E-JUST University - Computer Science and Information Technology Programs –
Artificial Intelligence and Data Science Program (AID)
Subject: Mathematics for Data Science (AID311)

MATH FOR DATA SCEINCE PROJECT (PHASE I& PHASE II COVER SHEET)

Project Name: ***Stress level prediction.***

Team Information ***(typed not handwritten, except for the attendance signature)*****:**

	ID [Ordered by ID]	Full Name	Attendance Signature [Handwritten]	Final Grade
1	320210037	Mahmoud Ossama Mustafa Mokhiamar		

Items		Actual Grade	Notes
Project Documentation **Please follow the style in the document file attached with this phase and be careful with the information. The documentation will check by similarity and plagiarism checker. Each phase has 2 marks.**	**16**		
Presentation and implementation **Each student has 10 minutes to present the idea, methodology and the interpretation of results.**	**4**		
Preprocessing			
Data Visualization **Missing Values Treatment** **Binning process** (If exist) **Data Analysis** (Min, Max, Mean, Variance, Standard Deviation, Skewness, Kurtosis). **Data Analysis** (Covariance matrix, Correlation, Heat map, Chi-square Test, Z-test or t-test, ANOVA)	**0**		
Feature Reduction **Linear Discriminate Analysis (LDA)** **Principle Component Analysis (PCA) and Kernel PCA (if data non-linear)** **Singular Value Decomposition (SVD)**	**0**		
Model Implementations			
Naive Bayesian	**0**		
Bayesian Belief Network	**0**		
Decision Tree (Entropy, and error estimation)	**0**		
LDA	**0**		
Neural Network	**0**		
K-NN (Different distances)	**0**		
Model evaluations			
Dataset splitting (80% training and 20% testing) and apply all evaluation matrix	**0**		

K-fold cross validation and average accuracy			
Confusion Matrix **Accuracy** **Error rate** **Precision** **Recall** **F-measure** **ROC**			
Interpret results of confusion matrix and show the model overfitten or underfitten	**o**		
Comparisons with other related work on the same domain -Table	**o**		
References (papers used in your domain of work and the studies uses the same data sets)	**o**		

Teaching-Assistant's Signature: _____

_____

20

***Abstract*— Psychological Stress can highly affect a person's health. Extended periods of stress exposure might have negative consequences that may necessitate costly therapies. Acute degree stress can be fatal for those who already have a diagnosis of schizophrenia or borderline personality disorder. A statistical predictive modeling approach was used in this project to accurately measure one's current stress level, which can be a proactive approach to avoid the cost effects of prolonged exposure to high stress. The models proposed are trained in over 2000+ thousand instances, it monitors a person's vital reading, and using these sensor readings the system estimates the level of stress. The vitals considered are body temperature, rate of motion, and sweat, irrelevant attributes to stress level such as heart rate are neglected due to their irrelevance on stress level in accordance with latest studies (In such scenarios, though the increase in heart rate might help in burning more calories, it cannot identify the stress level of the individual) [1]. The proposed models have achieved overall high accuracy (98%). Performance wise, the models proposed are very fast in training and testing, and computationally efficient which can be implemented in an embedded system for real-time stress detection, therefore measuring individual's stress and avoiding high-cost treatment resulting from prolonged exposures to stress.**

I. Introduction

Stress, a persuasive aspect of human experience, manifests in various forms, with classifications including eustress, neustress, and distress. Eustress, characterized as "good" stress, has the potential to inspire heightened performance and motivation [5]. Neustress, identified as neutral stress, lacks adverse effects on well-being and can be safely disregarded. Conversely, distress, a form of stress with negative implications for the human body, demands focused attention. Distress is further categorized based on its temporal characteristics into acute and chronic stress. Acute stress entails short but intense episodes, while chronic stress involves prolonged and intense levels, with the latter having profound consequences on human health [6]. Chronic stress exerts severe impacts on healthy living [7]. It heightens muscle tension, leading to impairment in daily physical activities. Moreover, escalated stress levels can precipitate complex mental illnesses such as borderline personality disorder (BPD). BPD manifests through dangerous mood swings, alterations in behavioral patterns, eating disorders, and may induce stressed individuals to make unhealthy decisions. Recognizing the nuanced distinctions between these types of stress is crucial for developing targeted interventions and strategies to mitigate the adverse effects on both physical and mental well-being.

In recent times, stress detection through machine learning techniques has been a growing area of interest. Various machine learning algorithms as decision trees, random forests and support vector machines SVMs have also been used to analyze data from wearable devices and speech signals in studies. One research employed both wearable devices and machine learning schemes to identify stress in people. 80% accuracy of the system in detecting stress has been found through this study (Molina-Ros et al., 2018). Another research applied machine learning algorithms on speech signals to determine stress [8]. This research revealed that pitch and tone were helpful acoustic features for stress detection. 80% accuracy was also obtained in the classification of speech signals as high or low stress through machine learning algorithms. Poulos et al., 2017 In general, these studies indicate that machine learning techniques can be used to accurately detect stress in people using wearable devices and speech signals [9]. Despite these achievements, additional research is required to overcome the obstacles and limitations associated with creating such reliable stress detection systems.

The main contribution provided in this project is using simple statistical models in evaluation of the stress level. Although most of the models used are simple and computationally efficient, they provided very high results in prediction of the stress level. The advantages of using simple statistical models in stress detection include Computational efficiency: The models used in this project are computationally efficient, making them suitable for real-time stress detection applications. Ease of interpretation: The models used in this project are easy to interpret, making them suitable for applications where interpretability is important. Robustness: The models used in this project are robust, making them suitable for applications where data quality may be compromised. Models used includes Gaussian Naive Bayes, KNN, LDA, and Decision trees to estimate the stress level for a proactive approach to mitigate the possible effects of a prolonged stress exposure. The computation efficiency of such models emphasizes the potential for a real-time prediction with the integration of IOMT.

This document is divided into first Abstract, then introduction which includes problem statement, literature review, and contribution to project, then methodology, proposed model, results, conclusions and future work.

II. Methodology

Predictive modeling was used on the Stress-lysis dataset to predict the stress level, the dataset incorporates features such as temperature, humidity, step count (activity), and ignores irrelevant attributes to stress level such as heart rate in compliance with the latest medical research. The data was preprocessed to remove any illogical or missing values, converting to appropriate data types, then reducing dimensionality using LDA, PCA and SVD for reduced risk of overfitting, and better results. Then different statistical and deep learning modules were implemented on the dataset.

- K-Nearest Neighbors (KNN): KNN is a powerful tool in local pattern recognition, it can predict stress levels by giving the neighbor closest to k. Its capability to detect subtle relationships in the feature space makes it a precious tool for stress prediction.
- Gaussian Naive Bayes (GNB): GNB, based on probabilistic principles that assume feature independence and compute the probability of each stress level. Despite all this simplicity, GNB often performs well especially when the assumptions about independence hold true. Decision Tree: Decision Trees define a hierarchical structure of decisions using feature splits and are considered interpretable in predicting stress levels. This transparency helps to comprehend the decision-making process, which is important for informed interventions.
- Linear Discriminant Analysis (LDA): LDA is concerned with identifying those linear combinations of features which best distinguish between different classes, optimized the level of stress prediction. Its ability to reduce the dimension while maintaining class-related information improves model interpretability.
- Neural Networks: Neural Networks with their capacity to learn complex patterns offer a versatile tool in stress prediction. A neural network with hidden layers allows capturing of complex relationships between various features, thus potentially enhancing prediction accuracy.

III. Proposed Model

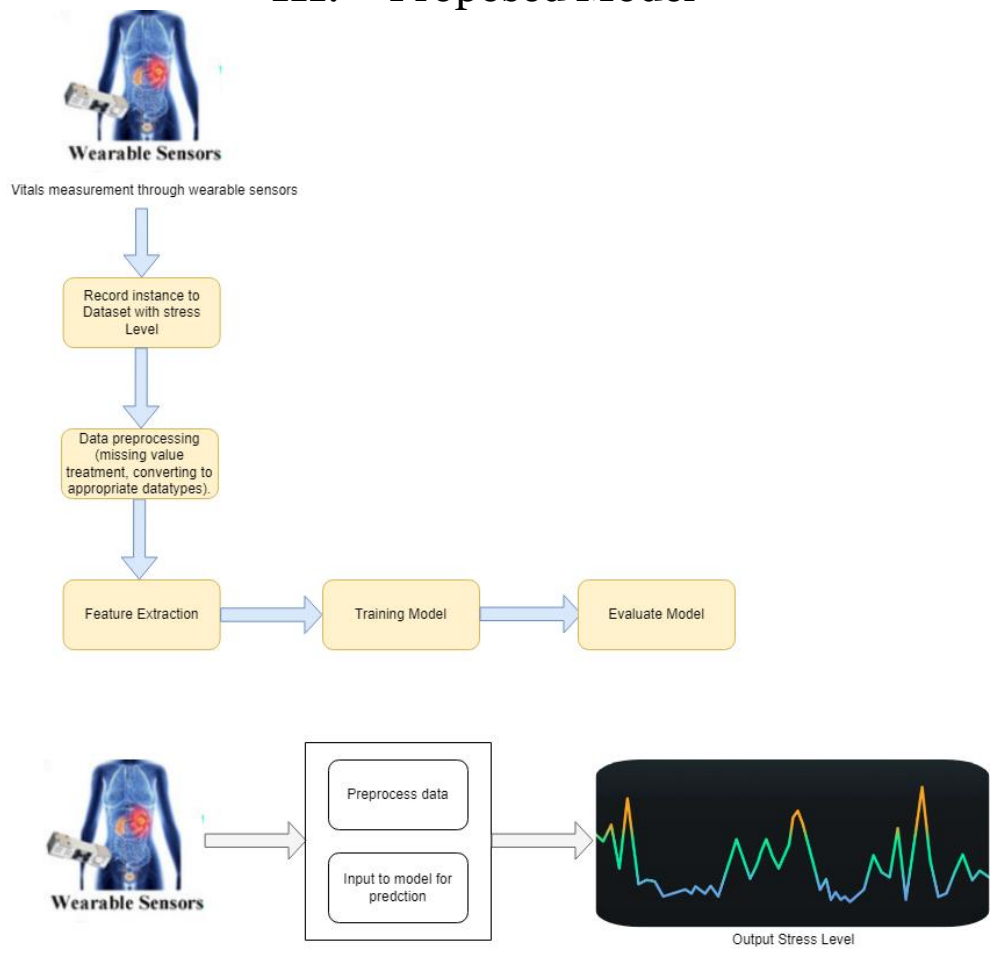


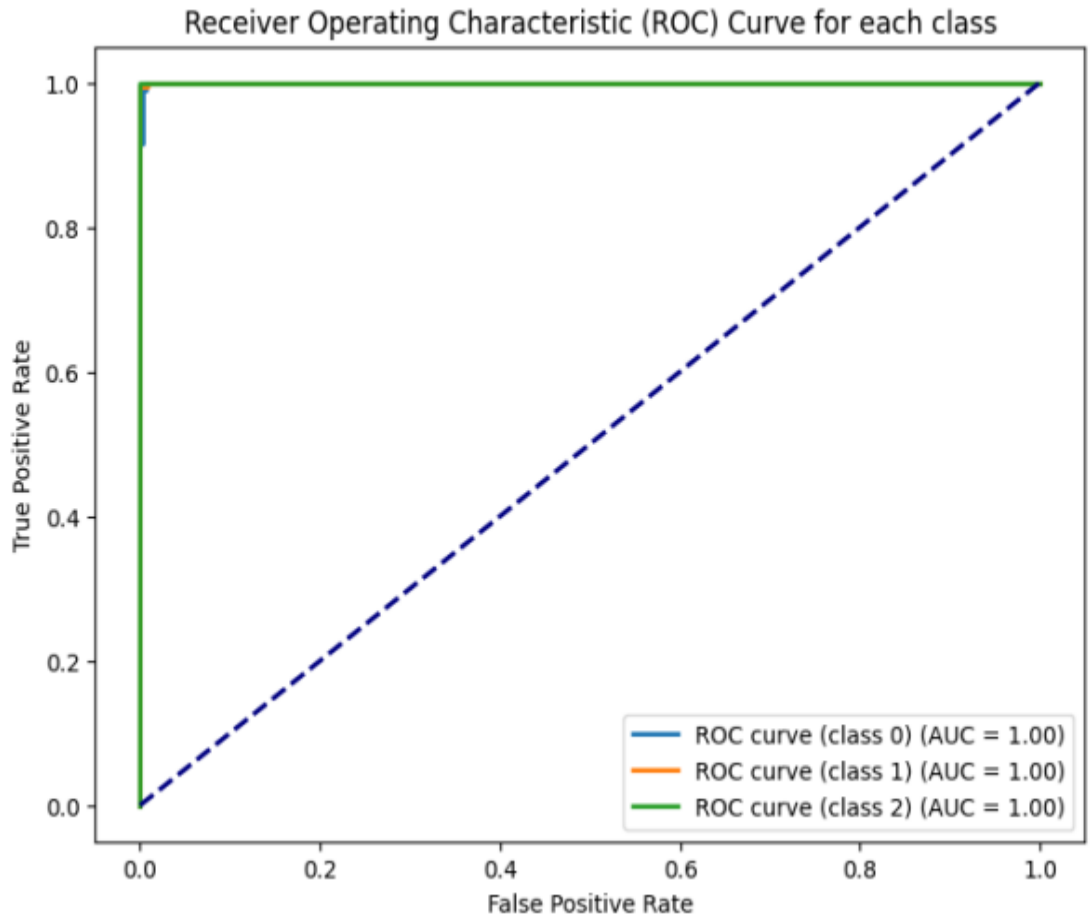
Fig 1. Proposed Model

The proposed solution included data preprocessing, dimensionality reduction for improved fitting of the data set and reducing the risk of overfitting, techniques used for dimensionality reduction were PCA, LDA and SVD. Features extraction was also used by PCA and SVD for selecting the most significant components that best describes the data set. After dimensionality reduction and feature extraction, processed data set was given to the models for learning. Different statistical and deep learning modules were implemented.

IV. Results

Model	Average K-Folds Accuracy
KNN	99.5
Decision Tree (with Error estimation)	99.7
Naïve Bayes	98.75
Bayesian Belief Network	99.4
LDA	96.5
Neural Networks	98.4

Evaluation metrics used were confusion matrix, specificity, recall, precision, K-folds cross validation, and ROC curves. Although the results are pretty high and could indicate overfitting, since the data set attributes are perfectly linear correlated, hence it would be easy for a model to find the best fitting line, and this makes the resultant accuracies reasonable.



ROC curve for Neural Networks Model

V. Conclusion and Future Work

The comprehensive analysis results in complex observations into the pros and cons of each model. Results show how different each algorithm is with local pattern recognition KNN, probabilistic modeling GNB, interpretability Decision Trees LDA for linear separation and complex pattern recognition Neural Network and emphasize on although that some models such as Neural Network would have higher accuracy than other due to its complex pattern recognition, it would perform poorly compared to the other statistical models, which also gave competitive results.

A future work for this study is to integrate the resultant models on embedded devices such IOMT for real-time stress level predictions and incorporating other significant and impacting features on stress level to the domain of the features.

References:

1. L. Rachakonda, S. P. Mohanty, E. Kougianos, and P. Sundaravadivel, “Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT,” IEEE Trans. Conum. Electron., vol. 65, no. 4, pp. 474–483, 2019.
2. L. Rachakonda, P. Sundaravadivel, S. P. Mohanty, E. Kougianos, and M. Ganapathiraju, “A Smart Sensor in the IoMT for Stress Level Detection”, in Proceedings of the 4th IEEE International Symposium on Smart Electronic Systems (iSES), 2018, pp. 141--145.
3. Mental Stress Level Prediction and Classification based on Machine Learning. In Smart Technologies, Communication and Robotics (STCR), 2021 (pp. 1-1). IEEE. DOI: 10.1109/STCR51658.2021.9588803
4. A. Ghosh, M. Danieli, and G. Riccardi, “Annotation and Prediction of Stress and Workload from Physiological and Inertial Signals,” in Proc. of 37th An. Int. Conf. of the IEEE Eng. in Med. and Bio. Soc. (EMBC), Aug 2015, pp. 1621–1624.
5. Sarada, P.A. & Ramkumar, B.. (2015). Positive stress and its impact on performance. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 6. 1519-1522.
6. Rasheed N. Prolonged Stress Leads to Serious Health Problems: Preventive Approaches. Int J Health Sci (Qassim). 2016 Jan;10(1):V-VI. PMID: 27004066; PMCID: PMC4791152.
7. Ahmadi K, Shahidi S, Nejati V, Karami G, Masoomi M. Effects of Chronic Illness on the Quality of Life in Psychiatric out patients of the Iraq - Iran War. Iran J Psychiatry. 2013 Mar;8(1):7-13. PMID: 23682246; PMCID: PMC3655233.
8. Molina-Ros, A., Moral, S., & Campos, J. (2018). Stress detection using wearable devices and machine learning techniques: A review. Sensors, 18(12), 3664.
9. Poulos, C. X., Bower, G. H., & Keltner, D. (2017). Machine learning for stress detection in speech. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017) (pp. 5562-5566). IEEE.