



# OPEN MyWear revolutionizes real-time health monitoring with comparative analysis of machine learning

Krishna Prakash<sup>1✉</sup>, Musam Naga Harshitha<sup>2</sup>, Golla Naga Lakshmi<sup>2</sup>, Pallem Moses<sup>2</sup>, Madala Sumanth Chowdary<sup>2</sup>, Shonak Bansal<sup>3✉</sup>, Mohammad Rashed Iqbal Faruque<sup>4✉</sup> & K. S. Al-Mugren<sup>5</sup>

This paper facilitates proactive health management, advanced patient care, and early identification of possible health hazards by using MyWear. It is a wearable T-shirt that continuously monitors and predicts physiological parameters such as stress and heart rate fluctuations. In particular, it is especially helpful for managing cardiovascular disease, tracking stress, improving athletic performance, and providing health care. The device was tested with several machine learning models, such as K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, Decision Tree, and Stochastic Gradient Descent (SGD) to identify irregular heart rhythms. Using the SVM model, the system detects problems with an average accuracy of 98%. In the future, MyWear—designed as a wearable T-shirt—will seamlessly integrate with mobile applications for real-time data visualization, enhancing patient outcomes and fostering greater user engagement.

**Keywords** Heart rate variability (HRV), Stress level detection, Machine learning, K-nearest neighbour (KNN), Support vector machine (SVM)

Over the years, the convergence of healthcare and technology has triggered a transformative paradigm shift towards personalized medicine and proactive health management. The emergence of wearables, such as the ability to read physiological parameters in real-time, and novel solutions for health improvement. This smart garment provides an intelligent analysis of physiological conditions such as heart rate, muscle activity, and any elements of stress measurement or detection<sup>1</sup>. Heart Rate Variability (HRV) refers to variation in time between two consecutive heartbeats, which is an important indicator of the function of the autonomic nervous system (ANS)<sup>2,3</sup>. In addition, HRV is also a critical indicator of stress, anxiety, and heart function<sup>4</sup>. Sethuraman et al.<sup>1</sup> developed MyWear, which precisely classifies the heartbeat anomalies and predicts the possible health risks, enabling proactive healthcare management. It is an alert system that sends real-time alerts to medical staff during emergencies, filling the gap between users and healthcare providers. The continuous operation of sensors and communication modules, such as Wi-Fi and Bluetooth, demands special focus on battery life as well as power consumption to allow the garment to track physiological information for an extended period without periodic recharging. However, MyWear seeks to disrupt personal healthcare for a better quality of life. Fred Shaffer et al. categorized HRV metrics into time-domain measures, frequency-domain measures, and non-linear measures<sup>2</sup>. They highlighted that normal HRV values vary across different populations. They found clinical implications of HRV; lower HRV causes increased morbidity and mortality. Further, Reena Tiwari et al. worked on the analysis of HRV. High HRV indicates better adaptability to stress and cardiovascular fitness associated with resistance against cardiac diseases, along with mortality risk, whereas low HRV represents an unhealthy heart<sup>3</sup>. The function of HRV is to convey information regarding the behaviour of the sympathetic and parasympathetic nervous systems. Furthermore, Haque et al. highlighted the significance of HRV and improvement in wearable

<sup>1</sup>Department of Electronics and Communication Engineering, NRI Institute of Technology, Agiripalli, Eluru, Andhra Pradesh 521212, India. <sup>2</sup>Department of CSE (AIML), NRI Institute of Technology, Agiripalli, Eluru, Andhra Pradesh 521212, India. <sup>3</sup>Department of Electronics and Communication Engineering, Chandigarh University, Gharuan, Punjab, India. <sup>4</sup>Space Science Centre (ANGKASA), Institute of Climate Change (IPI), Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor D. E., Malaysia. <sup>5</sup>Physics Department, Science College, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. ✉email: k\_krishna2k7@yahoo.co.in; shonakk@gmail.com; rashed@ukm.edu.my

sensors for continuous monitoring<sup>4</sup>. He examined various AI algorithms, including shallow and deep learning techniques. He also mentioned the importance of data pre-processing and multimodal integration, along with individual variability and data scarcity challenges.

On the other hand, Neil et al.<sup>5</sup> highlighted the effects of psychosocial stressors such as childhood trauma, life events, and chronic stress leading to anxiety, depression, and physical health issues. Further, Pia et al.<sup>6</sup> identified that self-efficacy significantly mediates the relationship between daily stressors and mental health outcomes. Specifically, self-efficacy serves as a stronger buffer for positive mental health as compared to negative mental health. Enhancing self-efficacy could be a valuable target to reduce the adverse effects of daily stress on mental health. Mario Salai et al.<sup>7</sup> developed a reliable and low-cost method for stress detection using heart rate data from a chest belt sensor. It consists of two phases: a comparison of the CardioSport TP3 sensor and gold standard Holter device, and a clinical study involving 46 participants to perform relaxing and stress-inducing tasks using the Stroop color test. They implemented data cleaning and synchronization techniques for high-quality data. Their study focused on time-domain features to detect stress. Their study found that low-cost heart rate sensors can effectively detect stress levels. Yaribeygi et al.<sup>8</sup> stated that stress affects brain function, particularly memory and cognition, the immune system, cardiovascular health, gastrointestinal function, and the endocrine system. Deepti et al.<sup>9</sup> conducted a study on the causes of stress and stress management among youth. Their analysis highlights that stress is common among youth due to academic pressures, interpersonal relationships, financial concerns, and career uncertainties. They identified symptoms of stress such as depression, insomnia, and unhealthy eating habits. The positive environment and outdoor activities help to manage the stress among youth. Later, Zafar et al.<sup>10</sup> highlighted that stress affects individuals differently, the role of stress hormone glucocorticoids, which alters brain function and causes mental health disorders. A balanced diet, yoga, and mindfulness practices reduce stress in individuals.

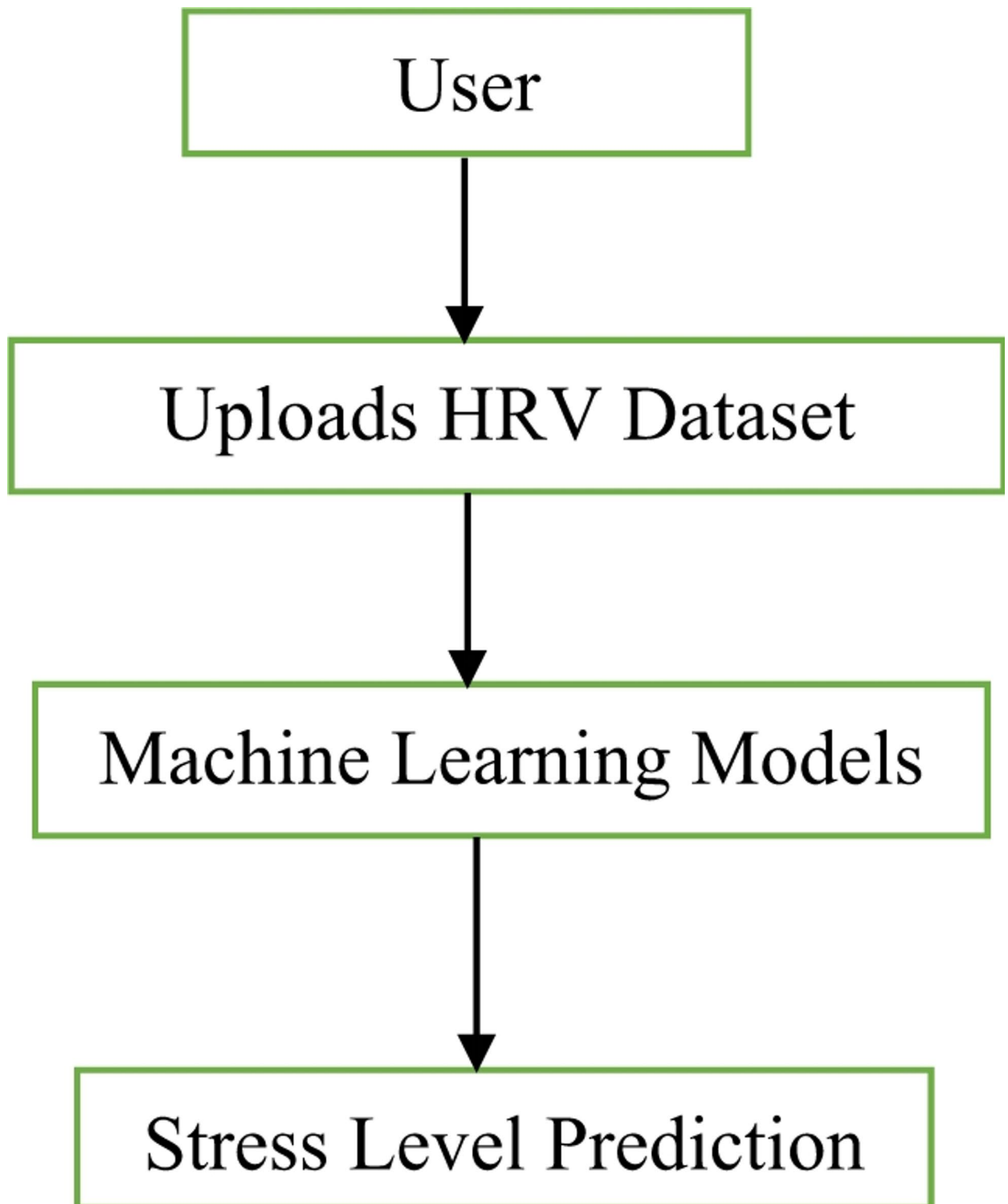
Jon et al.<sup>11</sup> developed a one-dimensional convolutional neural network (CNN) for multi-class stress classification, consisting of multiple layers for automatic feature extraction and learning. The model is trained using the ADAM optimizer and evaluated through evaluation metrics and Matthew's correlation coefficient (MCC). Using Hyperband tuning, hyperparameter optimization is performed for model performance. The Random Forest Model is utilized to manage complex relationships in the data<sup>12</sup>. Hamadneh, Tareq et al.<sup>13</sup> proposed a novel algorithm, Sales Training Based Optimization (STBO), designed to address Sustainable Lot Size Optimization (SLSO) in supply chain management. Inspired by sales training habits, STBO operates in two phases: exploration and exploitation. It outperforms twelve state-of-the-art metaheuristic algorithms across ten test scenarios by consistently achieving the lowest mean costs. The authors conclude that STBO is a stable optimization algorithm for SLSO problems and recommend future studies to explore its extension to other optimization tasks. In a separate study, Hamadneh, Tareq et al.<sup>14</sup> introduced the Orangutan Optimization Algorithm (OOA), inspired by the foraging and nesting behaviors of orangutans. OOA effectively balances exploration and exploitation to solve complex optimization problems. It demonstrates superior performance over twelve widely used metaheuristic algorithms when tested on the twenty-two problems in the CEC 2011 benchmark suite. The authors conclude that OOA is a robust optimization tool with potential for further research in binary and multi-objective problem domains.

The previously published paper on HRV dynamical properties forms the basis for MyWear. The development of the smart garment that has been proposed would make use of a web application developed with the Django framework which would allow for interaction with the users and visualization of data easily. MyWear protects health data storage and transmission via AES128 encryption and limits access to the user via a mobile app<sup>1</sup>. The system uses secure communication protocols and strong cloud security features to ensure against potential hacking. To encounter motion artifacts in MyWear, we used signal filtering and data pre-processing methods, allowing our machine learning algorithms to properly separate true physiological signals from noise due to user movements. The users of the system will be able to check their health parameters, get notifications in case of abnormal values, and track their stress. The major aims are to create an advanced smart garment that connects new sensors and employs machine learning algorithms to track health parameters, including HRV and stress levels, to promote proactive health management by giving users instant feedback about their physiological states, seeks to enable people to manage their health and enhance well-being. Existing wearable technology tends to concentrate mainly on physical parameters, and there is a huge gap in real-time stress tracking and comprehensive understanding of health, which MyWear aims to fill. A set of machine learning models, such as K-Nearest Neighbours (KNNs), Naive Bayes, Support Vector Machines (SVM), Logistic Regression, Decision Tree, and Stochastic Gradient Descent (SGD) are developed to increase the precision and accuracy of the health assessments and the recommendations supplied to the clients. The workflow of the proposed system is illustrated in Fig. 1.

## Model and methodology

MyWear's custom machine learning pipeline aims to effectively identify stress and heart rate changes via a scientific method. It includes gathering information from built-in sensors of the wearable T-shirt and then pre-processing to normalize the data. Appropriate features are derived, and several machine learning methods, such as K-Nearest Neighbours, Support Vector Machine, and Naive Bayes are used. Training is performed and validated using performance evaluation measures like F1-score and accuracy, before incorporating the optimal performing model within an easy-to-use application used for real-time feedback and tracking.

In machine learning, Naive Bayes is a group of classification algorithms derived from the Bayes theorem and assumes that the predictors are independent. It is very useful for classification problems if the input data contains categorical variables. In this work, the variant of MultinomialNB is used, which is preferable for text classification problems. It predicts the stress levels with a larger probability by calculating and comparing all classes with extracted features. SVM is a supervised machine learning algorithm that works by locating the best



**Fig. 1.** Working procedure of proposed system.

hyperplane in high-dimensional space that splits the multidimensional vector into two classes of points. This is to maximize the distance among the nearest points of classes (support vectors). By using the Linear SVC class, which is a linear Support Vector Classifier. This method then finds a linear decision boundary that will group the input data into one of many stress levels. Logistic regression is a statistical model that uses a logistic function to predict the outcome of a categorical dependent variable. It predicts that a certain input point is present in the advanced class. One of the two prediction models used in this work is to predict the stress levels under a given set of features. It predicts probabilities, and the threshold eventually converts that to class labels.

Decision Trees are tree-based models for classification and regression<sup>15</sup>. They divided the data into subsets by the value of input features and continued to branch recursively until some stopping criterion was reached (maximum depth or minimum samples per leaf, for example). Finally, a Decision Tree Classifier is utilized to create a model that makes decisions based on the heart rate data features and classifies the respective stress levels. Efficient implementation of linear classifiers (SVM and Logistic Regression) uses stochastic gradient descent to optimize the parameters of the model. Especially in the case of large datasets. In this work, it is used to approximate a linear relationship for forecasting stress from the features. KNN is a non-parametric, instance-based learning algorithm that compares new data points with the majority class (out of k-nearest neighbours) by considering their proximity in the feature space. Uses distance metrics like the Euclidean Distance to find neighbours. K-Nearest Neighbours Classifier is then used to identify the stress levels by examining the nearest points in the training set, using a simple majority vote among its neighbours.

## Results and discussion

The proposed garment integrates with multiple machine learning models using Naive Bayes, SVM, Logistic Regression, Decision Tree, SGD, and K-Nearest Neighbours to identify the stress levels based on heart rate data. The Naïve Bayes model has predicted the stress levels with an average accuracy of 97% and precision of 98%, as illustrated in Table 1 and Fig. 2A by using the following formula as mentioned in<sup>16</sup>:

$$P\left(\frac{C_i}{X}\right) = \frac{P\left(\frac{X}{C_i}\right) P(C_i)}{\sum_{i=1}^n P\left(\frac{X}{C_i}\right) P(C_i)} \quad (1)$$

where  $C_i$  represents a certain class,  $X$  represents features,  $i$  represents the index of a specific class,  $n$  represents a total number of classes, the model calculates the posterior probabilities, and makes predictions.

The SVM model has predicted the stress levels using heart rate data with an average accuracy of 98% and precision of 98%, as illustrated in Table 1 and Fig. 2B, by using the following formula as mentioned in<sup>17</sup>:

$$\min_{w, \xi_i} \|w\|^2 + C \sum_{i=1}^l \xi_i \text{ subject to : } y_i (w \cdot x_i + b) \geq 1 - \xi_i, \text{ for all } i \text{ and } \xi_i \geq 0 \quad (2)$$

where  $w$  represents a weight vector that defines the orientation of the hyperplane and is perpendicular to the hyperplane; its magnitude determines the margin.  $\|w\|^2$  represents the squared norm of the weight vector to maximize the margin among the classes,  $C$  denotes regularization parameter,  $\xi$  denotes a slack variable that allows misclassification,  $y$  represents class labels,  $x$  represents a feature vector,  $i$  represents a training sample,  $b$  represents bias that allows the hyperplane to shift away from the origin,  $t$  represents the total number of training samples.

The regression model is powerful as well as used significantly to reduce the time<sup>18,19</sup>. The regression model is used to investigate relationships among the input parameters and the output variables. Regression models can predict critical parameters<sup>15,20,21</sup>. The Logistic Regression model got an average accuracy of 98% and precision of 98% as illustrated in Table 1 and Fig. 2C, by using the following hypothesis formula as mentioned in<sup>22</sup>:

$$P_i(x) = \frac{\exp(\beta x)}{1 + \exp(\beta x)} \quad (3)$$

where  $i$  represents certain classes,  $\beta$  represents coefficients correlated with input features,  $x$  represents the feature vector, the model calculates probability, and makes predictions.

The Decision Tree model has predicted the different stress levels based on heart rate data with an average accuracy of 97% and precision of 98%, as illustrated in Table 1 and Fig. 2D, by using the following formula as mentioned in<sup>23</sup>:

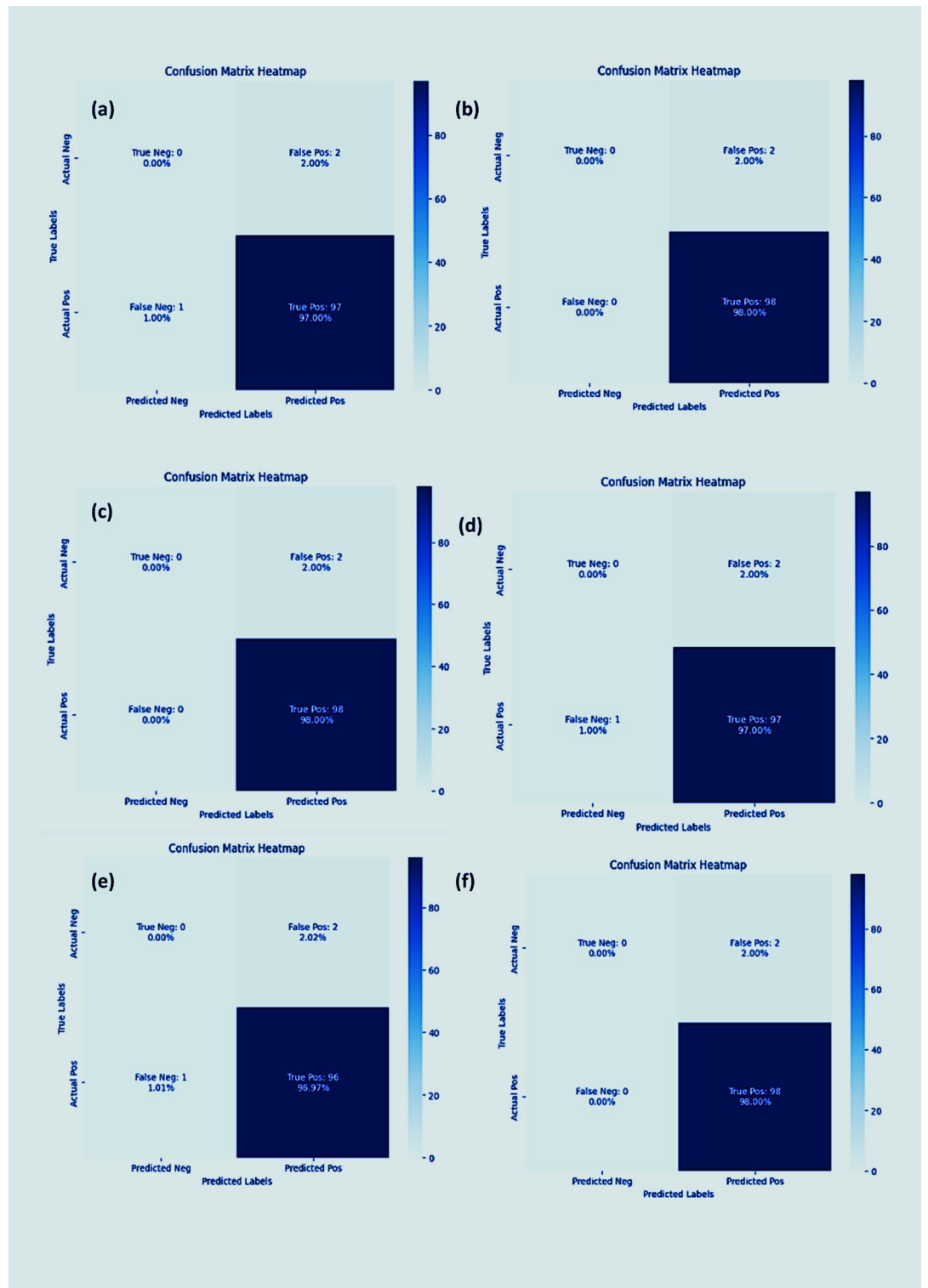
$$E(S) = -\sum_{i=1}^n p_i \log_2(p_i) \quad (4)$$

where  $E$  represents entropy that measures uncertainty in the dataset,  $S$  represents a dataset,  $n$  represents the total number of classes,  $i$  represents certain classes,  $p$  represents probability.

The SGD model got an average accuracy of 96% and precision of 98% as illustrated in Table 1 and Fig. 2E by using the following formula as mentioned in<sup>24</sup>:

Model	Accuracy	Precision	Recall	F1-score
Naïve Bayes	0.97	0.98	0.99	0.98
Support vector machine	0.98	0.98	1.00	0.99
Logistic regression	0.98	0.98	1.00	0.99
Decision tree	0.97	0.98	0.99	0.98
Stochastic Gradient Descent	0.96	0.98	0.98	0.98
K-Nearest Neighbours	0.98	0.98	1.00	0.99

**Table 1.** Comparisons of machine learning models.



**Fig. 2.** Heatmap performance visualization of (a) Naive Bayes Model Performance, (b) Support Vector Machine Model, (c) Logistic Regression Model, (d) Decision Tree Model, (e) Stochastic Gradient Descent Model, (f) K-Nearest Neighbours Model.

$$\omega_{t+1} = \omega_t - \alpha_t \nabla_{\omega_t} l(\omega_t; z_{i_t}) \quad (5)$$

where  $\omega_{t+1}$  represents the updated weight where  $\omega_t$  represents current weight,  $t$  represents the current iteration,  $i$  represents the index of the training example,  $z$  represents the vector of features,  $\alpha$  is the learning rate,  $\nabla$  represents a gradient of the loss function.

Model	Accuracy (%)	References
Simple stress detection algorithm	74.6	7
Decision tree	81	30
Naïve Bayes	90	31
SVM	92.8	32
Adaptive neuro-fuzzy inference system aided fire works grey wolf optimization (ANFIS-FWGO)	94	33
SVM	98	Proposed work

**Table 2.** Comparison table for previous published papers.

The K-Nearest Neighbours model has predicted the stress levels with an average accuracy of 98% and precision of 98%, as illustrated in Table 1 and Fig. 2F, by using the following Euclidean Distance formula as mentioned in<sup>25</sup>:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (6)$$

where  $d$  represents distance,  $i, j$  represents data points,  $k$  represents a feature,  $n$  represents the total number of features,  $x$  represents feature values.

The machine learning model performance is evaluated by using the following metrics, as mentioned in<sup>26,27</sup>:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (10)$$

where  $TP$  represents the true positive rate,  $TN$  represents the true negative rate,  $FP$  represents the false positive rate,  $FN$  represents the false negative rate.

Accuracy is the measure of the ratio of exact predicted values to the total values in the dataset, which evaluates how well the model has performed. Precision is the measure of accuracy of positive predicted values, the ratio of true positive predicted values to all the positive predicted values made by the model. Recall evaluates the ability of the model to observe all relevant samples and the ratio of true positive predicted values to all the actual positive samples. F1-score is the measure of a reciprocal average of precision and recall, which gives a single value to balance precision and recall, specifically used in imbalanced datasets.

The performance of various machine learning models in predicting stress levels is illustrated in Fig. 2. The Naive Bayes model achieved an accuracy of 97%, as shown in the heatmap in Fig. 2a. The Support Vector Machine model demonstrated a slightly higher accuracy of 98% (Fig. 2b). Similarly, the Logistic Regression model also attained an accuracy of 98% (Fig. 2c). The Decision Tree model achieved an accuracy of 97%, as depicted in Fig. 2d. The performance of the Stochastic Gradient Descent model is presented in Fig. 2e, with an accuracy of 96%. Lastly, the K-Nearest Neighbours model reached an accuracy of 98%, as illustrated in Fig. 2f.

The machine learning model accuracy varies across multiple runs due to randomness in training, variation in hyperparameters, model complexity, feature selection, and evaluation metrics<sup>28,29</sup>. The machine learning model like SGD involves randomness in the training of the dataset and it varies for different machine learning models. Each machine learning model has its own set of hyperparameters, variation in hyperparameters can significantly affect model performance. The accuracy of a machine learning model depends on the complexity of the model and it varies for different machine learning models. The model performance is affected by selected features because some machine learning models have the benefit of specific features that other models may not, inconsistent feature selection leads to accuracy varying. Choosing the right evaluation metrics affects model performance<sup>29</sup>. Some machine learning models like Naïve Bayes and Decision Tree got the same accuracy due to the model being trained with simple data of 499 samples and taken models having simple complexity. The difference in accuracy with different models of previously published papers can be observed as illustrated in Table 2.

The proposed smart garment has achieved the highest accuracy of 98% and precision of 98% using the SVM model. The model indicates its effectiveness in classifying stress levels by focusing on heart rate data. SVM performs well by finding an optimal hyperplane that separates various classes, by maximizing the margin among those classes. Data points are referred to as support vectors close to the hyperplane, a kernel trick that allows SVM to manage non-linear relationships by transforming input space into a higher-dimensional space. The regularization parameter stabilizes the change between maximizing margin and minimizing classification errors. The high accuracy indicates that SVM has learnt patterns well from the data, making it a reliable application for predicting stress levels.

## Conclusion

In this paper, a holistic system to calculate stress levels using heart rate variability parameters is presented using machine learning techniques. Used Naive Bayes, SVM, Logistic Regression, Decision Tree, SGD, and KNNs to classify people in high, medium, and low-stress levels. Got a better accuracy of 98% for the SVM model. Uploaded a heart rate data set to the web application and used these models for real-time prediction. These results proved that ensemble methods like Voting Classifiers are used for improving classification accuracy by combining the strengths of multiple machine learning models. Our approach not only provides users with information on stress management but also adds to the body of health informatics research by providing a platform for end-users to semi-continuously track their physiological responses in an easy-to-use fashion. Future work of this paper will focus on dealing with large datasets, incorporating features, and exploring advanced deep-learning techniques to improve accuracy, mass production, and adoption.

## Data availability

The datasets used and/or analysed during the current study are available at <https://www.kaggle.com/datasets/qi-riro/swell-heart-rate-variability-hrv/data>.

Received: 21 January 2025; Accepted: 8 May 2025

Published online: 16 May 2025

## References

- Sethuraman, S. C., Kompally, P., Mohanty, S. P. & Choppali, U. MyWear: A novel smart garment for automatic continuous vital monitoring. *IEEE Trans. Consum. Electron.* **67**(3), 214–222 (2021).
- Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017).
- Tiwari, R., Kumar, R., Malik, S., Raj, T. & Kumar, P. Analysis of heart rate variability and implication of different factors on heart rate variability. *Curr. Cardiol. Rev.* <https://doi.org/10.2174/1573403X16999201231203854> (2021).
- Haque, Y. et al. State-of-the-art of stress prediction from heart rate variability using artificial intelligence. *Cogn. Comput.* **16**(2), 455–481 (2024).
- Schneiderman, N., Ironson, G. & Siegel, S. D. Stress and health: Psychological, behavioral, and biological determinants. *Annu. Rev. Clin. Psychol.* **1**(1), 607–628 (2005).
- Schönfeld, P., Brailovskaia, J., Bieda, A., Zhang, X. C. & Margraf, J. The effects of daily stress on positive and negative mental health: Mediation through self-efficacy. *Int. J. Clin. Health Psychol.* **16**(1), 1–10 (2016).
- Salai, M., Vassányi, I. & Kósa, I. Stress detection using low cost heart rate sensors. *J. Healthc. Eng.* **2016**(1), 5136705 (2016).
- Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P. & Sahebkar, A. The impact of stress on body function: A review. *EXCLI J.* **16**, 1057 (2017).
- Bhargava, D. & Trivedi, H. A study of causes of stress and stress management among youth. *IRA-Int. J. Manag. Soc. Sci.* **11**(03), 108–117 (2018).
- Zafar, M. S. et al. Impact of stress on human body: A review. *Eur. J. Med. Health Sci.* **3**(3), 1–7. <https://doi.org/10.24018/ejmed.2021.3.3.821> (2021).
- Mortensen, J. A., Mollov, M. E., Chatterjee, A., Ghose, D. & Li, F. Y. Multi-class stress detection through heart rate variability: A deep neural network based study. *IEEE Access* **11**, 57470–57480. <https://doi.org/10.1109/ACCESS.2023.3274478> (2023).
- Prakash, K., James, A., Valeti, N. J. & Singha, M. K. Optimization and numerical studies with machine learning assisted graphene-based CuSbS<sub>2</sub> thin film solar cell for flexible electronics applications. *J. Phys. Chem. Solids* **199**, 112513. <https://doi.org/10.1016/j.jpcs.2024.112513> (2025).
- Hamadneh, T. et al. Sales training based optimization: A new human-inspired metaheuristic approach for supply chain management. *Int. J. Intell. Eng. Syst.* **17**(6), 1325–1334. <https://doi.org/10.22266/ijies2024.1231.96> (2024).
- Hamadneh, T. et al. Orangutan optimization algorithm: An innovative bio-inspired metaheuristic approach for solving engineering optimization problems. *Int. J. Intell. Eng. Syst.* **18**(1), 45–58 (2025).
- Bansal, S. et al. Optoelectronic performance prediction of HgCdTe homojunction photodetector in long wave infrared spectral region using traditional simulations and machine learning models. *Sci. Rep.* **14**, 28230 (2024).
- Chen, H., Hu, S., Hua, R. & Zhao, X. Improved naive Bayes classification algorithm for traffic risk management. *EURASIP J. Adv. Signal Process.* **2021**(1), 30 (2021).
- Evgeniou, T., & Pontil, M. Support vector machines: Theory and applications. In *Advanced Course on Artificial Intelligence* 249–257. (Springer Berlin Heidelberg, 1999).
- Jain, P. et al. Machine learning assisted hepta band THz metamaterial absorber for biomedical applications. *Sci. Rep.* **13**(1), 1792. <https://doi.org/10.1038/s41598-023-29024-x> (2023).
- Jain, P. et al. Machine learning techniques for predicting metamaterial microwave absorption performance: A comparison. *IEEE Access* **11**, 128774–128783 (2023).
- Kumar, P. R. et al. Dielectric characterization and machine learning-based predictions in polymer composites with mixed nanoparticles. *J. Macromol. Sci. Part B* <https://doi.org/10.1080/00222348.2024.2372522> (2024).
- Bansal, S. et al. Pt/ZnO and Pt/few-layer graphene/ZnO Schottky devices with Al ohmic contacts using Atlas simulation and machine learning. *J. Sci. Adv. Mater. Dev.* **9**(4), 100798. <https://doi.org/10.1016/j.jsamd.2024.100798> (2024).
- Cramer, J. S. The origins of logistic regression. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.360300> (2003).
- Mienye, Domor, I. & Jere, N. A survey of decision trees: Concepts, algorithms, and applications. *IEEE Access* (2024).
- Achlioptas, P. Stochastic gradient descent in theory and practice. (2019).
- Pulungan, A., Zarlis, M. & Suwilo, S. Performance analysis of distance measures in K-nearest neighbor. In *Proceedings of the 1st International Conference on Management, Business, Applied Science, Engineering and Sustainability Development, ICMASES 2019, 9–10 February 2019, Malang, Indonesia.* (2020).
- Rainio, O., Teuho, J. & Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **14**(1), 6086 (2024).
- Jonnala, N. S. et al. DSIA U-Net: Deep shallow interaction with attention mechanism UNET for remote sensing satellite images. *Sci. Rep.* <https://doi.org/10.1038/s41598-024-84134-4> (2025).
- Sarker, I. H. Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* <https://doi.org/10.1007/s42979-021-00592-x> (2021).
- Gupta, S., Saluja, K., Goyal, A., Vajpayee, A. & Tiwari, V. Comparing the performance of machine learning algorithms using estimated accuracy. *Meas. Sens.* **24**, 100432 (2022).
- Muñoz, S., Iglesias, C. Á., Mayora, O. & Osmani, V. Prediction of stress levels in the workplace using surrounding stress. *Inf. Process. Manag.* **59**(6), 103064 (2022).

31. Arya, S., Anju, A. & Ramli, N. A. Predicting the stress level of students using supervised machine learning and artificial neural network (ANN). *Indian J. Eng.* **21**(56), 1–24. <https://doi.org/10.54905/disssi.v21i55.e9ije1684> (2024).
32. Alruily, M. Sentiment analysis for predicting stress among workers and classification utilizing CNN: Unveiling the mechanism. *Alex. Eng. J.* **81**, 360–370 (2023).
33. Pankajavalli, P. B., Karthick, G. S. & Sakthivel, R. An efficient machine learning framework for stress prediction via sensor integrated keyboard data. *IEEE Access* **9**, 95023–95035 (2021).

## Acknowledgements

This research acknowledged by the Fundamental Research Grant Scheme (FRGS), MOE, Malaysia, Code: FRGS/1/2022/TK07/UKM/02/22. Moreover, this research acknowledges Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R10), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

## Author contributions

K Prakash, M N Harshitha, G N Lakshmi, made substantial contributions to development and numerically analysis through machine learning. P Moses, M S Chowdary, S Bansal participated in the conception, application and critical revision of the article for important intellectual content. M R I Faruque and K S A Mugren provided necessary instructions for analytical expression, data analysis for practical uses and critical revision of the article purposes.

## Declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to K.P., S.B. or M.R.I.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025