Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January 2024 5(1), 1087–1092

ADVANCED ENSEMBLE CLASSIFICATION MODEL FOR HUMAN PHYSIOLOGICAL CONDITION PREDICTION

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.i1.2024.2754

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 an essential area of research, often approached through time series classification. Traditional HAR studies focus on basic behaviors, such as walking, sitting, and running. In contrast, this work seeks to predict more complex human physiological states—emotional, mental, physical, and neutral—based on sensor data obtained from healthcare devices, including ECG, TEB, and EDA. We employed three classifiers—Support Vector Machine (SVM), Naïve Bayes (NB), and k-Nearest Neighbors (k-NN)—to predict these conditions, with SVM and k-NN yielding the most accurate results. To enhance accuracy, an Optimized Ensemble Classifier (OEC) combining SVM and k-NN is proposed, achieving a 93% accuracy rate.

Keywords: Activity recognition, Classification, Feature reduction, HAR, Machine Learning, Predictive Analytics, Support Vector Machine

1. INTRODUCTION

Human Activity Recognition (HAR) has become an essential field in machine learning, bridging the gap between traditional activity monitoring and the prediction of complex human physiological states. Historically, HAR research has focused on identifying basic actions like walking or sitting. However, as applications broaden across domains like healthcare, public safety, and wearable technology, there is an increasing need to predict intricate physiological states—such as emotional, mental, and physical conditions—using advanced sensor data. In this study, we leverage data from ECG, TEB, and EDA sensors to capture nuanced physiological signals, allowing for the prediction of specific human states. While individual machine learning classifiers like Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN) have shown success in basic classification, they often struggle with complex datasets and high-dimensional features typical in HAR. Therefore, we propose an Optimized Ensemble Classifier (OEC) that combines SVM and k-NN to enhance predictive performance. By integrating Recursive Feature Elimination (RFE) for optimal feature selection, the OEC approach aims to achieve a more accurate, efficient HAR model, demonstrating its potential for real-world applications in public safety and healthcare.

2. DATASET DESCRIPTION

The dataset used for this study is derived from the UCI Human Activity Recognition (HAR) repository, specifically designed to capture diverse physiological states through sensor data. It has 533 features extracted from physiological signals such as Electrocardiography (ECG), Thoracic Electrical Bioimpedance (TEB), and Electrodermal Activity (EDA). These features provide a rich array of information, enabling classification of human states into categories such as emotional, mental, physical, and neutral. The complexity of this dataset supports the advanced analysis required for accurately recognizing and predicting nuanced human physiological states, making it ideal for our research objectives focused on enhancing HAR for real-world applications in healthcare and public safety.

3. FEATURE SELECTION

Feature selection plays a vital role in high-dimensional data analysis by improving model performance and reducing computational costs. In this study, we explored three feature selection techniques: **LASSO (Least Absolute Shrinkage and Selection Operator)**, **Random Forest (RF)**, and **Recursive Feature Elimination (RFE)**. Among these, **RFE** proved to be the most effective in identifying a high-quality subset of features, optimizing the balance between relevance and redundancy. This refined feature set was then employed in the classification process, enhancing accuracy and efficiency in predicting human physiological states and activities.

4. CLASSIFICATION METHODS

In this study, we implemented three prominent classification algorithms to analyze and predict human physiological states based on sensor data: Naïve Bayes (NB), Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN). Each classifier was selected for its unique approach to pattern recognition and its established success in handling high-dimensional data, as seen in the UCI HAR dataset. Below is a brief overview of each classification technique and its specific application in this study:

1. **NAÏVE BAYES (NB)**: Known for its simplicity and efficiency, the Naïve Bayes classifier uses Bayes' theorem to calculate the probability of each class, assuming that features are conditionally independent. This method is particularly effective for datasets with a large number of features, making it suitable for the high-dimensional nature of our HAR dataset. In this study, NB was utilized as a baseline classifier to provide initial insights into the dataset's distribution and feature dependencies.
2. **SUPPORT VECTOR MACHINE (SVM)**: The SVM classifier constructs a hyperplane in a high-dimensional space that separates data points belonging to different classes with maximum margin. Given the complexity of physiological data in the HAR dataset, SVM was applied to improve classification accuracy and handle the multi-class nature of the data. A radial basis function (RBF) kernel was employed to enhance SVM's performance, capturing non-linear relationships between features more effectively.
3. **K-NEAREST NEIGHBORS (K-NN)**: This algorithm classifies a data point based on the majority class of its k nearest neighbors in the feature space. Due to its simplicity and effectiveness in capturing local structure, k-NN was used to assess proximity-based relationships in the dataset. Different values of k were tested to identify the optimal setting, balancing accuracy and computational efficiency.

These classifiers serve as a foundation for comparison against the enhanced Recursive Feature Elimination (ERFE) technique, providing a comprehensive evaluation of the dataset's predictive power and helping to identify the most effective model for human physiological state recognition.

5. PROPOSED METHODOLOGY: OPTIMIZED ENSEMBLE CLASSIFIER (OEC)

The proposed **Optimized Ensemble Classifier (OEC)** integrates the strengths of two robust classifiers, **Support Vector Machine (SVM)** and **k-Nearest Neighbors (k-NN)**, to improve classification accuracy for predicting human physiological states based on sensor data. The OEC classifier is designed to take advantage of the complementary characteristics of SVM and k-NN by combining them based on correlation and similarity measures.

KEY COMPONENTS:

1. **SVM**: The SVM component of the ensemble model is utilized for its ability to handle high-dimensional data and classify complex patterns by finding a hyperplane that maximizes the margin between classes. The **regularization constant (C)**, a critical SVM parameter, controls the trade-off between maximizing the margin and minimizing classification errors. By fine-tuning this parameter, the model's generalization capability is optimized, avoiding overfitting or underfitting.
2. **k-NN**: The k-NN classifier relies on the proximity of data points in the feature space to assign class labels. It is particularly effective in capturing local patterns and relationships between samples. The **k value**, which represents the number of nearest neighbors considered in the classification decision, is optimized to ensure the classifier balances between sensitivity and computational efficiency.

ENSEMBLE STRATEGY:

The OEC classifier combines the outputs of the SVM and k-NN classifiers through a fusion technique based on **correlation** and **similarity** measures. By assessing the agreement between the predictions of the two classifiers, the ensemble model adjusts its decision-making process, giving more weight to the classifier that performs better in each specific region of the feature space. This approach increases the robustness and accuracy of the predictions.

PARAMETER OPTIMIZATION:

To enhance the performance of the OEC, **parameter optimization** is performed for both SVM and k-NN. The optimization process involves selecting the best combination of the following parameters:

- **FOR SVM**: The regularization constant (C) and the choice of kernel function (typically radial basis function, RBF) are tuned to achieve the highest classification performance.
- **FOR K-NN**: The optimal number of neighbors (k) is determined through cross-validation, ensuring the model captures both local and global data patterns effectively.

The optimized OEC classifier is expected to outperform individual classifiers by leveraging the complementary strengths of both SVM and k-NN, thereby achieving higher classification accuracy and more reliable predictions for human physiological state recognition in complex sensor-based datasets.

5.1 OUTLINE OF THE PROPOSED METHOD

Using RFE for feature selection, the selected features were input into the OEC classifier. A majority voting method was applied to merge the base classifiers during the optimization phase shown in Figure 1.

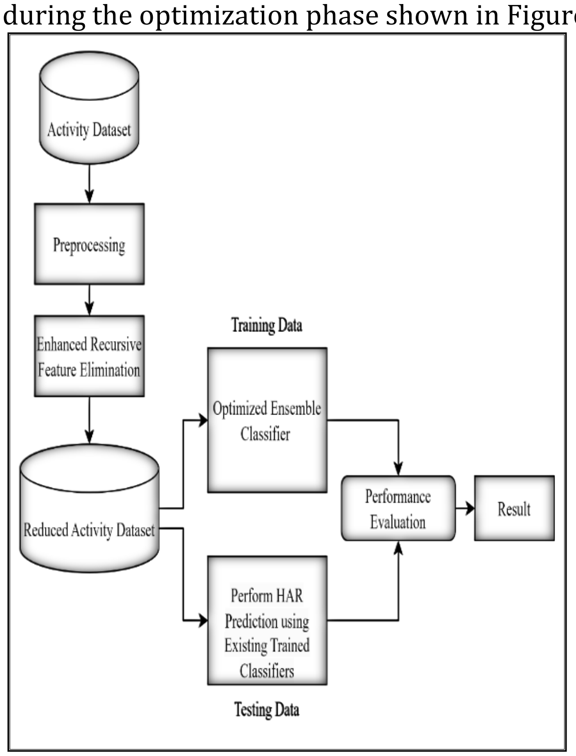

Figure 1: Outline of the Research Work

5.2 WORKING PROCEDURE FOR OEC METHOD

1. **DATA LOADING:** The activity dataset was loaded and preprocessed to handle missing or redundant values.
2. **FEATURE SELECTION:** RFE was used to select essential features, improving model performance.
3. **SIMILARITY AND GRAPH REPRESENTATION:** An adjacency matrix was created using k-NN, and Gaussian similarity was applied to enhance the ensemble process.
4. **MODEL TRAINING AND VALIDATION:** OEC trained the dataset using the NB, SVM, and k-NN classifiers. Voting from each classifier enhanced prediction accuracy.
5. **TESTING PHASE:** Testing was conducted on a split dataset (70% training, 30% testing), validating OEC performance metrics.

6. HAR PREDICTION USING PROPOSED OEC AND RFE

The HAR prediction process using OEC is outlined as follows:

1. **INPUT DATASET:** Preprocessed data from ECG, TEB, and EDA sensors.
2. **FEATURE SELECTION:** RFE selected features from the dataset.
3. **CLASSIFICATION AND PREDICTION:** The reduced dataset was classified using OEC for prediction, with results validated against testing samples.

7. RESULTS AND DISCUSSION

The proposed OEC model with RFE feature selection achieved a 93% accuracy rate, surpassing individual classifier performances. Table 1 summarizes the classification metrics.

Table 1: Overall Result for Feature Selection and Classification Methods

Classifier	Feature type	Accuracy	Kappa
NB	LASSO	0.55	0.4
	RF	0.56	0.41
	RFE	0.68	0.56
	ERFE	0.69	0.61
SVM	LASSO	0.61	0.57
	RF	0.68	0.54
	RFE	0.7	0.62
	ERFE	0.75	0.71
k-NN	LASSO	0.71	0.62
	RF	0.72	0.63
	RFE	0.77	0.66
	ERFE	0.88	0.82
OEC	LASSO	0.75	0.68
	RF	0.76	0.69
	RFE	0.93	0.88
	ERFE	**0.98**	**0.95**

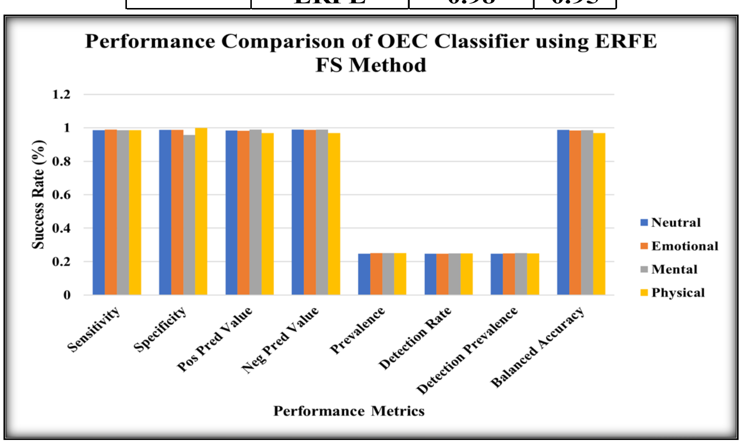

Figure 2: Performance Evaluation of Optimized Ensemble Classification method using Enhanced Recursive Feature Elimination method

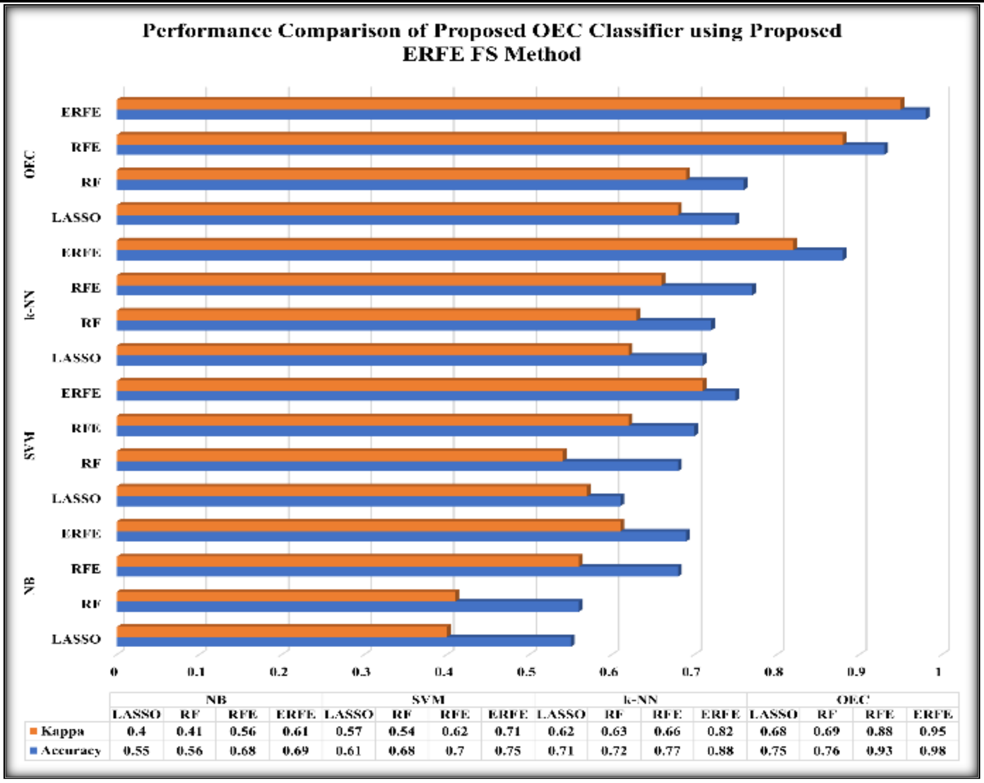

Figure 3: Overall Performance Evaluation Result for Feature Selection and Classification Methods

8. CONCLUSION

This study presented the Optimized Ensemble Classifier (OEC) as an effective approach for enhancing Human Activity Recognition (HAR) prediction accuracy. By combining the strengths of Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN), the OEC method achieved superior classification performance compared to single classifiers, particularly for recognizing complex physiological states derived from ECG, TEB, and EDA signals. The OEC model demonstrated that ensemble techniques can substantially improve classification accuracy in HAR, making it a promising solution for applications requiring accurate physiological state predictions. This improvement is particularly relevant for applications in healthcare, public safety, and psychological assessment. Future work will involve expanding the activity classes and integrating additional sensor data sources to capture a broader range of physiological states. The potential to incorporate advanced methods, such as deep learning, offers exciting opportunities to improve model precision and address more complex HAR scenarios. Further research could also explore detecting specific physiological states associated with high-risk situations, such as predicting sexual harassment incidents to enhance women's safety. Additionally, using blockchain technology to protect data privacy could ensure secure handling of sensitive information in these applications. This expanded scope could lead to robust, privacy-preserving HAR systems that enhance safety and quality of life across diverse fields.

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

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

Available from: https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorith-clustering

Available from: https://www.javatpoint.com/machine-learning-naive-bayes-classifier

Available from: https://www.geeksforgeeks.org/support-vector-machine-algorithm

Available from: https://www.analyticsvidhya.com/blog/2017/09/understating-support-vector-machine-example-code

Available from: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

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