ADVANCED FEATURE SELECTION FOR HUMAN PHYSIOLOGICAL STATE PREDICTION USING ERFE

Thangapriya[1], Nancy Jasmine Goldena[2]

[1] Research Scholar, Reg No: 20211242282010, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India
[2] Associate professor, Department of Computer Applications and Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

Corresponding Author
Thangapriya,
priyasaravananrex@gmail.com

DOI
10.29121/shodhkosh.v5.i3.2024.2755

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Human Activity Recognition (HAR) is becoming increasingly important in healthcare as the volume of sensor data grows. Medical practitioners often struggle to quickly and accurately interpret this data to recognize physiological states. Machine learning and feature selection methods can help address this challenge by pinpointing essential features, thereby reducing processing time and enhancing accuracy. This paper introduces an Enhanced Recursive Feature Elimination (ERFE) method for refining feature selection in HAR prediction. Experimental results demonstrate that the ERFE method achieves an 88% classification accuracy, surpassing traditional approaches like LASSO, Random Forest (RF), and standard Recursive Feature Elimination (RFE).

Keywords: Classification, ERFE, Feature Selection, Human Activity Recognition, LASSO, Random Forest, RFE, Performance Evaluation

1. INTRODUCTION

Health is central to quality of life. Advances in healthcare research often yield new treatments, devices, and protocols. Human Action Recognition (HAR) contributes significantly to understanding physiological states and has broad applications in healthcare, wellness, and public safety. Traditional machine learning methods predict common behaviors like walking or running; however, this study focuses on recognizing nuanced states such as emotional, mental, and physical conditions. Feature Selection (FS) is essential in high-dimensional data sets, especially in HAR, where eliminating redundant features can simplify models and enhance efficiency. This paper presents the Enhanced Recursive Feature Elimination (ERFE) method, which identifies optimal HAR features with superior accuracy.

2. RELATED WORKS

HAR has broad implications for human-computer interaction (HCI), elderly care, and healthcare monitoring. HAR methods generally use sensor-based, device-free, or device-worn systems, each suited to specific applications. For instance, smartphone sensors and wearables have shown promise in transportation safety and daily behavior recognition. **Fang et al.** utilized a k-NN classifier with RFE for action identification, demonstrating practical applications for public safety. **Hsu et al.** developed an inertial sensor-based HAR model with Non-parametric Weighted Feature Extraction (NWFE), achieving accurate classification of everyday behaviors. **Chen et al.** proposed BPSO for feature selection, validated with UCI datasets to enhance healthcare classification. These studies underscore the importance of effective FS in HAR, motivating the need for an improved RFE method.

3. EXISTING FEATURE SELECTION TECHNIQUES

HAR data is inherently high-dimensional, necessitating effective feature selection (FS) to isolate only the most relevant features for accurate prediction. Common FS techniques in this domain include:

- **LASSO (LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR):** Reduces regression coefficients toward zero, thus effectively eliminating low-contribution features and helping to mitigate overfitting.
- **RANDOM FOREST (RF):** A wrapper-based FS technique that leverages multiple decision trees to prioritize high-impact features, offering robustness and interpretability in feature selection.
- **RECURSIVE FEATURE ELIMINATION (RFE):** Uses backward selection to iteratively identify optimal feature subsets. While effective, standard RFE can be computationally intensive, which limits its scalability. The Enhanced Recursive Feature Elimination (ERFE) method introduced in this paper addresses this issue, offering a more efficient approach tailored for HAR data.

ERFE reduces computational overhead while maintaining or improving predictive accuracy, thus making it a promising alternative in high-dimensional HAR applications. data is high-dimensional, making it essential to identify only the most relevant features. Common FS techniques include:

4. OUTLINE OF THE WORK

The HAR dataset from the UCI repository, containing 532 features, is preprocessed and fed into the Enhanced Recursive Feature Elimination (ERFE) model. ERFE selects the most relevant features, reducing dimensionality while retaining critical information. These selected features are then classified using supervised machine learning methods, leading to improved prediction accuracy and efficiency in recognizing physiological states.The general workflow is shown in Figure 1.

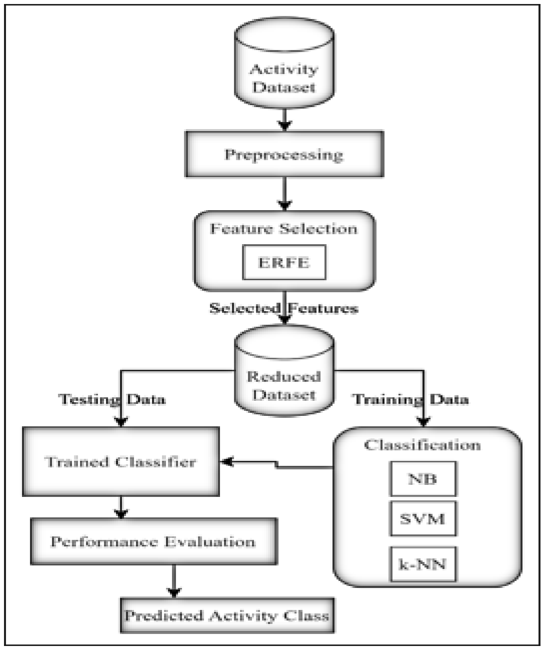

Figure 1: Outline of the Work

5. PROPOSED METHODOLOGY

5.1 ENHANCED RFE (ERFE)

ERFE improves on RFE by refining feature selection and computational efficiency. ERFE uses a brute-force selection process followed by a convergence bagging method, significantly reducing computational time shown in Figure 2.

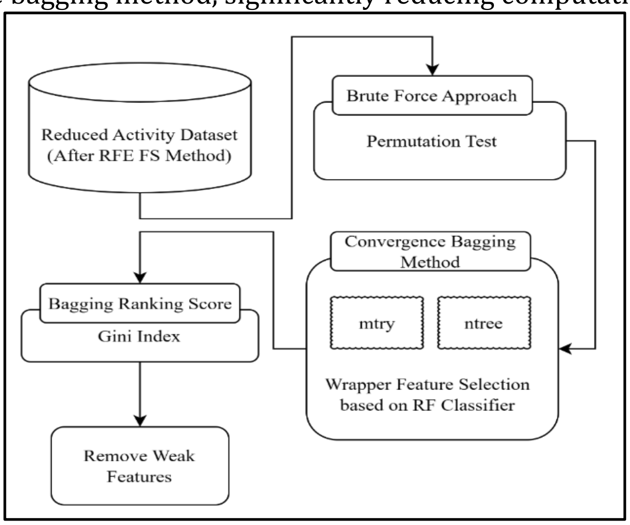

Figure 2: ERFE Method

5.2 ERFE IMPLEMENTATION STEPS:

1. **LOAD DATASET:** Preprocess the data to eliminate missing and redundant information.
2. **PERMUTATION AND MERGING:** Reorganize data using brute-force and shuffle techniques.
3. **RF CLASSIFIER FOR FEATURE IMPORTANCE:** Using decision trees, RF estimates missing values and selects essential features.
4. **BAGGING SCORE CALCULATION:** Computes the absolute bagging score to rank features.
5. **GINI INDEX CALCULATION:** Ranks features by importance, removing weak features.
6. **OPTIMAL FEATURE SELECTION:** Identifies the best features for HAR classification using RF parameters (ntree, mtry, nodesize).

6. CLASSIFICATION METHODS

In this study, after applying the ERFE method to select the most relevant features from the high-dimensional HAR dataset, three supervised classifiers—Support Vector Machine (SVM), Naive Bayes (NB), and k-Nearest Neighbor (k-NN)—are used to categorize the data based on these selected features. The data is split into two sets: 70% is used for training, where the classifiers learn to recognize patterns in the ERFE-selected features, and 30% is reserved for testing, where the classifiers' predictive accuracy and performance are measured. By using the same ERFE-selected features across all three classifiers, this setup allows for a direct comparison of each classifier's effectiveness in terms of accuracy, processing time, and suitability for HAR. This approach helps determine which classifier works best with the ERFE-enhanced features, contributing to more efficient and accurate human physiological state recognition.

7. PERFORMANCE EVALUATION RESULTS

The ERFE approach accelerates model training and enhances accuracy. It selects the most effective subset of features, outperforming other methods. As shown in Figures 3-6, performance evaluation metrics such as accuracy and kappa indicate that ERFE combined with k-NN yields optimal results, with an overall accuracy of 88%.

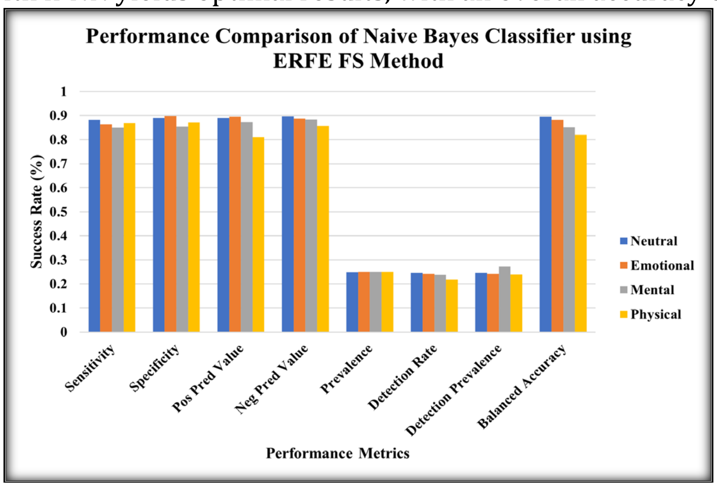

Figure 3: Performance Evaluation Result for Naive Bayes Classifier using ERFE Feature Selection Method

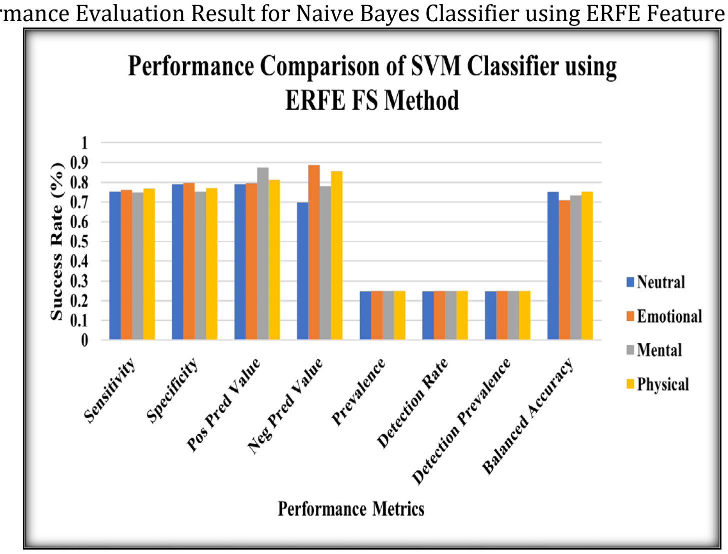

Figure 4: Performance Evaluation Result for Support Vector Machine Classifier using ERFE Feature Selection Method

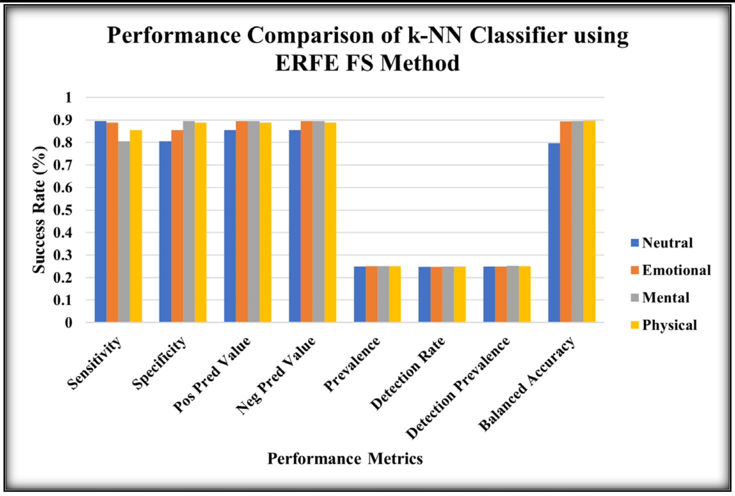

Figure 5 Performance Evaluation Result for k-Nearest Neighbour Classifier using ERFE Feature Selection Method

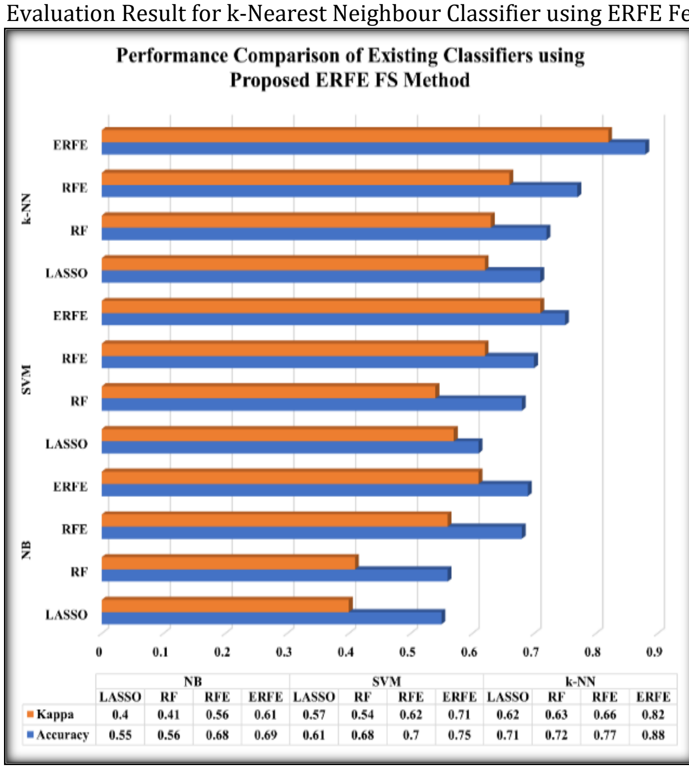

Figure 6: Overall Performance Evaluation Result of Existing Classifiers using Proposed ERFE Feature Selection Method

8. CONCLUSION

In the conclusion, the study highlights the advantages of the Enhanced Recursive Feature Elimination (ERFE) method in improving feature selection for Human Activity Recognition (HAR) tasks. By focusing on selecting only the most relevant features, ERFE allows the machine learning models to train faster and with higher accuracy, addressing the challenges of high-dimensional HAR data. This selective approach also reduces model complexity, making it easier to interpret and reducing the likelihood of overfitting, where a model performs well on training data but poorly on new data. The results show that ERFE consistently performs better than established methods like LASSO, Random Forest (RF), and standard Recursive Feature Elimination (RFE). This consistent improvement suggests ERFE's potential as a more effective approach for HAR prediction. For future research, the study suggests exploring ways to further optimize ERFE by integrating it with real-time data processing. This enhancement could make ERFE even more suitable for applications that require immediate or ongoing data analysis, such as continuous monitoring in healthcare and real-time surveillance.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Tamil Nadu State Council for Science and Technology (TNSCST) for their support of the Human Activity Recognition (HAR) project under the Research Fund for Research Scholars (RFRS) Scheme.

REFERENCES

Fan, Changjun, and Fei Gao. "Enhanced Human Activity Recognition Using Wearable Sensors via a Hybrid Feature Selection Method." Sensors, vol. 21, no. 19, MDPI AG, Sept. 2021, p. 6434. Crossref, https://dx.doi.org/10.29121/shodhkosh.v5.i3.2024.2755

Li, Frédéric, et al. "Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors." Sensors, vol. 18, no. 3, MDPI AG, Feb. 2018, p. 679. Crossref, https://doi.org/10.3390/s18020679.

Cilia, Nicole Dalia, et al. "Comparing Filter and Wrapper Approaches for Feature Selection in Handwritten Character Recognition." Pattern Recognition Letters, vol. 168, Elsevier BV, Apr. 2023, pp. 39–46. Crossref, https://doi.org/10.1016/j.patrec.2023.02.028.

A. Badawi, A. Al-Kabbany and H. Shaban, "Daily Activity Recognition using Wearable Sensors via Machine Learning and Feature Selection," 2018 13th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 2018, pp. 75-79, doi: 10.1109/ICCES.2018.8639309.

Wang, Aiguo, et al. "Evaluation of Random Forest for Complex Human Activity Recognition Using Wearable Sensors." 2020 International Conference on Networking and Network Applications (NaNA), IEEE, Dec. 2020. Crossref, https://doi.org/10.1109/nana51271.2020.00060.

Badawi, Abeer A., et al. "Sensor Type, Axis, and Position-Based Fusion and Feature Selection for Multimodal Human Daily Activity Recognition in Wearable Body Sensor Networks." Journal of Healthcare Engineering, vol. 2020, Hindawi Limited, June 2020, pp. 1–14. Crossref, https://doi.org/10.1155/2020/7914649.

Ahmed, Nadeem, et al. "Enhanced Human Activity Recognition Based on Smartphone Sensor Data Using Hybrid Feature Selection Model." Sensors, vol. 20, no. 1, MDPI AG, Jan. 2020, p. 317. Crossref, https://doi.org/10.3390/s20010317.

Mohino-Herranz, Inma, et al. "Activity Recognition Using Wearable Physiological Measurements: Selection of Features From a Comprehensive Literature Study." Sensors, vol. 19, no. 24, MDPI AG, Dec. 2019, p. 5524. Crossref, https://doi.org/10.3390/s19245524.

Badshah, Mustafa. Sensor - Based Human Activity Recognition Using Smartphones. San Jose State University Library. Crossref, https://doi.org/10.31979/etd.8fjc-drpn.

Othman, N.A., Aydin, I. (2021). Challenges and limitations in human action recognition on unmanned aerial vehicles: A comprehensive survey. Traitement du Signal, Vol. 38, No. 5, pp. 1403-1411. https://doi.org/10.18280/ts.380515

Alzahrani, Mona Saleh and Salma Kammoun. "Human Activity Recognition: Challenges and Process Stages." International Journal of Innovative Research in Computer and Communication Engineering 2016 (2016): n. pag.

Sunny, Jubil T et al. "Applications and Challenges of Human Activity Recognition using Sensors in a Smart Environment." (2015).

Available from: https://www.linkedin.com/pulse/what-recursive-feature-elimination-amit-mittal

Available from: https://bookdown.org/max/FES/recursive-feature-elimination.html

Available from: https://machinelearningmastery.com/rfe-feature-selection-in-python

Available from: https://topepo.github.io/caret/recursive-feature-elimination.html

Available from: https://towardsdatascience.com/feature-selection-in-machine-learning-using-lasso-regression

Available from: https://medium.com/@23.sargam/lasso-regression-for-feature-selection-8ac2287e25fa

Available from: https://corporatefinanceinstitute.com/resources/knowledge/other/lasso

Available from: https://chrisalbon.com/code/machine_learning/trees_and_forests/feature_selection_using_random_forest

Available from: https://blog.datadive.net/selecting-good-features-part-iii-random-forests

L. Fang, S. Yishui and C. Wei, "Up and down buses activity recognition using smartphone accelerometer," 2016 IEEE Information Technology, Networking, Electronic and Automation Contrl Conference, Chongqing, China, 2016, pp. 761-765, doi: 10.1109/ITNEC.2016.7560464.

Y. -L. Hsu, S. -L. Lin, P. -H. Chou, H. -C. Lai, H. -C. Chang and S. -C. Yang, "Application of nonparametric weighted feature extraction for an inertial-signal-based human activity recognition system," 2017 International Conference on Applied System Innovation (ICASI), Sapporo, Japan, 2017, pp. 1718-1720, doi: 10.1109/ICASI.2017.7988270.

Y. Chen, Y. Wang, L. Cao and Q. Jin, "CCFS: A Confidence-Based Cost-Effective Feature Selection Scheme for Healthcare Data Classification," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 3, pp. 902-911, 1 May-June 2021, doi: 10.1109/TCBB.2019.2903804.

Activity recognition using wearable physiological measurements. (2019). UCI Machine Learning Repository. https://doi.org/10.24432/C5RK6V