 Performance Evaluation Metrics

Metrics are used to monitor and measure the performance of a model during training and testing. Performance evaluation metrics are a key part of every machine learning project. The metrics inform the researchers if there is progress and gives a measurement to it. All machine learning models developed needs a metric to judge the projects performance. Accuracy is the most commonly used evaluation metrics though it only works on a balanced class and not an unbalanced class. To find accuracy first the error rate is calculated using the formula which gives accuracy as a difference of error rate;

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3.7}$$

Where TP: True positives

TN: True Negative

FP: False Positive

FN: False Negative

The metrics that will be used to test and measure the accuracy of the model will be Recall, Precision and F1-Score. Recall provides the functionality of a classification model that identifies all data points within a class of interest. Recall calculates the percentage of the positives that the model predicts well. To get the percentage the number of well predicted positives is divided by the total number of positives.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \qquad (3.8)$$

Where TP: True positives

FN: False Negative

Precision is a metric that quantifies the number of correct positive predictions. Precision therefore computes minority class accuracy. It is calculated as the proportion of correctly predicted positive cases divided by the total number of positive cases predicted.

The number of positive predictions well made is calculated.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \qquad (3.9)$$

Where TP: True positives

FP: False Positive

F1 score is a machine learning metric that measures the accuracy of a model. Precision ad recall cannot completely evaluate the model. So F1-Score combines the scores of the precision and recall of a model. The accuracy metric calculates the number of times the model made correct predictions across the dataset.

$$\text{F1 - Score} = \frac{2 * P * R}{P + R} \times 100\% \qquad (3.10)$$

Where P: Precision

R: Recall

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Data Import Stage

The model development is carried out within the Google Collaboratory environment. Google colab is a powerful platform that is used learn and rapidly develop machine learning models in Python. Data is imported into the jupyter lab environment.

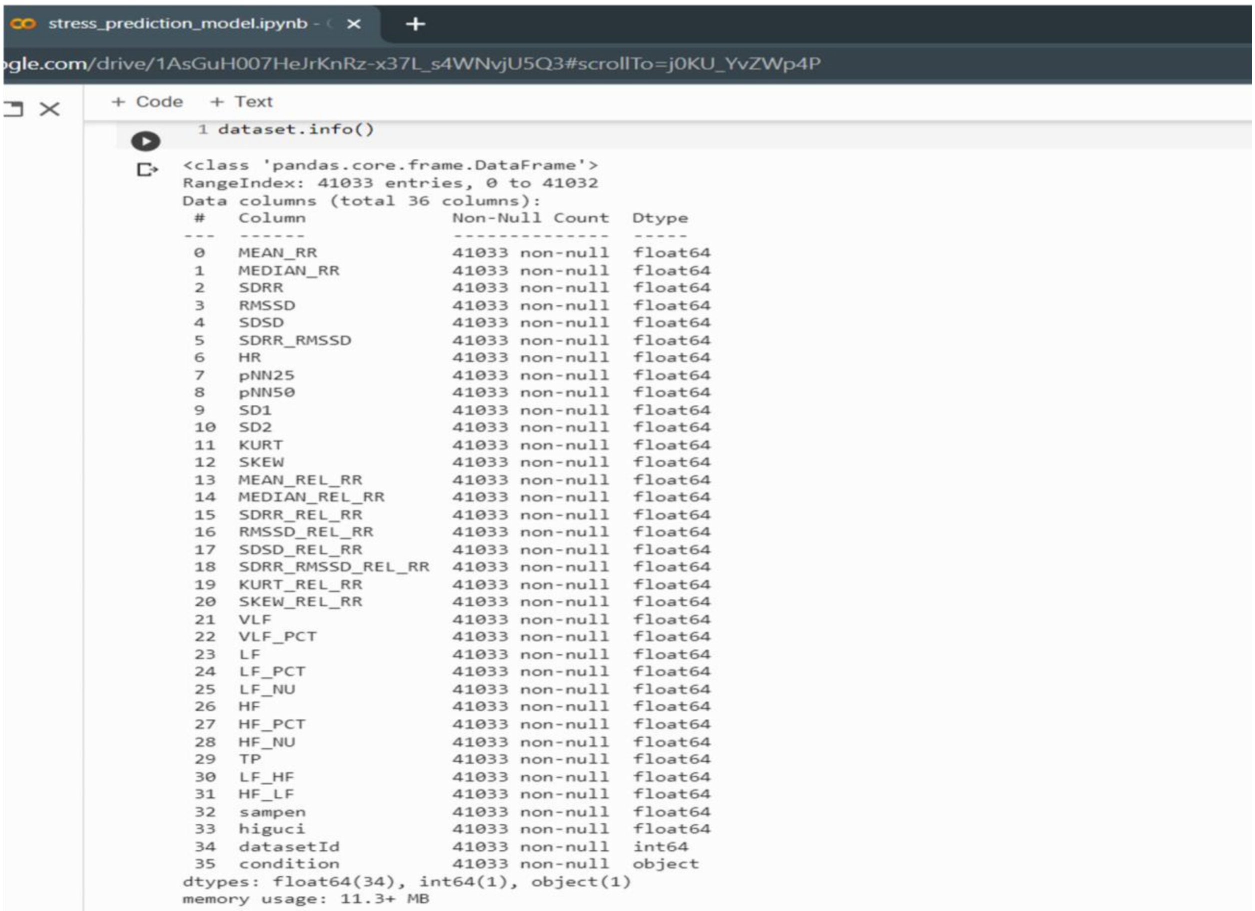
The Heart Rate Variability (HRV) dataset is imported using the panda's python library. The dataset is read into the Google Collaboratory environment with a helper method '*read_csv'*, which is used in importing the csv file from local or remote repository. The csv file fully known as a comma-separated values file which is a plain text file that due to how specific its format is enables data to be saved in a table structured format.

However, five sample datapoint is viewed using the *'head'* helper method.

4.2 Data Exploration Stage

The exploratory data analysis enables better understanding of the dataset, and some essential meta data about the dataset. In respect to Figure 4.2, it can be identified that the dataset consists of 41,033 rows and 36 columns. Additionally, it can be identified that the dataset values are in numeric format.

The downloaded dataset contains 41,033 datapoint (rows) and 36 features (columns). The meta data about each column is essential to be explored and viewed based on Figure 4.1, because it states the numbers of null (empty) and not null (not-empty) value in each column. In addition, it also states the data type of each columns. Generally, the name of each attribute, data type, number of attribute, null, and not-null value are showcased



```
stress_prediction_model.ipynb
gle.com/drive/1AsGuH007HeJrKnRz-x37L_s4WNvjU5Q3#scrollTo=j0KU_YvZWp4P
+ Code  + Text
1 dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41033 entries, 0 to 41032
Data columns (total 36 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   MEAN_RR            41033 non-null  float64
 1   MEDIAN_RR          41033 non-null  float64
 2   SDRR               41033 non-null  float64
 3   RMSSD              41033 non-null  float64
 4   SDSD               41033 non-null  float64
 5   SDRR_RMSSD         41033 non-null  float64
 6   HR                 41033 non-null  float64
 7   pNN25              41033 non-null  float64
 8   pNN50              41033 non-null  float64
 9   SD1                41033 non-null  float64
 10  SD2                41033 non-null  float64
 11  KURT               41033 non-null  float64
 12  SKEW               41033 non-null  float64
 13  MEAN_REL_RR        41033 non-null  float64
 14  MEDIAN_REL_RR      41033 non-null  float64
 15  SDRR_REL_RR        41033 non-null  float64
 16  RMSSD_REL_RR       41033 non-null  float64
 17  SDSD_REL_RR        41033 non-null  float64
 18  SDRR_RMSSD_REL_RR  41033 non-null  float64
 19  KURT_REL_RR        41033 non-null  float64
 20  SKEW_REL_RR        41033 non-null  float64
 21  VLF                41033 non-null  float64
 22  VLF_PCT            41033 non-null  float64
 23  LF                 41033 non-null  float64
 24  LF_PCT             41033 non-null  float64
 25  LF_NU              41033 non-null  float64
 26  HF                 41033 non-null  float64
 27  HF_PCT             41033 non-null  float64
 28  HF_NU              41033 non-null  float64
 29  TP                 41033 non-null  float64
 30  LF_HF              41033 non-null  float64
 31  HF_LF              41033 non-null  float64
 32  sampen             41033 non-null  float64
 33  higuci             41033 non-null  float64
 34  datasetId          41033 non-null  int64
 35  condition          41033 non-null  object
dtypes: float64(34), int64(1), object(1)
memory usage: 11.3+ MB
```

Figure 4.1: The HRV Dataset meta data information

The no stress class consist of 22,158 datapoint, 7093 datapoint belongs to stress or time pressure class, and 11782 datapoint for the interruption class.

Figure 4.2, visually indicates the three main classes present in the HRV dataset using Bar Chart, this includes; no stress, time pressure or stress, and interruption. Based on the Figure 4.2, the no stress class contain the highest datapoint with about 20,000 data sample, while the time pressure or stress class have the minimum datapoint with about 5,000 data sample and finally interruption class consist of 10,000 data point.

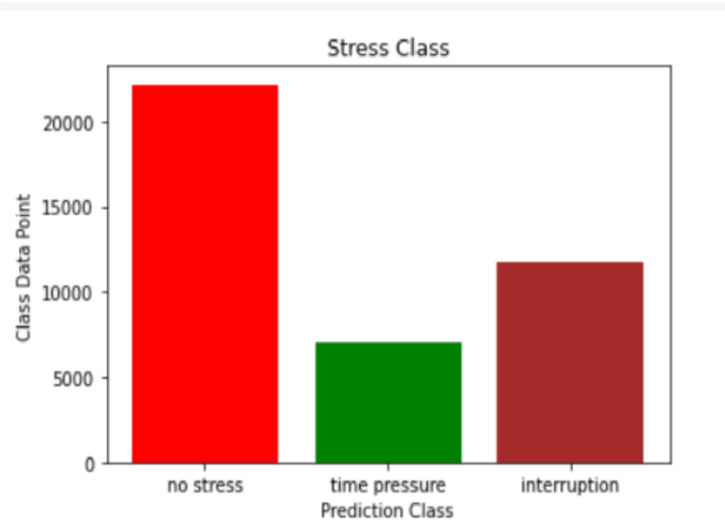


Figure 4.2. The HRV Dataset (Classification Class) Bar chart

4.3 Data Preprocessing Stage

The condition (independent feature) columns consist of “no stress, time pressure, and interruption” unique values. It’s essential to convert or encode them into numeric representation for machine learning algorithm’s understanding. Hence, no stress class is represented with ‘0’, time pressure or stress class with ‘1’, and finally ‘2’ represent the interruption class.

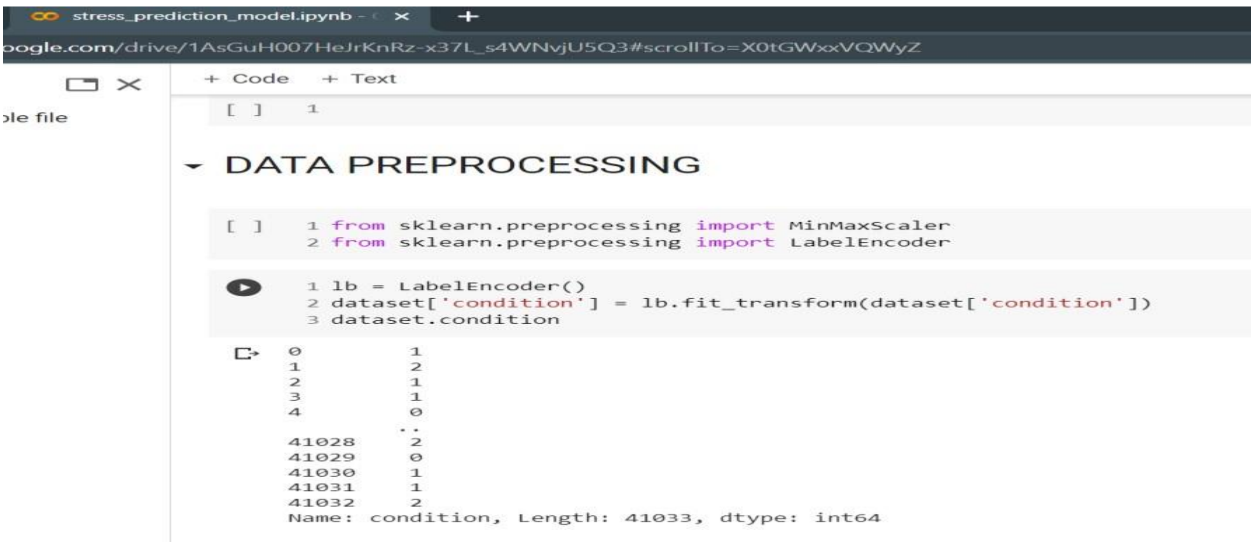


Figure 4.3: Encoding Textual Condition Attribute

4.4 Data Splitting (Training and Testing Set)

DATA SPLITTING (TESTING AND TRAINING SET)

```
[ ]  from sklearn import model_selection
     X_train, X_test, y_train, y_test = model_selection.train_test_split(dataset,y, test_s
```

```
[ ]  print("Train x : ", X_train.shape)
     print("Train y : " , y_train.shape)
     print("test x : ", X_test.shape)
     print("test y : " , y_test.shape)

Train x :  (32826, 34)
Train y :  (32826,)
test x :  (8207, 34)
test y :  (8207,)
```

Figure 4.4: HRV Dataset Splitting (Training and Testing)

This preprocessing stage is essential to improve the machine learning classification accuracy. The dataset scaling standardizes the numeric value between the range of '0' and '1'. This ensure dataset not been loosely distributed.

After dataset preprocessing and dataset is split into 80 percent training data sample and 20 percent testing data sample. Based on Figure 4.4, 32,826 is used for training the proposed model and 8,207 datapoint for testing the trained model.

4.5 HRV Model Training and Testing

The grid search cv is used in training the model with different parameter tuning or configuration, based on the Figure 4.5, the support vector machine is trained with parameter configuration of 'gamma = 10', 'C = 1, 10 and 20 as value' and 'kernel = rbf or linear'. However, the resulted cross fold validation scores and mean test scores are visualized in a tabular form, the essential columns and best parameter configuration is specified in Figure 4.6.

```
[ ]
svm_gridcv = GridSearchCV(SVC(gamma=10), {
    "C":[1, 10, 20],
    'kernel' : ['rbf', 'linear']
}, cv=3, return_train_score=False)

svm_gridcv.fit(X_test, y_test)
```

```
GridSearchCV(cv=3, estimator=SVC(gamma=10),
             param_grid={'C': [1, 10, 20], 'kernel': ['rbf', 'linear']})
```

```
[ ]
import pandas as pd
pd.DataFrame
comparison_result = pd.DataFrame(svm_gridcv.cv_results_)
comparison_result
```

std_fit_time	mean_score_time	std_score_time	param_C	param_kernel	params	split0_test_score
0.526324	0.566521	0.070174	1	rbf	{'C': 1, 'kernel': 'rbf'}	0.987208
0.007305	0.334377	0.009538	1	linear	{'C': 1, 'kernel': 'linear'}	0.635599
0.003567	0.363817	0.003538	10	rbf	{'C': 10, 'kernel': 'rbf'}	0.997076

notely or in another tab. Show diff

✓ 2s completed at 10:11 PM

Figure 4.5: HRV Model Training And Testing

Based on the result analysis of Figure 4.6, the Grid-Search-CV reveals that the row 4 combination of parameter yield the best result, by using the 20 for param_c, and 'rbf' for param_kernel a mean test score of 0.9957 % accuracy is gotten. And this will be It's used in training and evaluating the final model.

+ Code + Text

	param_C	param_kernel	mean_test_score
0	1	rbf	0.986109
1	1	linear	0.635677
2	10	rbf	0.995613
3	10	linear	0.640307
4	20	rbf	0.995735
5	20	linear	0.639576

```
comparison_result.iloc[4]
```

```
mean_fit_time                           0.8077
std_fit_time                          0.008853
mean_score_time                       0.356274
std_score_time                        0.003528
param_C                                     20
param_kernel                               rbf
params               {'C': 20, 'kernel': 'rbf'}
split0_test_score                     0.997076
split1_test_score                     0.995249
split2_test_score                     0.994881
mean_test_score                       0.995735
std_test_score                         0.00096
rank_test_score                              1
Name: 4, dtype: object
```

Figure 4.6: Evaluating the SVM Grid-Search-CV Results

illustrated in fig 4.7, how the training dataset is trained using Support Vector Machine

```
[ ]  svc = SVC(gamma=10, C=20, kernel='rbf')
     svc.fit(X_train, y_train)
     print('training complete')

     training complete

[ ]  svc.score(X_test, y_test)

     0.995857195077373
```

Figure 4.7: HRV Model Training

(SVM). However, the SVM parameter is tuned to gamma=10, c=20, and kernel ='rbf'.

Finally, an accuracy level of 99% is achieved. Table 4.1 shows the

corresponding accuracy, precision and recall gotten for each class.

Table 4.1 HRV Model Classification Report.

	PRECISION	**RECALL**	**F1-SCORE**	**SUPPORT**
0	98.50	99.00	1.00	2368
1	99.50	98.00	1.00	4489
2	99.51	98.00	1.00	1350
Accuracy			99.99	8207
Macro Avg	1.00	1.00	1.00	8207
Weighted Avg	1.00	1.00	1.00	8207

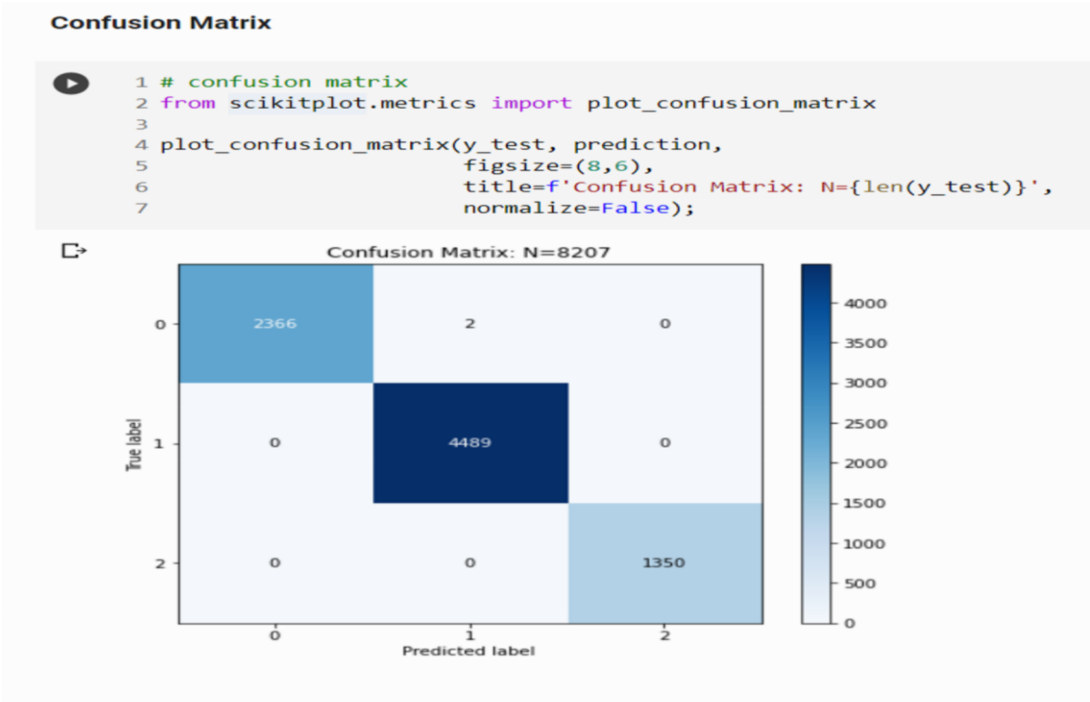


Figure 4.8: Confusion Matrix

The Figure 4.8 shows the correctly predicted or classified datapoint against the datapoint that are misclassified. The diagonal numbers indicate the numbers of correctly classified datapoint, while the other data in the confusion matrix denote the misclassified datapoints.

The accuracy presented by the SVM classifier with grid search-cv performs better than the Deep Convolutional Neural Network (DCNN) and Multi-output regression (MOR) used by the research of Jin *et al* (2021). In this study the existing models, (Jin et al., 2021) model and the proposed model is compared using Table 4.2;

Table 4.2: Result Comparison Analysis

S/N	Approach	Accuracy (%)
1	Random Forest	93%
2	SVM	94 %
3	MO-SVR	95.8%
4	DCNN	95.2%
5	MOR-DCNN (Jin et al., 2021)	98.2%
6	Proposed Model	99.9 %

Based on the accuracy comparisons of the existing models and the proposed model, its proof that the proposed model with (99.9 %) accuracy perform better than the existing approach.

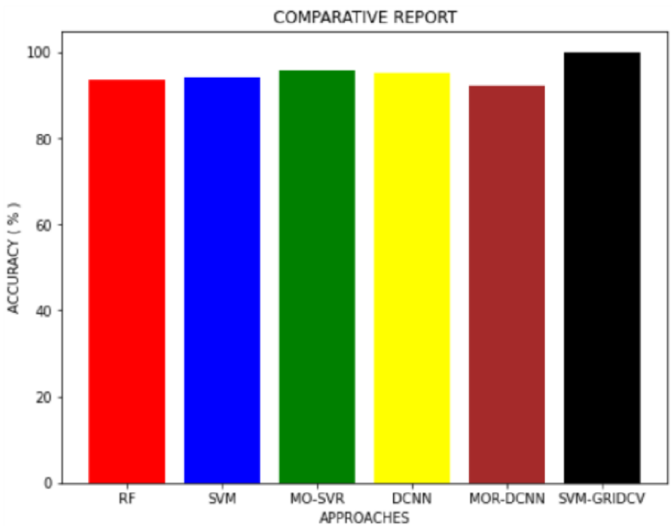


Figure 4.9: Comparison of existing model and proposed model

Figure 4.9 shows the models' accuracy rate in a bar chart to compare which is the highest. The proposed model represented by 'SVM-GRIDCV' with the black color is the closest to 100%.

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The data collected was standardized after importation which is essential in improving the machine learning classification accuracy. The dataset scaling standardizes the numeric value between the range of ‘0’ and ‘1’ to ensure that the dataset was not loosely distributed which is shown in Chapter Four (4.3).

The dataset was split into 80% for training set and 20% for testing set (4.4), it was then finetuned with SVM classifier using grid search cv to obtain the optimal parametric combination.

The developed model is trained using the best parametric combination achieved in Figure 4.10 and evaluated with the standard performance metrics using the testing dataset to achieve a 99% accuracy as shown in Figure 4.13.

5.2 Recommendations

This study provides an efficient and more accurate prediction model for detecting stress level using Heart Variability Rate. The proposed approach has proven better in comparison to the existing study that is evaluated. A different stress measurement could be adopted in future stress prediction model development. Also, the developed model can be adopted by sectors such as the financial, agricultural and medical sectors to manage their workers’ stress levels.

5.3 Contributions to Knowledge

The study came up with the combination of SVM and Grid Search-CV to develop the model proved to be better than multi-output regression and deep convolutional nueral network (MOR-DCNN) developed in the existing research. An improved model for stress detection was developed with an accuracy of 99% which is higher than the compared research (98.2%).

REFERENCES

Aristizabal, S., Byun, K., Wood, N., Mullan, A.F., Porter, P.M., Campanella, C., Jamrozik, A., Nenadic, I.z., & Bauer, B.A. (2021). The Feasibility of Wearable and Self-Report Stress Detection Measures in a Semi-Controlled Lab Environment. *IEEE Access.* 9. DOI: 10.1109/ACCESS.2021.3097038

Bardram, J.E., & Matic, A. (2019). A Decade of Ubiquitous Computing Research in Mental Health. *IEEE Pervasive Computing,* 1536-1268. DOI 10.1109/MPRV.2019.2925338

Bryant, N., Alice, R., Faisal, A., & Hui, Y. (2020), Sensor-Based Virtual Reality for Clinical Decision Support in the Assessment of Mental Disorders. *IEEE Xplore.* pp. 666-669, DOI: 10.1109/CoG47356.2020.9231896

Bu, N., Fukami, M., & Fukuda, O. (2021). Pattern recognition of mental stress levels from differential RRI time series using LSTM networks. *IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech).* 1(21). DOI:10.1109/LIFETECH52111.2021.9391853

Can, Y.S., Chalabianloo, N., Ekiz, D., Fernandez-Alvarez, J., Repetto, C., Riva, G., Iles-Smith, H., & Ersoy, C. (2020) Real-Life Stress Level Monitoring using Smart Bands in the Light of Contextual Information. *IEEE Sensors Journal.* DOI 10.1109/JSEN.2020.2984644

Charith S., Mahsa S., & Mo S. (2019), Predictive Modelling in Mental Health: A Data Science Approach. *IEEE Xplore, 2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET).* pp. 6-11, DOI: 10.1109/CSUDET47057.2019.9214626

Choudhury A., Asan O., & Alelyani, T. (2021). Exploring the Role of the Internet, Care Quality and Communication in Shaping Mental Health: Analysis of the Health Information National Trends Survey. *IEEE Journal of Biomedical and Health Informatics,* 2168-2194. DOI 10.1109/JBHI.2021.3087083

Clark, J., Nath, K.R., & Thapliya, H. (2021). Machine Learning Based Prediction of Future Stress Events in a Driving Scenario. *2021 IEEE 7th World Forum on Internet of Things (WF-IoT).* pp. 455-458, DOI: 10.1109/WF-IoT51360.2021.9595098

Figueroa, C.A., & Aguilera, A.(2020). The need for a mental health technology revolution in the COVID-19 pandemic. *Front. Psychiatry* 11(523). doi:10.3389/fpsyt.2020.00523

Gao, J., & Shi, G. (2021). Mental Health Evaluation of College individuals Based on Similar Trajectory Clustering Algorithm. *2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA).* 6(21). DOI 10.1109/ICMTMA52658.2021.00189

Gupta, D., Bhatia, M.P.S., & Kumar A. (2021). *Resolving Data Overload and Latency Issues in Multivariate Time-Series IoMT Data for Mental Health Monitoring, IEEE Sensors Journal.* 1530(437). DOI 10.1109/JSEN.2021.3095853

Haowei, G., & Ting, W. (2020). Analysis of Community Mental Health Services in the Context of Big Data. *2020 International Conference on Big Data and Social Sciences (ICBDSS).* 7(20). DOI: 10.1109/ICBDSS51270.2020.00016

Hasanbasic, A., Spahic, M., Bosnjic, D., Hadzic, H., Mesic, V., & Jahic, O. (2019). Recognition of stress levels among individuals with wearable sensors. *18th International Symposium INFOTEH-JAHORINA,* 20-22. DOI: 10.1109/INFOTEH.2019.8717754

Hooshyar, D., Pedaste, M., & Yang, Y. (2019). Mining Educational Data to Predict individuals' Performance through Procrastination Behavior. *Entropy* 22(12). doi:10.3390/e22010012

Jin, N., Zhang, X., Hou, Z., Sanz-Prieto, I., & Mohammed, B.S. (2021). IoT based psychological and physical stress evaluation in sportsmen using heart rate variability *Aggression and Violent Behavior. Elsevier,* https://doi.org/10.1016/j.avb.2021.101587

Jing, L. (2021) Optimal Modeling of College individuals' Mental Health Based on Brain-Computer Interface and Imaging Sensing. *Proceedings of the Fifth International Conference on Intelligent Computing and Control Systems (ICICCS 2021,)* IEEE Xplore. pp. 772-775, doi: 10.1109/ICICCS51141.2021.9432220

Kaga, S. & Kato, S. (2019). Extraction of useful features for stress detection using various biosignals doing mental arithmetic. *2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech 2019)* pp. 153-154, DOI: 10.1109/LifeTech.2019.8883967

Karaman, M.A., Lerma, E., Vela, C.J., & Watson, J.C. (2019). Predictors of Academic Stress Among College individuals. *Journal of College Counseling* 22. DOI: 10.1002/jocc.12113

Kavitha, K.R., Rajan, A.K.V., & Pillai, A. (2019). An Improved Feature Selection and Classification of Gene Expression Profile using SVM. *2019 2nd International Conference on Intelligent Computing, Instrumetation and Control Technologies (ICICICT)* DOI:10.1109/ICICICT46008.2019.8993358

Kim, T., Seo, Y., Lee, J., Chae, S., & An, J. (2021). Brain to music: Musical Representation from stress-induced EEG. *2021 9th international Winter Conference on Brain-Computer Interface (BCI).* pp. 1-6. DOI: 10.1109/BCI51272.2021.9385354

Klyueva, I. (2019). Improving Quality of the Multiclass SVM Classification Based on the Feature Engineering. *2019 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA), Lipetsk, Russia* 2019, pp. 491-494. DOI: 10.1109/SUMMA48161.2019.8947599

Kolenik, T., & Gams, M. (2021), Persuasive Technology for Mental Health: One Step Closer to (Mental Health Care) Equality? *IEEE Technology and Society Magazine.* 0278(0097). DOI 10.1109/MTS.2021.3056288

Kong, S. T., (2019), A qualitative analysis of user reviews on mental health apps: who used it? for what? and why?*2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), Danang, Vietnam.* pp. 1-4. DOI: 10.1109/RIVF.2019.8713726

Laijawala, V., Aachaliya, A., Jatta, H., & Pinjarkar, V. (2020). Classification Algorithms based Mental Health Prediction using Data Mining. *Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020) IEEE Conference Record # 48766;* pp. 1174-1178, DOI: 10.1109/ICCES48766.2020.9137856

Liang, Z., Tatha, O., & Andersen, L.E. (2020). Developing mHealth App for Tracking Academic Stress and Physiological Reactions to Stress. *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech),* 3(20). DOI:10.1109/LifeTech48969.2020.1570618580

Lin, B., Wang, L., Ding, H., & Feng, X. (2021). Improving L2 English Rhythm Evaluation With Automatic Sentence Stress Detection. *2021 IEEE Spoken Language Technology Workshop (SLT).* 4(20). DOI: 10.1109/SLT48900.2021.9383455

Lin, C., & Yang, H. (2020). Research on the Value of Physical Education in to the Physical & Mental Health Development of Teenagers. *2020 International Conference on Modern Education and Information Management (ICMEIM).* 7(20). DOI: 10.1109/ICMEIM51375.2020.00161

Liu, H., Xiao, X., Li, Y., Mi., & Yang, Z. (2019). Effective Data Classification via Combining Neural Networks and SVM. *2019 Chinese Control And Decision Conference (CCDC).* 4(19). pp. 4006-4009. DOI: 10.1109/CCDC.2019.8832442

Mahsa, H., Jaejoon, L., Christopher, B., Jiangtao, W., and Sumi, H. (2020). What Happens in Peer-Support, Stays in Peer-Support: Software Architecture for Peer-Sourcing in Mental Health. *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)* 0(20). DOI 10.1109/COMPSAC48688.2020.0-184

Manav, P.J., Soumyaprakash, S. D., & Stevina C., (2020), Mental Health State Detection Using Open CV and Sentimental Analysis, *Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020] IEEE Xplore* 3, pp. 465-470, DOI: 10.1109/ICISS49785.2020.9315984

Nazari, N., & Far, D.M. (2019). The Relationship between Teaching Skills, Academic Emotion, Academic Stress and Mindset in adult Academic Achievement Prediction: A PLS-SEM Approach. *Journal of Intellectual Disability - Diagnosis and Treatment.* 7(3). 119-133. DOI:10.6000/2292-2598.2019.07.03.9

Nilufar B., Sylvia H., Imran K., Lehan S., Jamuna K., John N., Hai-Ning L., Hamid S. Glass, J., & McGregor, C. (2020). Towards Player Health Analytics in Overwatch. *IEEE Xplore.* 6(20). pp. 1-5, DOI: 10.1109/SeGAH49190.2020.9201733

Pankajavalli, P.B., Karthick, G.S., & Sakthivel, R. (2021). An Efficient Machine Learning Framework for Stress Prediction via Sensor Integrated Keyboard Data. *IEEE Access.* DOI: 10.1109/ACCESS.2021.3094334

Park, K., Kim, M. J., Kim,J., Kwon, O. C., Yoon, D., & Kim, H. (2021). Requirements and Design of Mental Health System for Stress Management of Knowledge Workers. 9(20). pp. 1829-1832, DOI: 10.1109/ICTC49870.2020.9289464.

Priya, A., Garg S., & Tigga, N.P. (2019). Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science* 167. 1258–1267, 1877(0509). DOI: 10.1016/j.procs.2020.03.442

Shen, Y. (2021). Analysis on Mental Health Education Strategy of Higher Vocational College individuals Based on Information Technology. *2021 2nd International Conference on Education, Knowledge and Information Management (ICEKIM).* 0(21). DOI: 10.1109/ICEKIM52309.2021.00151

Sulistiana, & Muslim, M.A. (2020). Support Vector Machine (SVM) Optimization Using Grid Search and Unigram to Improve E-Commerce Review Accuracy. *Journal of Soft Computing Exploration,* *https://shmpublisher.com/index.php/joscex*. 1(1). DOI: https://doi.org/10.52465/joscex.v1i1.3

Thomas, N., Perumalla, A., Rao, S., Thangaraj, V., Ravi, S.K., Geethanath, S., Kim, H., & Srinivasan, G. (2020). Fully Automated End-to-End Neuroimaging Workflow for Mental Health Screening. *2020 IEEE 20th International Conference on BioInformatics and BioEngineering (BIBE).* 9(20). DOI:10.1109/BIBE50027.2020.00109

Trigueros, R., Padilla, A.M., Aguilar-Parra, J.M., Rocamora, P, Morales-Gazquez, M.J., & Lopez-Liria, R. (2020). The Influence of Emotional Intelligence on Resilience, Test Anxiety, Academic Stress and the Mediterranean Diet. A Study with individuals. *International Journal of Environmental Research and Public Health.* 17(6). DOI:10.3390/ijerph17062071

Tumpa, Z. N., Saifuzzaman, M., Rabby, S.K. F., Crearie, L., & Stansfield, M. (2021). Understanding Business Intelligence in The Context of Mental Healthcare Sector of

Bangladesh for Improving Health Services. *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC).* DOI:10.1109/R10-HTC49770.2020.9357023

Wang, H., Cheng, L., Wang, L., & Sang, Y. (2020), The visualization analysis of network mental health education literature in china based on citespace. *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT).* 0(20). DOI: 10.1109/ISCTT51595.2020.0030

Wang, R., Wang, J., Liao, Y., & Wang, J. (2020). Supervised Machine Learning Chatbots for Perinatal Mental Healthcare. *2020 International Conference on Intelligent Com puting and H um an-Com puter Interaction (ICHCI). 8(20).* DOI 10.1109/ICHCI51889.2020.00086

Wang, Z., Jiang, Y., & Hu, X. (2020). A Leaf Type Recognition Algorithm Based on SVM Optimized by Improved Grid Search Method. *2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*. 3(20). DOI 10.1109/ICECTT50890.2020.00076

Yan, H. (2020) Improve the mental health of college individuals by means of computer network technology. *2020 International Conference on Computer Engineering and Intelligent Control (ICCEIC).* 6(20). DOI 10.1109/ICCEIC51584.2020.00008

Yijun, Z. (2020). Research on innovative ways of mental health education for vocational individuals under the internet background. *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)* . 4(20). DOI: 10.1109/ICMCCE51767.2020.00372

Yueheng, W., Kun, Q., Jacob, N., Hiromichi, Y., Akifumi, K., Kenji, M., & Yoshiharu, Y., (2020), Can Affective Computing Better the Mental Status of the Electronic Games Player? A Perspective. *2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech 2020).* 3(20). DOI: 10.1109/LifeTech48969.2020.1570620404

Zamkah, A., Hui, T., Andrews, S., Dey, N., Shi, F., & Sherratt, R.S. (2020). Identification of Suitable Biomarkers for Stress and Emotion Detection for Future Personal Affective Wearable Sensors. *Biosensors.* 10(40). DOI:10.3390/bios10040040

Zhaorong, Y., & Yiwen, C. (2019) Design and Application of Adolescent Mental Health Data Mining and Management System. *2019 International Conference on Smart Grid and Electrical Automation (ICSGEA).* 4(19). DOI: 10.1109/ICSGEA.2019.00117

APPENDICES

EXPLORATORY DATA ANALYSIS (EDA)

```
1 dataset
```

	MEAN_RR	MEDIAN_RR	SDRR	RMSSD	SDSD	SDRR_RMSSD	HR	pNN25	pNN50	SD1	...	HF	HF_PC
0	721.901897	727.267280	74.722315	12.361264	12.361069	6.044877	84.121868	4.933333	0.000000	8.743513	...	66.617057	3.92186
1	843.538633	844.407930	58.499429	19.298880	19.298795	3.031234	71.478642	21.000000	0.200000	13.650863	...	26.500086	1.12341
2	958.523868	966.671125	132.849110	21.342715	21.342653	6.224565	63.874293	24.133333	1.800000	15.096571	...	16.024935	0.37020
3	824.838669	842.485905	117.822094	11.771814	11.771248	10.008830	74.330531	4.733333	0.533333	8.326307	...	17.581470	0.61593
4	756.707933	747.941620	143.968457	13.357748	13.356388	10.777899	82.092049	5.933333	0.666667	9.447545	...	35.199054	0.66287
...	...	...	...	...	...	...	...	...	...	...	...	...	..
41028	1118.406543	1117.857050	113.955632	18.592177	18.592071	6.129225	54.234182	18.800000	0.266667	13.150967	...	0.347514	0.00681

```
[ ]  1 row , column = dataset.shape
     2 print('Number of Datapoints : ', row)
     3 print('Number of Features : ', column)
```

```
Number of Datapoints :  41033
Number of Features :  36
```

```
1 dataset.columns
```

```
Index(['MEAN_RR', 'MEDIAN_RR', 'SDRR', 'RMSSD', 'SDSD', 'SDRR_RMSSD', 'HR',
       'pNN25', 'pNN50', 'SD1', 'SD2', 'KURT', 'SKEW', 'MEAN_REL_RR',
       'MEDIAN_REL_RR', 'SDRR_REL_RR', 'RMSSD_REL_RR', 'SDSD_REL_RR',
       'SDRR_RMSSD_REL_RR', 'KURT_REL_RR', 'SKEW_REL_RR', 'VLF', 'VLF_PCT',
       'LF', 'LF_PCT', 'LF_NU', 'HF', 'HF_PCT', 'HF_NU', 'TP', 'LF_HF',
       'HF_LF', 'sampen', 'higuci', 'datasetId', 'condition'],
      dtype='object')
```

IMPORTING HRV DATASET

```
import pandas as pd
dataset = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/DATASET/hvr_test.csv')
dataset.head()
```

	MEAN_RR	MEDIAN_RR	SDRR	RMSSD	SDSD	SDRR_RMSSD	HR	pNN25	pNN50	SD1	...
0	721.901897	727.267280	74.722315	12.361264	12.361069	6.044877	84.121868	4.933333	0.000000	8.743513	...
1	843.538633	844.407930	58.499429	19.298880	19.298795	3.031234	71.478642	21.000000	0.200000	13.650863	...
2	958.523868	966.671125	132.849110	21.342715	21.342653	6.224565	63.874293	24.133333	1.800000	15.096571	...
3	824.838669	842.485905	117.822094	11.771814	11.771248	10.008830	74.330531	4.733333	0.533333	8.326307	...
4	756.707933	747.941620	143.968457	13.357748	13.356388	10.777899	82.092049	5.933333	0.666667	9.447545	...

5 rows × 36 columns

```
no_stress = len(dataset[dataset['condition'] =='no stress'])
time_pressure = len(dataset[dataset['condition'] =='time pressure'])
```

stress_prediction_model.ipynb

com/drive/1AsGuH007HeJrKnRz-x37L_s4WNvjU5Q3#scrollTo=j0KU_YvZWp4P

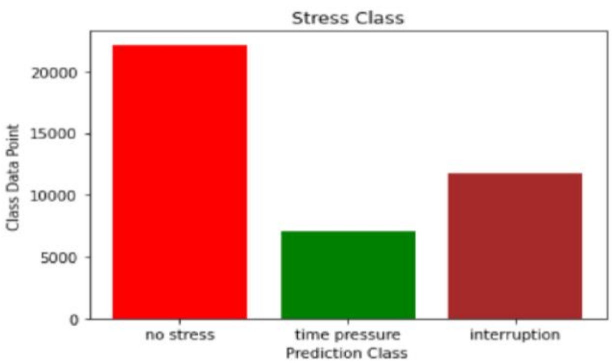
+ Code + Text

```
interruption = len(dataset[dataset['condition'] == 'interruption'])
y_bar_data = [no_stress, time_pressure, interruption]
```

Data Visualization

```
import matplotlib.pyplot as plt

plt.xlabel('Prediction Class')
plt.ylabel('Class Data Point')
plt.title('Stress Class')
plt.bar(x_bar_data, y_bar_data, color=['red', 'green', 'brown'])
plt.show()
```



+ Code + Text

Data Standadization

```python
min_max = MinMaxScaler()
# train_x = min_max.fit_transform(train_x)
# test_x = min_max.fit_transform(test_x)

for c in dataset.columns:
   dataset[c] = min_max.fit_transform(dataset[[c]])

dataset.head()
```

	MEAN_RR	MEDIAN_RR	SDRR	RMSSD	SDSD	SDRR_RMSSD	HR	pNN25	pNN50	SD1	...	LF_PCT
0	0.225313	0.184864	0.088443	0.324583	0.324578	0.065286	0.544238	0.125212	0.000000	0.324578	...	0.452452
1	0.382444	0.287996	0.058163	0.654295	0.654295	0.007079	0.349779	0.532995	0.037037	0.654295	...	0.852726
2	0.530982	0.395639	0.196938	0.751429	0.751430	0.068757	0.232820	0.612521	0.333333	0.751430	...	0.607359
3	0.358287	0.286304	0.168890	0.296569	0.296546	0.141849	0.393642	0.120135	0.098765	0.296546	...	0.206464
4	0.270276	0.203066	0.217693	0.371941	0.371880	0.156703	0.513019	0.150592	0.123457	0.371880	...	0.102285

5 rows × 34 columns

	precision	recall	f1-score	support
0	98.50	99.00	1.00	2368
1	99.50	98.00	1.00	4489
2	99.51	98.00	1.00	1350
accuracy			99.99	8207
macro avg	1.00	1.00	1.00	8207
weighted avg	1.00	1.00	1.00	8207