**Stress Detection Through Image Classification: A Comparative Study of Naive Bayes and Support Vector Machine**

**Mike Rasell Carale Dago-oc1**

Computer Science and Information Technology Department,

Bachelor of Science in Computer Science (BSCS)

Negros Oriental State University (NORSU), Negros Oriental, Philippines

|  |  |  |
| --- | --- | --- |
| **Article Info** |  | **ABSTRACT** |
| ***Keywords:***  Stress  Naïve Bayes  Support Vector Machine (SVM)  Inception-V3  Accuracy  Precision  Recall  F1-Score |  | This study explores the application of machine learning algorithms for classifying facial images into stressed and non-stressed categories. The dataset used comprises 355 facial images, sourced from Kaggle, of which a subset of 280 images (84 non-stressed and 196 stressed) was utilized due to hardware limitations. Pre-processing steps included image resizing to 299×299 resolution, normalization, and alignment with the input requirements of the Inception-V3 model for feature extraction. A comparative analysis was conducted between two traditional machine learning algorithms: Naive Bayes (NB) and Support Vector Machine (SVM). The models were evaluated using 10-fold cross-validation. The results indicated that the SVM model outperformed NB, achieving an accuracy of 82.86%, with precision, recall, and F1-score values of 0.8312, 0.8286, and 0.81, respectively. The NB model demonstrated an accuracy of 74.29%, with precision, recall, and F1-score values of 0.7797, 0.7429, and 0.75, respectively. These findings highlight the effectiveness of SVM for stress detection using facial images, though not reliable enough for real-world scenarios it still offers a promising non-invasive approach to stress monitoring and classification. |
|  |

1. **INTRODUCTION**

Stress has become a pervasive issue affecting individuals across various ages and professions. As of August 2024, 31% of adults surveyed worldwide believed that stress was the biggest health problem in their country, indicating a growing recognition of stress as a major health concern [1]. It is also reported that 51% of adults who felt stressed also reported feeling depressed, and 61% reported feeling anxious [2]. It indicates that higher levels of perceived stress are significantly associated with greater variability in negative emotional states, highlighting the complex relationship between stress and emotions [3]. Not only is stress related to the person's emotions, but it also negatively affects the person's body. Chronic stress can lead to high blood pressure, increasing the risk of heart attack and stroke, this shows that stress is also a significant risk factor for various cardiovascular conditions [4][5].

This study aims to make a machine-learning model that uses facial images as inputs to classify a person's emotional state whether the person is stressed or not stressed. While deep learning models have shown success in emotion recognition, simpler and interpretable machine learning models, such as Naive Bayes, Support Vector Machine, and Logistic Regression, remain underexplored in the context of stress detection using only feature-extracted facial images. Therefore, a comparison of these model's performance on stress classification is necessary to determine their viability and practicality in real-world applications. With this tool, stress can be detected early which can lead to timely interventions, improving quality of life and preventing adverse health outcomes. Additionally, traditional stress detection methods, such as surveys and biometric sensors, can be intrusive, expensive, or time-consuming [6][7]. This tool offers a non-invasive, cost-effective alternative that can be applied in real-time scenarios.

1. **LITERATURE REVIEW**

Stress detection through machine learning algorithms has gained significant attention due to its non-invasive applications in mental health monitoring. Traditional machine learning models, such as Naive Bayes (NB) and Support Vector Machine (SVM), offer interpretable and computationally efficient approaches compared to complex deep learning models. This section explores existing literature on stress detection using NB and SVM, highlighting methodologies, performance, and applicability.

**2.1. Naïve Bayes Algorithm**

The Naive Bayes algorithm, rooted in Bayes' theorem, is known for its simplicity and speed. It performs well in scenarios with limited data and when preprocessing is optimized to address class imbalances or feature inconsistencies. NB's use in stress detection has shown promise, with studies reporting precision scores as high as 78% when normalized features are employed. However, NB's assumption of feature independence can lead to limitations, especially in datasets like facial images where features are interdependent [8]. Despite this limitation, NB has achieved notable results in stress detection tasks. In another example, researchers obtained an accuracy of 74% using NB to classify stress from facial images, underlining its potential in preliminary analyses and computationally constrained settings [9].

**2.2. Support Vector Machine Algorithm**

Machine learning approaches have been extensively employed in stress detection tasks due to their ability to process and classify complex data effectively. SVM is widely recognized for its robust performance in binary classification tasks by optimizing hyperplanes in high-dimensional spaces. A study employing SVM to classify stress levels using facial landmark features achieved an accuracy of 83%, demonstrating its capacity to capture stress-related features such as facial tension and asymmetry. SVM relies on kernel functions to transform data into higher-dimensional spaces, enabling it to draw optimal decision boundaries even in overlapping feature distributions. Comparative studies have shown SVM outperforming NB, achieving higher precision, recall, and F1-scores. For instance, SVM achieved an accuracy of 83% in stress detection tasks, outperforming NB across all metrics [9][10].

1. **METHODOLOGY**

**3.1. Materials**

**3.1.1 Datasets**

This study employs a dataset from Kaggle which is the Stress Non-Stress Images dataset, a file with 355 jpeg files which are facial images categorized based on a person's emotional state which is Non-Stressed or Stressed. The Non-Stressed class includes emotions such as being happy and neutral, while the Stressed class includes emotions such as being sad and angry [11]. Because of hardware limitations, this study uses 280 images with 84 Non-Stressed and 196 Stress images.

**3.1.2. Hardware**

The study was conducted on a system running Windows 10 with a 64-bit operating system. The hardware specifications included an Intel Core™ i5-7400 CPU and 16GB of Random Access Memory (RAM).

**3.1.3. Software**

This study utilized Jupyter and Python version 3.9.15 as the primary programming environment for data analysis and model implementation. The following libraries were employed: Math, NumPy, Pandas, PIL, OS, Scikit-Learn, Matplotlib, Seaborn, TensorFlow, and Keras.

**3.2. Methods**

**3.2.1. Data Pre-processing**

Data processing is an essential aspect of model development. Data acquired in their raw form contain noise and anomalies, which can affect the performance and training process of the model being schooled. In addition to this when preparing input data for a Inception-V3, it's essential to adhere to specific requirements regarding input shape, color channels, and preprocessing steps to ensure optimal performance [14]. The researchers employed several techniques to clean the data, which included image pre-processing and data normalization:

Image pre-processing is an essential part of image classification to ensure that input images align with the model's expectations. Inception-V3 requires input images in 299x299 resolution and should also have three color channels (red, green, and blue) [15][16]. And since the image datasets have higher resolution than the required input, the image is resized to have a resolution of 299x299. After this, the image is preprocessed using TensorFlow Keras API preprocess\_input so that the images align with the model's training conditions.

Data normalization is a pre-processing technique primarily intended to manage numerical features and is applied to numerical features before the application of classification algorithms. Normalization is crucial to prevent the effect of certain features from being concealed by others, particularly when the ranges of the features are inconsistent [17]. After feature extraction using Inception-V3, the features are normalized to improve model training [18]. The normalization techniques used are Min-Max Normalization for Naive Bayes (NB) and Z-Score Normalization for the Support Vector Machine (SVM).

**3.2.2. Feature Extraction**

The main goal and objectives of the proposed system are to diagnose and create a tool that can identify whether a person has pneumonia through chest X-ray images. This study uses Inception-V3 for feature extraction of the X-ray images.

Inception models were developed for the first time in 2014. The structures of inception models and the conventional convolutional neural networks (CNN) model are different because they are inception blocks which means lapping the same input tensor with multiple filters and concatenating their results. In 2015, a researcher proposed a new version of the inception models named Inception-V3, an improved version of the previous versions of inception models which are Inception-V1 and Inception-V2, and possesses 24M parameters. Inception-V3 improves the efficiency and performance of CNNs by introducing clever factorization techniques. Instead of directly using large convolutions (e.g., n x n), it breaks them into smaller, more manageable operations. For example, a 5 x 5 convolution is replaced by two 3 x 3 convolutions, and an n x n convolution is split into asymmetric 1 x n and n x 1 convolutions, reducing computation. Additionally, 7 x 7 convolutions are replaced with multiple 3 x 3 convolutions. Each inception block processes the input in parallel through multiple convolutional filters of different sizes (1 x 1, 3 x 3, and 5 x 5), as well as 3 x 3 max pooling, capturing features at various scales. These outputs are then concatenated and passed to the next module, allowing the network to learn complex features efficiently [19].

**3.2.3. Machine Learning Algorithms**

This section focuses on the machine learning classification models utilized in this study. After image and data pre-processing, the machine learning workflow progresses to the model training stage, where an algorithm is taught to learn from data and produce predictions. This algorithm is specifically responsible for the classification of facial images whether the person is stressed. To find the best classifier for this task, two classifiers, which include Naive Bayes (NB) and Support Vector Machine (SVM) were tested through a variety of tests.

**3.2.3.1. Naïve Bayes**

The Naïve Bayes algorithm has its foundation rooted in the Bayes theorem by Thomas Bayes. One of the strengths of this model is its ability to handle missing values. And Unlike other models, Naïve Bayes conserves processing and training time [20]. The term ‘naive’ is used due to this algorithm's uncertain independence. With this, stated that with this ability it's able to converge quicker when compared to several others [20].

(1)

Where P is the probability, X is the training set of attributes and Y is the given class.

**3.2.3.2. Support Vector Machine**

Support Vector Machine (SVM) is a binary linear classifier. As a non-probabilistic supervised learning algorithm, it utilizes training data and employs a high-dimensional space to construct a set of hyperplanes for data classification. While only the features of test data are provided, the model is trained on the training data to predict the target values. For effective classification of problem instances, SVM relies on selecting the optimal hyperplane [21].

The formula for the Support Vector Machine (SVM) decision boundary can be expressed as:

(2)

Where is the weight vector, determining the orientation of the hyperplane,  is the feature vector of the input data,  is the bias term, shifting the hyperplane.

**3.2.4 Model Evaluation Metrics:**

Evaluation measures are metrics used to assess the results of an experiment [22]. In the context of classification models, different evaluation metrics are used to measure their output. In this study, the main performance evaluation metric is “Accuracy”. However, additional metrics such as recall, precision, f-measure, and confusion matrices are also used to supplement the evaluation of the model's performance. Each model identifies learner engagement levels when assessed using these metrics. A brief description of these metrics is provided below.

Accuracy is a common evaluation metric for classification models. It's calculated as the ratio of well-predicted samples to the total sample of prediction. For a balanced dataset, accuracy is a reliable measure of the model's performance.

(3)

Where the numerator reflects the total number of correct predictions, while the denominator represents the total number of predictions made by the model. A higher accuracy value suggests that the model is more effective at correctly classifying both classes [21].

Precision measures the proportion of correctly predicted positive observations out of all the predicted positive observations. A high precision score indicates strong class predictions, while a low precision score reflects weak class predictions.

(4)

Where True Positives refer to the instances that were correctly identified as positive, while False Positives are the instances where the model wrongly predicted the positive class [21].

Recall is the ratio of correctly predicted positive results to all actual positive samples, also known as the detection rate. It's calculated by dividing the true positive samples by the sum of the positive samples.

(5)

Where False Negatives occur when the model mistakenly classifies a positive instance as belonging to the negative class [21].

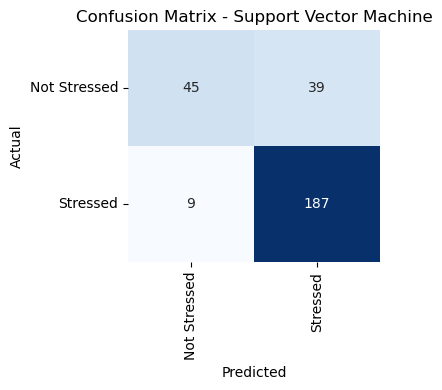
F1-score is the mean value for recall and precision. It offers an indicator of mistakenly graded results [21]. It is regarded as the best metric for measuring the performance of models on an imbalanced dataset. It ranges from 0 to 1, with higher values indicating better model performance.

(6)

Where Precision evaluates the accuracy of positive predictions and Recall assesses the model's ability to identify all relevant instances, the F1-Score combines these two metrics into a single value, offering a balanced measure of the model's accuracy and completeness [21].

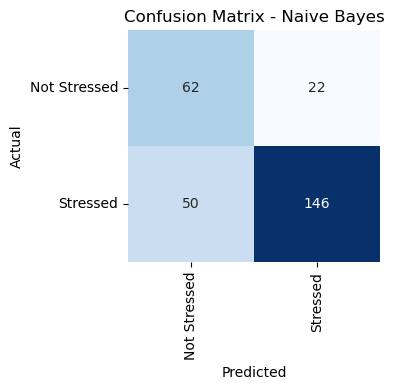
1. **RESULTS AND DISCUSSION**

**4.1. Confusion Matrix**

****

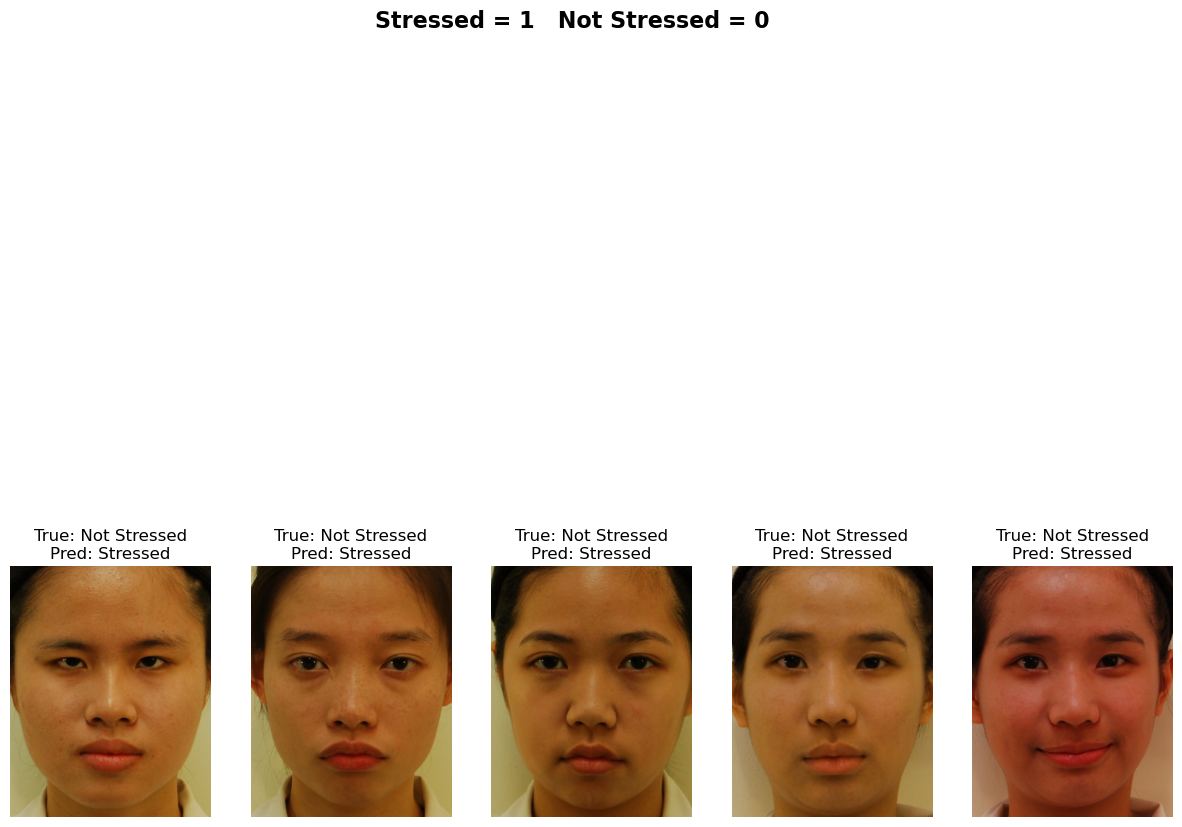
**Figure 1.** SVM Confusion Matrix

In Figure 2, the confusion matrix for the SVM model shows that it performs well in classifying the “Stressed” class with a high true positive count (187) and a low false negative count (9), which indicates strong recall for the stressed class. However, it struggles more with the “Not Stressed” class, as seen in the higher number of false positives (39), where “Not Stressed” was misclassified as “Stressed”. Overall, SVM seems better at identifying “Stressed” instances than “Not Stressed” ones, which might suggest a bias towards the “Stressed” class.

****

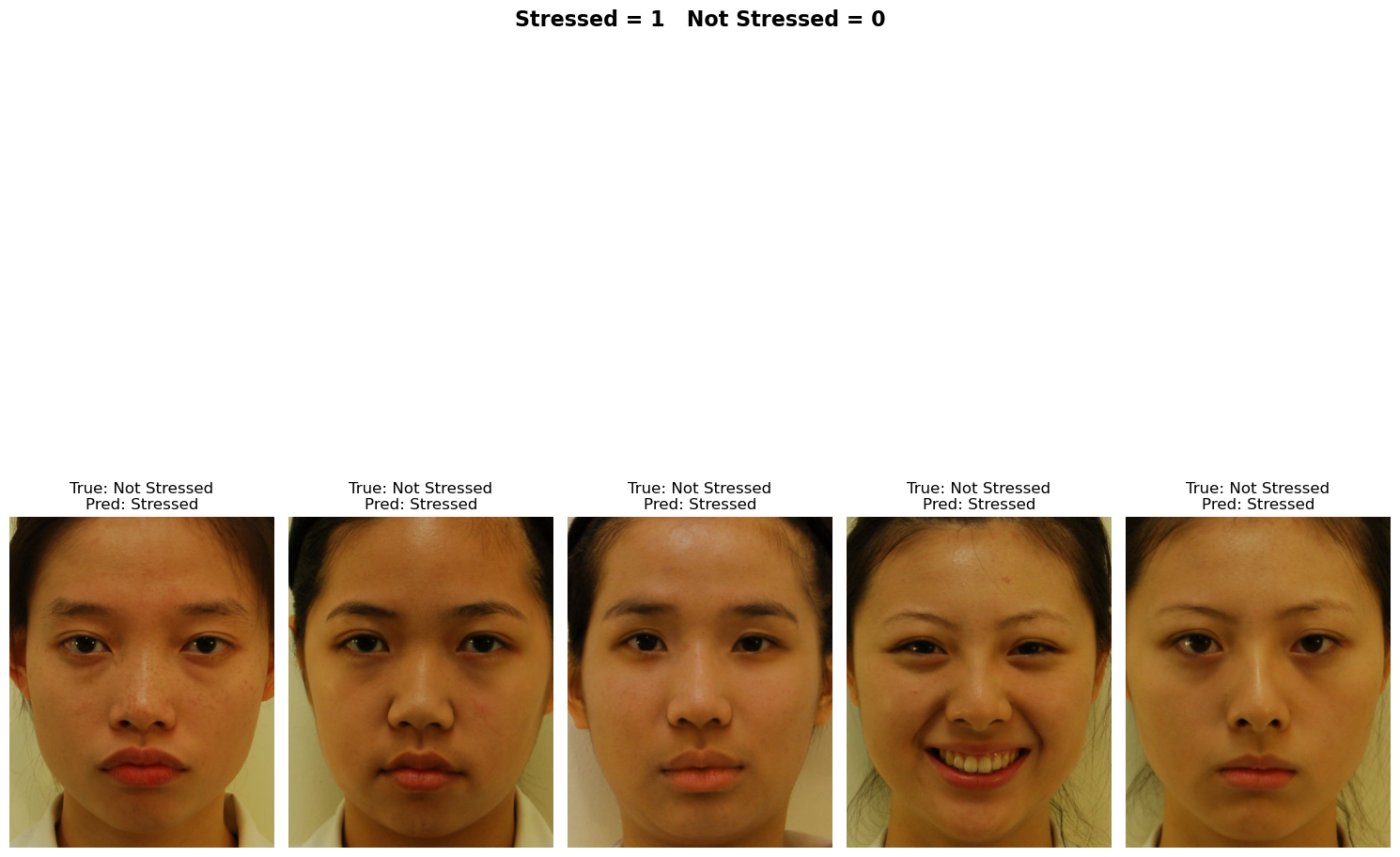
**Figure 2.** NB Confusion Matrix

In contrast in Figure 3, NB shows better performance in identifying the “Not Stressed” class compared to SVM, with more true negatives (62) and fewer false positives (22). However, it performs worse for the “Stressed” class, as it has a lower true positive count (146) and a much higher false negative count (50) than SVM. This suggests Naive Bayes might underperform in capturing “Stressed” instances, potentially due to its assumption of feature independence, which may not hold well in this dataset.

**4.2 Misclassified Image Analysis**

SVM has a consistent misclassification of “Not Stressed” as ”Stressed”. This suggests the model is overly sensitive to features it associates with stress like certain facial expressions, redness, or tension in the face, leading to false positives. The model might also be picking up on specific features such as slight frowns, facial asymmetry, or other expressions, which may not truly indicate stress but are being interpreted as such by the SVM. And since the expressions in the “Not Stressed” category appear ambiguous or might share similarities with features the model associates with stress, like redness in the face, this contributed to the misclassification. Also in some images redness or lighting variations might be interpreted by the model as stress, even when it’s unrelated.

**Figure 3.** SVM Misclassified Images

 NB mostly misclassified “Stressed” images as “Not Stressed” which is the opposite of SVM. This shows that some faces lack common stress indicators like furrowed brows and tensed jaws, making it harder for the model to distinguish. And in the misclassified “Not Stressed” images as”Stressed”, because of misleading cues. For instance, intense facial expressions or lighting effects may have contributed to false positives. The misclassifications from NB also appear more random or less pattern-dependent compared to SVM, which uses a decision boundary to classify. This shows that the misclassification rate is due to its inability to model feature dependencies. Stress indicators like furrowed brows and mouth tension are likely correlated, and NB cannot effectively handle these interdependencies. NB tends to be simpler and faster, though it lacks the complexity required for nuanced tasks like facial expression analysis.

**Figure 4.** NB Misclassified Images

**4.3. Model Performance**

**Table 1**. Performance Metrics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1-Score |
| SVM | 0.8286 | 0.8312 | 0.8286 | 0.81 |
| NB | 0.7429 | 0.7797 | 0.7429 | 0.75 |

In the evaluation of the two machine learning models using a 10-fold cross-validation, in Table 1 it can be observed that the performance of all the model's accuracy ranges from 74% to 83%. The result shows that all the models can identify whether the facial images of the subject are stressed. And with a thorough analysis of the results, indicates that SVM provided the highest accuracy of 82.86%, with a precision of 0.8312, a recall of 0.8286, and an F1-score of 0.81. The NB model, underperformed showed lower accuracy compared to SVM, with an accuracy of 74.29%, a precision of 0.7797, a recall of 0.7429, and an F1-score of 0.75. These results suggest that SVM outperforms Naive Bayes across all metrics, suggesting it is better suited for this task. The higher accuracy and balanced precision-recall values make it more reliable. While NB still performs reasonably well, it underperforms compared to SVM due to its simplicity.

1. **CONCLUSION**

The SVM model demonstrates good overall performance in the image classification task, achieving consistently high metrics around 81-83%. The model effectively identifies “Stressed” images but struggles with “Not Stressed” images, leading to high false positives and low precision for this class. In its current state, the model may not be reliable enough for real-world scenarios where correctly identifying "Not Stressed" individuals is critical.

While the NB model exhibits lower performance compared to SVM with metrics ranging between 74–77%. It is more balanced compared to SVM in identifying both classes but tends to miss “Stressed” cases while performing better for the “Not Stressed” class. It also has higher misclassification rates due to its inability to model feature dependencies. Stress indicators like furrowed brows and mouth tension are likely correlated, and NB cannot effectively handle these interdependencies. Overall, NB underperforms in this case because the facial features extracted are unlikely to be independent, and the assumptions of NB may not hold true.

This study confirms that stress is a complex state and can manifest in ways not easily captured in static facial expressions alone. This might explain why the models struggle to distinguish between subtle cases of stress and non-stress. This study also confirms that SVM is the most effective model for classifying facial images if the person is stressed, this is expected since SVM handles complex, high-dimensional data like Inception V3 features better.

**REFERENCES:**

[1] Statista, "Stress as a top health concern worldwide 2023," Jan. 2025. [Online]. Available: https://www.statista.com/statistics/1498280/stress-as-a-top-health-concern-worldwide/. [Accessed: Jan. 20, 2025].

[2] Mental Health Foundation, "Stress statistics," [Online]. Available: https://www.mentalhealth.org.uk/explore-mental-health/statistics/stress-statistics. [Accessed: Jan. 20, 2025].

[3] National Center for Biotechnology Information, "The Relationship between Psychological Stress and Emotional State in Chinese University Students during COVID-19: The Moderating Role of Physical Exercise" PubMed Central, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10001233/. [Accessed: Jan. 20, 2025].

[4] American Heart Association, "Chronic stress can cause heart trouble," Feb. 4, 2020. [Online]. Available: https://www.heart.org/en/news/2020/02/04/chronic-stress-can-cause-heart-trouble. [Accessed: Jan. 20, 2025].

[5] National Center for Biotechnology Information, "The effects of stress on physical health and its implications," PubMed Central, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7603890/. [Accessed: Jan. 20, 2025].

[6] MDPI, "Electronics | Free Full-Text | A New Era in Stress Monitoring: A Review of Embedded Devices and Tools for Detecting Stress in the Workplace," [Online]. Available: https://www.mdpi.com/2079-9292/13/19/3899. [Accessed: Jan. 20, 2025].

[7] National Center for Biotechnology Information, "Stress: Concepts, Cognition, Emotion, and Behavior," PubMed Central, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6749249/. [Accessed: Jan. 20, 2025].

[8] National Center for Biotechnology Information, "The Impact of Stress on Health: A Review of the Literature," PubMed Central, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8625615. [Accessed: Jan. 20, 2025].

[9] ScitePress, "Stress Detection Using Machine Learning: A Review," [Online]. Available: https://www.scitepress.org/PublishedPapers/2021/104742/104742.pdf. [Accessed: Jan. 20, 2025].

[10] National Center for Biotechnology Information, "The Impact of Stress on Health: A Review of the Literature," PubMed Central, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC8625615. [Accessed: Jan. 20, 2025].

[11] P. Bhagat, "Stress vs Non-Stress Images Dataset," Kaggle, [Online]. Available: https://www.kaggle.com/datasets/preritbhagat/stress-non-stress-images. [Accessed: Jan. 20, 2025].

[12] C. Szegedy, W. Liu, Y. Jia et al., “Going deeper with convolutions,” Technical report, Cornell University, New York, NY, USA, 2014.

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, “Rethinking the inception architecture for computer vision,” 2015, https:// arxiv.org/abs/1512.00567.

[14] Cmrad, "The Ultimate Guide to Preprocessing Medical Images: Techniques, Tools, and Best Practices for Enhanced Diagnosis," [Online]. Available: https://about.cmrad.com/articles/the-ultimate-guide-to-preprocessing-medical-images-techniques-tools-and-best-practices-for-enhanced-diagnosis. [Accessed: Jan. 20, 2025].

[15] Keras, "InceptionV3: Pre-trained model for image classification," [Online]. Available: https://keras.io/api/applications/inceptionv3/. [Accessed: Jan. 20, 2025].

[16] PyTorch, "Inception V3 | PyTorch," [Online]. Available: https://pytorch.org/hub/pytorch\_vision\_inception\_v3/. [Accessed: Jan. 20, 2025].

[17] M. C. M. Pasague, "Classifying Garbage Images: A Comparative Study of Naive Bayes, SVM, and Logistic Regression Models," Bachelor of Science in Computer Science, Negros Oriental State University, 2025. [Online]. Available: cristinapasague27@gmail.com.

[18] MDPI, "Deep Learning for Stress Detection: A Survey," Healthcare, vol. 6, no. 2, p. 11, [Online]. Available: https://www.mdpi.com/2306-5729/6/2/11. [Accessed: Jan. 20, 2025].

[19] E. F. Buraimoh, “Predicting Student Success Using Student Engagement in the Online Component of a Blended-Learning Course,” pp. 1–65, 2021.

[20] G. Theophilus and C. I. Eke, "Machine learning-based e-learners' engagement level prediction using benchmark datasets," Int. J. Appl. Inf. Syst., vol. 12, no. 41, Sept. 2023.

[21] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, “Student Performance Prediction Model based on Supervised Machine Learning Algorithms,” IOP Conf. Ser. Mater. Sci. Eng., vol. 928, no. 3, 2020, doi: 10.1088/1757- 899X/928/3/032019.

[22] C. I. Eke, A. A. Norman, and L. Shuib, Multi-feature fusion framework for sarcasm identification on twitter data: A machine learning based approach, vol. 16, no. 6