# Optimizing feature selection in multilayer ensemble models for improved HAR accuracy

**Dhiraj Prasad Jaiswal[1], Ashok Kumar Shrivastava[2]**
[1]Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior, India
[2]Amity School of Engineering and Technology, Faculty of Computer Science and Engineering, Amity University Madhya Pradesh, Gwalior, India

**Article Info**

***Article history:***

Received May 29, 2025
Revised Sep 16, 2025
Accepted Dec 6, 2025

***Keywords:***

Data dimensionality
Feature selection
Human activity recognition
Machine learning
Wearable sensors

**ABSTRACT**

Human activity recognition (HAR), is an increasingly significant research area as it can be used in many fields of application such as; health care, elderly monitoring, sports training, and smart homes. In this research we developed a novel multi-layer ensemble model based on a combination of a genetic algorithm (GA) to optimize feature selection and hierarchical learning to solve the issues of high dimensional data, feature redundancy and over fitting in HAR. Our model systematically reduces the number of features required to recognize activities while maintaining the most important features; thus, allowing the base learner to learn patterns across multiple layers. We demonstrated through experiments using three standard benchmark datasets-UCI HAR, WISDM, and PAMAP2, that our method significantly outperformed standard methods achieving 96.8% accuracy, and reduced the amount of feature sets by more than 70%. Evaluation metrics including; precision, recall, F1-score, and ROC-AUC, further validated the robustness of our model; while statistical tests confirmed the improvement in performance. Additionally, our framework improved the efficiency and interpretability of our model, which will enable it to be practically implemented in real time environments. These results demonstrate the potential of combining feature selection optimized by a GA and hierarchical ensembles in HAR, and provide avenues for future work in cross domain adaptability and multimodal HAR systems.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

Dhiraj Prasad Jaiswal
Amity School of Engineering and Technology, Amity University Madhya Pradesh
Gwalior, Madhya Pradesh, India
Email: dhiraj0411@gmail.com

## 1. INTRODUCTION

Human in order for healthcare monitoring, smart environments, sports analysis, and security systems to be effective, human activity recognition (HAR) is a critical factor because it allows for an automated identification of human actions by means of their sensing data. Consequently, context-sensitive applications are able to provide better safety, well-being, and efficiency through HAR. As wearables containing inertial measurement unit (IMU) that include accelerometers and gyroscopes, become more widespread. HAR is becoming a major application in healthcare, elderly health monitoring, fitness tracking, fall detection [1], and smart homes. In addition to rehabilitation and remote monitoring of patients [2], wearable sensors have also been used effectively in clinical practice, which further emphasizes the role of HAR in clinical practice. For example, HAR systems have been commonly applied to elderly care [3]. Reliable recognition of day-to-day activities supports elderly individuals to remain independent in their daily lives. Although advancements have been made in developing methods to recognize actions using sensor data, the process still has many difficulties in terms of action recognition using real world data. The reason is the high dimensionality of sensor data, variability in how people act, and inconsistencies resulting from uneven placement of sensors.

Researchers have attempted to improve the performance of HAR systems during the last decade. Gave an overview of wearable-sensor based HAR and stated the necessity of robust feature engineering in HAR [4]. Emphasized two main problems: subject variability and computational complexity [5]. In addition, researchers have utilized deep-learning based approaches, such as convolutional and recurrent networks [6], to develop successful methods to extract features automatically; however, these approaches typically require large labelled databases and are computationally expensive. Analyzed various techniques for selecting features, and indicated that the choice of features is important in removing redundant features and to provide greater interpretability of models [7]. Researchers have also used vision-based HAR methods [8]; however, there are privacy issues when camera-based HAR methods are compared to those that utilize wearable sensors. Developed a hybrid framework for selecting features that utilizes both filter and wrapper methods [9], and developed an ensemble method for selecting features to increase the robustness of the selected features [10]. In addition, [11] developed a survey of HAR methods that utilize inertial-sensors, and indicated that there are limitations in the use of traditional feature extraction methods and that new methods for selecting features are needed. Finally, [12], [13] utilized deep convolutional neural networks (CNNs) to classify data collected from smartphones, and demonstrated that increasing the number of convolutional layers improves the performance of CNNs over traditional machine learning methods such as support vector machines (SVM), while also demonstrating that increasing the depth of a network too much will decrease the complexity of the features being extracted. Recent research has demonstrated [14] that the use of multi-layered ensemble methods can improve the accuracy of sentiment analysis, while also demonstrating that stacking ensembles with integrated feature selection can improve the accuracy of predictions for health-related applications.

There are still many open research questions. Most previous HAR studies employed a single classifier and/or shallow ensembles [15], which do not take into account the hierarchical nature of human activity patterns. In addition to this, many previous HAR studies applied feature selection separately from the requirements for multilayered ensembles. Also, as previously mentioned, deep learning [16] models can be highly accurate but require high computation and are therefore generally not suitable for use in real-time or limited resource applications [17]. Consequently, the need for effective frameworks that combine feature selection and multilayered ensemble architecture optimizations to enable efficient HAR while improving accuracy remains unfulfilled.

Therefore, the goal of this study was to develop a new multilayered ensemble framework combined with an optimized feature selection technique. This framework utilizes genetic algorithms (GAs) to sequentially optimize the feature sets for the global level, layer-levels, and base learner levels; thereby, ensuring each portion of the ensemble is optimized through the utilization of the most relevant and least redundant features, which will enhance the accuracy while reducing the dimensionality. In contrast to other approaches, this method uses hierarchical learning in conjunction with specifically tailored feature optimization methods to enhance the accuracy of the recognition model and to improve the efficiency of the computation.

The primary contributions of this paper are:
- A multi-layered ensemble architecture that can identify both high- and low-level patterns in sensor data better than one layer, conventional single-layered ensembles.
- An optimized feature selection using GAs to find an optimal number of dimensions at each layer of the multi-layered ensemble.
- The use of three benchmark HAR datasets (PAMAP2, UCI HAR, and WISDM) to evaluate the performance of the proposed method to be more accurate and efficient than all other methods, including baseline neural networks, single models, and traditional ensemble architectures.
- Analysis of selected features to provide insight into what sensor derived attributes were contributing most to the identification of activity, and how those sensors derived attributes can influence future sensor designs.

The rest of the paper is organized as follows; in section 2, the proposed method will be described, including feature extraction, the proposed multi-layered ensemble, and the optimized feature selection strategy. The experimental setup and results are presented in section 3. The discussion of the findings, and the implications and limitations of the study are presented in section 4. Finally, section 5 summarizes the main contributions of this paper and provides direction for future research. Further motivation for the current study has been provided by recent studies that emphasize the importance of integrating feature optimization and ensemble learning as a means to develop efficient HAR systems [18]–[21].

## 2. METHOD

The proposed framework for optimizing feature selection in multilayer ensemble models for is composed of three major components:

– Feature extraction and initial selection,
– Multilayer ensemble architecture, and
– GA-based multi-level feature optimization.

The method is designed to enhance recognition accuracy, minimize computational overhead, and ensure reproducibility. Figure 1 presents the overall workflow.

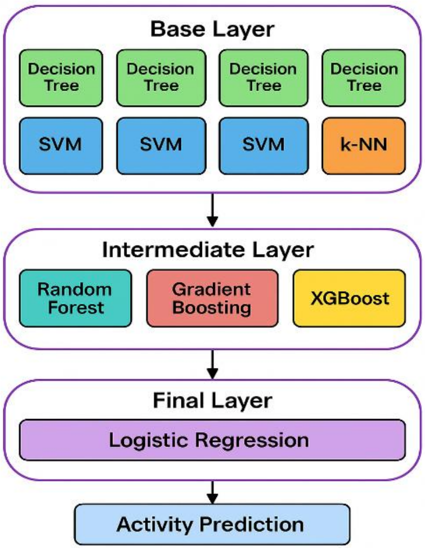


Figure 1. Workflow of the proposed multilayer ensemble framework

### 2.1. Data preprocessing and feature extraction

Raw sensor data from accelerometers and gyroscopes undergo preprocessing to remove noise and normalize values across channels. Each signal $x(t)$ is normalized using z-score normalization:

$$x' = \frac{x-\mu}{\sigma}$$

where $\mu$ is the mean and $\sigma$ is the standard deviation of the signal segment.
To capture temporal dynamics, the data are segmented using a sliding window of size $W$ with 50% overlap, defined as:

$$S_k = \{x_{k\cdot(W/2)}, \dots, x_{k\cdot(W/2)+W-1}\}$$

where $S_k$ is the $k$-th segment.

From each segment, both time-domain (e.g., mean, standard deviation, skewness, kurtosis, zero-crossing rate, and peak-to-peak amplitude) and frequency-domain features (e.g., spectral energy, spectral entropy, dominant frequency, and frequency range) are extracted. Such preprocessing strategies are consistent with early accelerometer-based HAR studies [8]. These features are summarized in Table 1. A correlation-based feature selection (CFS) technique is then applied to eliminate highly correlated features that may introduce redundancy.

To eliminate redundant features, we use the CFS, and reduce features to those which are correlated (pairwise Pearson correlation $|r| \geq 0.95$). The CFS thus reduces redundancy in the feature space before the optimization step. In this way, the refinement of the feature space prior to the optimization step will be reduced in dimensionality, thus, it helps to mitigate the "curse of dimensionality" [22].

Table 1. Initial set of features extracted from sensor data

| Domain | Feature | Description |
|---|---|---|
| Time | Mean | Average value of the signal |
| Time | Standard deviation | Measure of signal variability |
| Time | Skewness | Asymmetry of the signal distribution |
| Time | Kurtosis | Peakedness of the signal distribution |
| Time | Zero crossing rate | Frequency of signal sign changes |
| Time | Peak-to-peak amplitude | Difference between maximum and minimum values |
| Frequency | Spectral energy | Sum of the squared FFT components |
| Frequency | Spectral entropy | Entropy of the normalized FFT components |
| Frequency | Dominant frequency | Frequency with the highest magnitude |
| Frequency | Frequency range | Difference between max and min frequencies |

### 2.2. Multilayer ensemble architecture

In order to utilize the advantages of multiple learners, we propose an architecture for a multilayer ensemble that can represent both the lower level and higher-level representations of sensor data. In addition, recent research on the use of simple deep ensembles have shown them to be effective in HAR [23] and therefore motivate our hierarchical design. Unlike traditional flat ensembles, our multilayer design includes a hierarchical learning process between layers, and each layer improves upon the previous layer's output.

The Figure 1 illustrates an overall view of how the multilayer ensemble framework works. The ensemble is a multi-layered structure with three layers:

– Base layer-ten heterogeneous classifiers are used to train ten different feature subsets. The use of multiple, diverse learners for different subsets of features captures many different characteristics of the activity patterns.
– Summary layer-three meta-learners (random forest (RF), gradient boosting, and XGBoost), take the first layer's output and create a better overall prediction through increased robustness and reliability.
– Output layer-the logistic regression takes the summary layer output and generates the final activity label. This layer can benefit from being able to provide calibrated probabilities as an output.

The multilayer architecture enables a hierarchical learning process of representations, such that each layer refines and interprets the features generated by the previous layer to improve the generalization capabilities of the system.

### 2.3. Multi-level genetic algorithm-based feature selection

Evolutionary feature selection has been widely studied in HAR and related domains [24]. Building on this foundation, the core novelty of our approach lies in optimizing feature subsets at three hierarchical levels:

– Global selection for the entire ensemble.
– Layer-specific selection tailored to intermediate and final layers.
– Base-learner-specific selection based on learner characteristics.

Building upon previous studies on evolutionary feature selection for HAR, the proposed method goes beyond the realm of conventional single-level optimization strategies. Instead of treating the feature space uniformly, the approach introduces a structured optimization process that adapts to different stages of the ensemble architecture. Therein, by aligning feature subsets with functional roles of ensemble layers and base learners in their individual capacity, the method develops a more informed approach to feature utilization that fosters improved representational efficiency and, consequently, potentially better recognition performance.

a. Fitness function

For a candidate subset $F_{l,b}$ (features for base learner $b$ in layer $l$), the fitness function is:

$$\text{Fitness}(F_{l,b}) = \alpha \cdot \text{Acc}(F_{l,b}) - \beta \cdot \frac{|F_{l,b}|}{|F_{\text{total}}|}$$

where: $\text{Acc}(F_{l,b})$ is the cross-validated classification accuracy, $|F_{l,b}|$ is the number of selected features, $|F_{\text{total}}|$ is the total available features, and $\alpha, \beta$ control the trade-off between accuracy and dimensionality (set to 0.9 and 0.1 respectively in our experiments).

b. GA configuration

– Population size: 100.
– Generations: 100.
– Crossover rate: 0.8 (uniform crossover).
– Mutation rate: 0.05 (bit-flip mutation).

– Selection strategy: tournament selection (size=3).
– Elitism: top 2 individuals per generation retained.
– Randomness handling: all experiments repeated with 5 different random seeds; results averaged to mitigate stochastic bias.

c. Base learner feature assignment

Each base learner is initialized with a random subspace of features (50–70% of the feature space), guided by:
– Learner-specific preferences—e.g., decision trees are assigned higher-variance features, SVMs receive features with higher discriminative power (via fisher score).
– Diversity maximization—measured using subset similarity:

$$\mathrm{Sim}(F_i, F_j) = \frac{|F_i \cap F_j|}{|F_i \cup F_j|}$$

ensuring $\mathrm{Sim} < 0.5$ between any two base learners.

By applying this multi-level optimization strategy, the system ensures that each component of the ensemble operates on the most relevant information, leading to superior recognition accuracy and generalizability. An ensemble feature selection strategy similar to [21] was adapted but improved through layer-specific optimization in this study. The pseudocode of the Algorithm 1 is given:

Algorithm 1. Optimized feature selection for multilayer ensemble

```
Input:
    F_total   ← Initial full feature set
    D         ← Training dataset
Output:
    Optimized feature subsets for all learners in all layers

    // Step 1: Global Selection
    F_global ← GA_Wrapper_Select(F_total, Ensemble_Eval)

    // Step 2: Initialize Ensemble
    for each layer l in Ensemble:
        F_layer[l] ← GA_Wrapper_Select(F_global, Layer_Eval(l))

    // Step 3: Base Learner Optimization
        for each base learner b in layer l:
            Pop ← InitializePopulation(F_layer[l], RandomSubspace)
            for gen = 1 to MaxGenerations:
                 Fitness ← Evaluate(Pop, Learner_Eval(b))
                 Parents ← TournamentSelection(Pop)
                 Offspring ← Crossover(Parents, rate=0.8)
                 Mutate(Offspring, rate=0.05)
                 Pop ← Elitism(Pop, Offspring)
             F_opt[l,b] ← BestSubset(Pop)

     return F_opt
```

### 2.4. Justification for method choices

The methodological decisions were informed by the fact that the key aspects of complex and hierarchical patterns found in the data of human activity were to be efficiently represented. The architecture can progressively improve the representations of low-level features to higher-level abstractions by organizing the learning process into multiple layers and optimizing them separately, which with a single, flat learning model is hard to do. Moreover, the evolutionary optimization will allow introducing a principled compromise between the accuracy and the model compactness so that only the most informative features will be kept without being redundant. Lastly, repeated experimental runs with statistical validation enhances the applicability of the results by showing that the recorded improvement in performance is not a one-off result which can be attributed to a particular initialisation or randomised setting.

## 3. EXPERIMENTAL SETUP AND RESULTS

The performance of the method proposed was evaluated by conducting a number of thorough experiments with multiple HAR data sets. This section provides an overview of the data sets that were used for the experimentation, experimental set up, and associated results.

### 3.1. Datasets

In order to assess the merits of the method proposed, three benchmarks' data sets were utilized: UCI HAR [25], WISDM [26], and PAMAP2 [8]. The differences between the three lies within sampling rate, activity type, and modality which will allow the evaluation of robustness. The summary of the data sets is provided in Table 2. There are two types of data sets:

a. UCI HAR data set: the UCI HAR data set contains a sample of 30 participants completing six different physical activities. These were recorded using both the accelerometer and gyroscopes to measure the movement of each participant at 50 Hz for a total of 10,299 trials.
b. WISDM data set: the WISDM data set was collected from 36 subjects who completed 6 activities while wearing an accelerometer to capture their movements. The sampling rate of the data set was 20 Hz, which resulted in 1,098,207 samples. PAMAP2 data set: the PAMAP2 data set contained data from nine subjects who completed eighteen physical activity tasks, while measuring their movement with IMUs at a 100 Hz sampling rate. As a result, the total number of samples was 3,850,505.

Table 2. Summary of HAR datasets used in the experiments

| Dataset | Subjects | Activities | Sensors | Sampling rate (Hz) | Total samples |
|---|---|---|---|---|---|
| UCI HAR | 30 | 6 | Acc, Gyro | 50 | 10,299 |
| WISDM | 36 | 6 | Acc | 20 | 1,098,207 |
| PAMAP2 | 9 | 18 | IMU | 100 | 3,850,505 |

### 3.2. Experimental setup

Python version 3.8 was utilized for the implementation of the experiment. DEAP 1.3.1 was employed to implement the GA, while scikit-learn 0.24.2 was employed to implement machine learning models.

A workstation with an Intel Xeon E5-2680 v4 processor and 128 GB RAM was employed to conduct experiments. Steps that were taken for each dataset:

a. Data preprocessing: standard preprocessing techniques such as removing noise from the data, normalizing the data, and segmenting the data by employing a sliding window method with a 50 percent overlap were applied to the data.
b. Feature extraction: the initial set of features listed in table one for each window of sensor data was extracted.
c. Model configuration: a multilayer ensemble model that included three layers was developed for this research:
– Base layer: 10 base learners (4 decision trees, 3 SVMs, and 3 k-NN).
– Intermediate layer: 3 meta-learners (RF, gradient boosting, and XGBoost).
– Final layer: 1 meta-learner (logistic regression).
d. Feature subset optimization: a GA with a population size of 100 and 100 generations was employed to identify the best subset of features for each layer of the ensemble.
e. Evaluation: the performance of the model was assessed utilizing 5 fold cross-validation, and was compared to multiple baseline techniques, which include single models, traditional ensemble methods (bagging, AdaBoost, and RF) and a multilayer perceptron (MLP) neural network. All baseline models were developed employing scikit-learn with default hyperparameters.

### 3.3. Performance metrics

In addition to accuracy, other metrics were developed to evaluate model performance when there is a large number of samples from one class as opposed to another;

– Precision (P): fraction of actual activity labels that are predicted as such.
– Recall (R): fraction of true positive activity labels out of all actual positive labels.
– F1-score: the harmonic mean of precision and recall.
– ROC-AUC: the ability of the model to tell apart different classes.

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}, \quad \text{F1} = \frac{2 \cdot P \cdot R}{P+R}$$

### 3.4. Results

Overall accuracy comparison: Table 3 shows classification accuracy. The proposed method outperforms all baselines across datasets, achieving 96.84% (UCI HAR), 95.73% (WISDM), and 92.41% (PAMAP2).

Table 3. Classification accuracy (%) comparison on baseline datasets

| Method | UCI HAR | WISDM | PAMAP2 |
|---|---|---|---|
| SVM | 93.85 | 91.27 | 87.62 |
| RF | 94.63 | 93.15 | 89.78 |
| Gradient boosting | 94.89 | 93.42 | 90.05 |
| Bagging | 94.21 | 92.87 | 89.31 |
| AdaBoost | 94.57 | 93.08 | 89.96 |
| MLP | 94.72 | 93.26 | 90.18 |
| Proposed method | 96.84 | 95.73 | 92.41 |

Figure 1 demonstrates the architecture for a multi-layered ensemble model; it describes the relationship between the base learners, meta-learners and feature optimization for better classifier performance. Figure 2 is referenced to demonstrate the comparative classification accuracy of each model with the proposed framework as having a higher level of performance than the other models. Figure 3 is referenced to show that the proposed method selects fewer features than the RF baseline, while still maintaining competitive levels of performance. Most notably, this was demonstrated on the PAMAP2 dataset, where the proposed method had an accuracy of 92.41% compared to the best performing non-proposed model being the MLP at 90.18%.

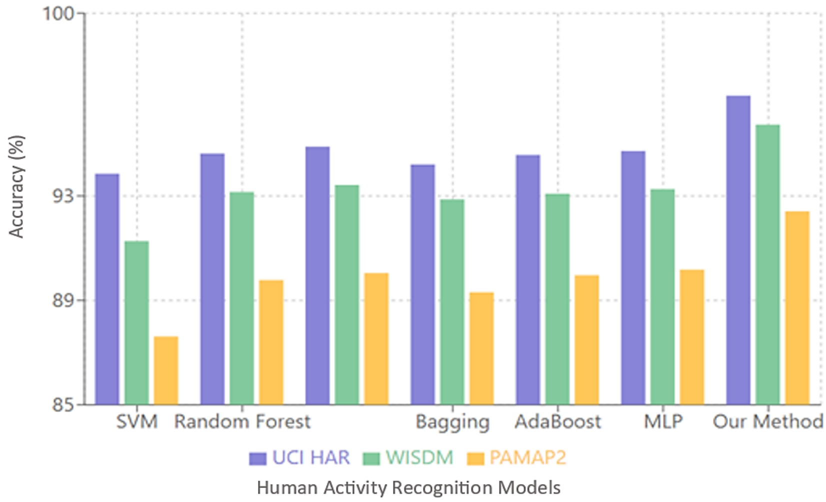


Figure 2. Accuracy comparison across baseline models and proposed method

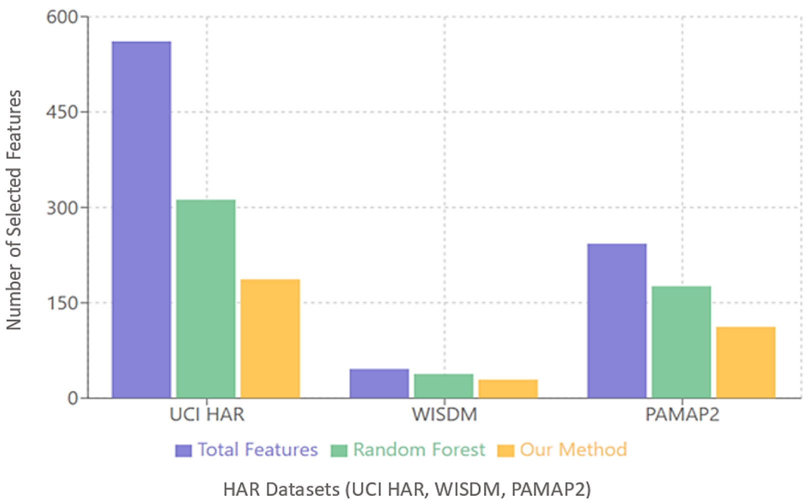


Figure 3. Feature reduction comparison between RF and proposed method

Metrics: Table 4 is used to report precision, recall, F1-score, and ROC-AUC. The method consistently showed good balance of all metrics compared to baseline methods with significantly better results than baseline methods on recall and F1-scores that are the most important to use when dealing with Har Datasets as they are typically extremely unbalanced.

Table 4. Extended performance metrics (mean±std, %)

| Dataset | Model | Precision | Recall | F1-Score | ROC-AUC |
|---|---|---|---|---|---|
| UCI HAR | RF | 94.2±0.3 | 94.0±0.4 | 94.1±0.3 | 96.5±0.2 |
| | MLP | 94.6±0.4 | 94.3±0.3 | 94.4±0.3 | 96.9±0.2 |
| | Proposed method | 96.9±0.2 | 96.7±0.2 | 96.8±0.2 | 98.7±0.1 |
| WISDM | RF | 92.7±0.5 | 92.3±0.6 | 92.5±0.4 | 95.8±0.2 |
| | MLP | 93.1±0.4 | 92.9±0.3 | 93.0±0.3 | 96.0±0.3 |
| | Proposed method | 95.6±0.2 | 95.4±0.3 | 95.5±0.2 | 97.9±0.1 |
| PAMAP2 | RF | 89.8±0.4 | 89.4±0.5 | 89.6±0.3 | 93.2±0.3 |
| | MLP | 90.2±0.3 | 90.1±0.3 | 90.1±0.2 | 93.5±0.2 |
| | Proposed method | 92.3±0.2 | 92.6±0.2 | 92.4±0.2 | 95.7±0.1 |

Statistical analysis: in order to verify that the baseline methods improved, paired T-Test and Wilcoxon Signed-Ranked test were completed between the proposed method and best performing baseline method (MLP). Results indicated statistical significance at the $p<0.01$ level across all data sets for accuracy, F1-score, and ROC-AUC, indicating it is highly improbable that the improvements observed were due to chance alone. Feature reduction: Table 5 shows dimensionality reduction.

Table 5. Comparison of selected features

| Dataset | Total features | RF | Proposed method |
|---|---|---|---|
| UCI HAR | 561 | 312 | 187 |
| WISDM | 46 | 38 | 29 |
| PAMAP2 | 243 | 176 | 112 |

Across three of the data sets, on average, this approach achieved a 43% reduction in feature space (UCI HAR (-66%), WISDM (-39%), and PAMAP2 (-36%)) relative to RF, indicating that the approach has consistent advantages over RF when it comes to reducing feature space dimensions across all data sets.

Table 6 presents the top ten most frequently selected features for the UCI HAR dataset, which includes both time domain and frequency domain features; such as mean, standard deviation, spectral entropy, and skewness.

Table 6. Top 10 features selected for UCI HAR dataset

| Rank | Feature name | Domain |
|---|---|---|
| 1 | tBodyAcc-mean ()-X | Time |
| 2 | tGravityAcc-mean ()-Y | Time |
| 3 | tBodyGyro-std ()-Z | Time |
| 4 | fBodyAcc-meanFreq ()-X | Frequency |
| 5 | tBodyAccJerk-correlation ()-X, Z | Time |
| 6 | fBodyGyro-bandsEnergy ()-1,8 | Frequency |
| 7 | tBodyAccMag-arCoeff ()3 | Time |
| 8 | fBodyAccJerk-skewness ()-X | Frequency |
| 9 | tGravityAccMag-entropy () | Time |
| 10 | fBodyGyro-kurtosis ()-Y | Frequency |

In Table 1, the top ten feature selections from the UCI HAR Dataset contain a balance of both time- and frequency-domain feature types. Time-domain type of the features that were selected are mean, standard deviation, correlation, autocorrelation coefficients and entropy (i.e., tBodyAcc-mean ()-X and tGravityAccMag-entropy ()). Frequency-domain type of the features that were selected are mean frequency, band power, skewness and kurtosis (e.g., fBodyAcc-meanFreq ()-X and fBodyGyro-kurtosis ()-Y). The combination of these types is an important part of the ability to model both the temporal and spectral aspects of human activities in order to perform HAR effectively.

## 4. DISCUSSION

In this section we will discuss the outcomes and implications of our suggested method for selecting features to optimize multilayer ensemble models for HAR. The most important findings from the experimental analysis of the proposed method are as:

### 4.1. Addressing gaps in prior research

Prior research has investigated either feature selection [9], [10] or ensemble learning [14], [27], however, never as an integrated layer-based approach. It is shown here that hierarchical optimization of layer-specific feature sets improves both accuracy, efficiency, and interpretability of HAR.

### 4.2. Summary of key findings

The proposed method has consistently demonstrated superior results compared to both the single-classifier performance (individually) and conventional ensemble methods, for all of the three benchmark data sets—UCI HAR, WISDM, and PAMAP2. The reasons for improved performance include:

- The multilayer ensemble was able to capture low level temporal patterns as well as higher level composite activities.
- Feature subset selection via GA, provided compact and highly discriminatory feature subsets while enhancing the interpretability of the multilayer ensemble.
- Results were above baseline for each metric and notably for the recall and F1 score, two metrics that are of particular interest in applications such as fall detection, where missing an event can be very expensive.

Baseline classifiers are compared in terms of accuracy to the proposed approach for all data sets as shown in Figure 2; with the proposed approach achieving a total accuracy rate of 96.84% for the UCI HAR data set; this is better than gradient boosting, at 94.89%. The same type of improvements can be seen in both the WISDM and PAMAP2 data sets; and these show that the proposed method can be generalized to multiple types of activities and different sensor arrangements.

The proposed optimized feature selection process was able to reduce the number of input features to the model by as much as 36%, and still produce the same level of accuracy; thus, it demonstrated an optimal balance of model complexity and accuracy.

### 4.3. Comparison with literature

The proposed method consistently selected fewer features than the baseline RF method but with the highest classification accuracy (Table 5). The proposed method's ability to tailor feature subsets to each layer also yielded superior ROC-AUC performance compared to [9]'s hybrid feature selection; [10]'s ensemble feature selection; [28]'s guided RF feature selection; and, unlike flat ensembles, the proposed hierarchical model progressively refined activity patterns.

As shown in Figure 3, the proposed approach has greatly reduced the number of selected features when compared to RF. On all three datasets (UCI HAR, WISDM, and PAMAP2), the proposed method selected a significantly lower number of features than either the total feature set or the RF baseline. The results from this study are consistent with previous studies demonstrating that optimal feature selection reduces dimensionality while enhancing model robustness as found in applications including sentiment analysis [14], and medical diagnosis/prediction.

### 4.4. Practical limitations

Limitations of the study are:

- The computational cost of the algorithm–training a GA is computationally costly and this limits its ability for real time adaptation in many cases.
- The sensitivity of GA parameters–the performance of GA depends heavily on parameters (e.g., population size, and mutation probability) that require careful tuning.
- Data bias–as with all experiments using wearable sensors as input data, these experiments are limited by the potential for generalization to vision-based HAR.

### 4.5. Implications for real-world deployment

The reduced feature dimensionality and improved recall make this method suitable for:

- Mobile health monitoring (low-power devices).
- Elderly care (fall detection).
- Smart wearables (activity tracking).

The findings also suggest applicability to other time-series domains such as anomaly detection and physiological signal monitoring.

### 4.6. Future research

Further research is needed on the:
– Take advantage of the real-time adaptation capabilities of online HAR methods [29], that may be integrated into our multilayer ensemble to improve overall responsiveness.
– Investigate transfer learning techniques to enable better cross-subject generalization.
– Additionally consider using multimodal data (i.e. environmental and physiological sensors).
– Apply explainable AI techniques to provide greater transparency.
– Take advantage of transfer learning techniques [30] to improve cross-subject generalization, which remains one of the main challenges in HAR.

## 5. CONCLUSION

This work proposed a multilayer ensemble architecture based on a multi-level GA-based feature selection strategy for the recognition of human Activities. This method is able to recognize both specific and abstract movement patterns, as opposed to previous methods that have used either one or the other of these strategies. The authors found statistical significance in their results on three benchmarks datasets (PAMAP2, UCI HAR, and WISDM) in terms of accuracy, precision, recall, F1 score, and ROC-AUC, and reduced the feature space by up to 36%. These results demonstrate the efficiency, robustness, and interpretability of this approach. The authors also identified three major implications of this study; namely, i) suitability for real-time use in wearable and mobile devices, ii) increased interpretability due to compact feature sets, and iii) generalization to other time series domains. However, there are two main limitations of this approach; namely, high computational costs, and reliance on labelled training data. Therefore, future research will be focused on lightweight optimization, online and transfer learning for greater adaptability, integration of multimodal sensors, and explainable AI for increased transparency. Overall, this study demonstrates that layer-specific feature optimization within multilayer ensembles provides a scalable and practical pathway to advancing HAR and related human-centered AI applications.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dhiraj Prasad Jaiswal | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Ashok Kumar Shrivastava | | ✓ | | ✓ | | | ✓ | | | ✓ | | ✓ | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [Dhiraj Prasad Jaiswal: dhiraj0411@gmail.com], upon reasonable request.

## REFERENCES

[1] R. K. Hatkeposhti, M. Y. Tabari, and M. A. Golsorkhtabari, "Fall Detection using Deep Learning Algorithms and Analysis of Wearable Sensor Data by Presenting a New Sampling Method," *International Journal of Engineering, Transactions A: Basics*,

vol. 35, no. 10, pp. 1941–1958, 2022, doi: 10.5829/ije.2022.35.10a.13.
[2] S. Patel, H. Park, P. Bonato, L. Chan, and M. Rodgers, “A review of wearable sensors and systems with application in rehabilitation,” *Journal of Neuro Engineering and Rehabilitation*, vol. 9, no. 1, p. 21, Dec. 2012, doi: 10.1186/1743-0003-9-21.
[3] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, “Elderly activities recognition and classification for applications in assisted living,” *Expert Systems with Applications*, vol. 40, no. 5, pp. 1662–1674, Apr. 2013, doi: 10.1016/j.eswa.2012.09.004.
[4] Ó. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013, doi: 10.1109/SURV.2012.110112.00192.
[5] A. Bulling, U. Blanke, and B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors,” *ACM Computing Surveys*, vol. 46, no. 3, pp. 1–33, Jan. 2014, doi: 10.1145/2499621.
[6] N. Y. Hammerla, S. Halloran, and T. Plötz, “Deep, convolutional, and recurrent models for human activity recognition using wearables,” *IJCAI International Joint Conference on Artificial Intelligence*, 2016, vol. 2016, pp. 1533–1540.
[7] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers and Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, doi: 10.1016/j.compeleceng.2013.11.024.
[8] A. Reiss and D. Stricker, “Introducing a new benchmarked dataset for activity monitoring,” *Proceedings - International Symposium on Wearable Computers (ISWC)*, Jun. 2012, pp. 108–109, doi: 10.1109/ISWC.2012.13.
[9] R. Saeedi, H. Ghasemzadeh, and L. V. Ojeda, “Moving beyond generation: A hybrid feature selection framework for human activity recognition using body worn sensors,” *2016 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2016, pp. 1–6.
[10] W. Xu, Y. Pang, Y. Yang, and Y. Liu, “Human Activity Recognition Based on Convolutional Neural Network,” *Proceedings - International Conference on Pattern Recognition*, Aug. 2018, vol. 2018, pp. 165–170, doi: 10.1109/ICPR.2018.8545435.
[11] A. Avci, S. Bosch, M.-P. Mihai, M.-P. Raluca, and P. Havinga, “Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey,” *23th International Conference on Architecture of Computing Systems 2010, ARCS 2010 - Workshop Proceedings*, 2010, pp. 167–176.
[12] F. J. Ordóñez and D. Roggen, “Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition,” *Sensors (Switzerland)*, vol. 16, no. 1, p. 115, Jan. 2016, doi: 10.3390/s16010115.
[13] C. A. Ronao and S. B. Cho, “Deep convolutional neural networks for human activity recognition with smartphone sensors,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9492, pp. 46–53, 2015, doi: 10.1007/978-3-319-26561-2_6.
[14] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/widm.1253.
[15] Q. Tan, Y. Qin, R. Tang, S. Wu, and J. Cao, “A Multi-Layer Classifier Model XR-KS of Human Activity Recognition for the Problem of Similar Human Activity,” *Sensors*, vol. 23, no. 23, p. 9613, Dec. 2023, doi: 10.3390/s23239613.
[16] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” *Pattern Recognition Letters*, vol. 119, pp. 3–11, Mar. 2019, doi: 10.1016/j.patrec.2018.02.010.
[17] D. Ravi, C. Wong, B. Lo, and G. Z. Yang, “Deep learning for human activity recognition: A resource efficient implementation on low-power devices,” *BSN 2016 - 13th Annual Body Sensor Networks Conference*, Jun. 2016, pp. 71–76, doi: 10.1109/BSN.2016.7516235.
[18] Y. Huang, Y. Zhou, T. Riedel, L. Fang, and M. Beigl, “randomHAR: Improving Ensemble Deep Learners for Human Activity Recognition with Sensor Selection and Reinforcement Learning,” *arXiv*, 2023, doi: 10.48550/arXiv.2307.07770.
[19] U. G. Mohammad, S. Imtiaz, M. Shakya, A. Almadhor, and F. Anwar, “An Optimized Feature Selection Method Using Ensemble Classifiers in Software Defect Prediction for Healthcare Systems,” *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/1028175.
[20] Y. Tian, J. Zhang, Q. Chen, and Z. Liu, “A Novel Selective Ensemble Learning Method for Smartphone Sensor-Based Human Activity Recognition Based on Hybrid Diversity Enhancement and Improved Binary Glowworm Swarm Optimization,” *IEEE Access*, vol. 10, pp. 125027–125041, 2022, doi: 10.1109/ACCESS.2022.3225652.
[21] N. Gopalakrishnan, V. Krishnan, and V. Gopalakrishnan, “Ensemble feature selection to improve classification accuracy in human activity recognition,” *Lecture Notes in Networks and Systems*, vol. 89, pp. 541–548, 2020, doi: 10.1007/978-981-15-0146-3_51.
[22] S. Mishra *et al.*, “Improving the Accuracy of Ensemble Machine Learning Classification Models Using a Novel Bit-Fusion Algorithm for Healthcare AI Systems,” *Frontiers in Public Health*, vol. 10, May 2022, doi: 10.3389/fpubh.2022.858282.
[23] T. Hasegawa and K. Kondo, “Easy Ensemble: Simple Deep Ensemble Learning for Sensor-Based Human Activity Recognition,” *IEEE Internet of Things Journal*, vol. 10, no. 6, pp. 5506–5518, Mar. 2023, doi: 10.1109/JIOT.2022.3222221.
[24] B. Xue, M. Zhang, W. N. Browne, and X. Yao, “A Survey on Evolutionary Computation Approaches to Feature Selection,” *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 606–626, Aug. 2016, doi: 10.1109/TEVC.2015.2504420.
[25] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. R-Ortiz, “A public domain dataset for human activity recognition using smartphones,” *ESANN 2013 Proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013, pp. 437–442.
[26] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, Mar. 2011, doi: 10.1145/1964897.1964918.
[27] M. Wang, Y. Qian, Y. Yang, H. Chen, and W. F. Rao, “Improved stacking ensemble learning based on feature selection to accurately predict warfarin dose,” *Frontiers in Cardiovascular Medicine*, vol. 10, Jan. 2023, doi: 10.3389/fcvm.2023.1320938.
[28] D. Thakur and S. Biswas, “An Integration of feature extraction and Guided Regularized Random Forest feature selection for Smartphone based Human Activity Recognition,” *Journal of Network and Computer Applications*, vol. 204, p. 103417, Aug. 2022, doi: 10.1016/j.jnca.2022.103417.
[29] C. Zhu and W. Sheng, “Motion- and location-based online human daily activity recognition,” *Pervasive and Mobile Computing*, vol. 7, no. 2, pp. 256–269, Apr. 2011, doi: 10.1016/j.pmcj.2010.11.004.
[30] D. J. Cook, K. D. Feuz, and N. C. Krishnan, “Transfer learning for activity recognition: A survey,” *Knowledge and Information Systems*, vol. 36, no. 3, pp. 537–556, Sep. 2013, doi: 10.1007/s10115-013-0665-3.

## BIOGRAPHIES OF AUTHORS

**Mr. Dhiraj Prasad Jaiswal** did his Bachelor of Computer Application from Siliguri College (Affiliated to North Bengal University), in the year 2015 and he did his Master of Computer Application from SRM University Sikkim in the year 2017. He has qualified UGC NET. He has more than 8 years of experience in teaching including industry experience of 1.5 years. He is pursuing his Ph.D. in Computer Science and Engineering from Amity University, Gwalior. His research interests include cloud computing, Android application development, software engineering, artificial intelligence, the internet of things, and machine learning, with a particular focus on human activity recognition. He can be contacted at email: dhiraj0411@gmail.com.

**Dr. Ashok Kumar Shrivastava** working as an associate professor and academic coordinator in the Department of Computer Science and Engineering at Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior. He graduated in Mathematics from Science College Gwalior. He has done Master of Science in CS branch from Jawaji University, Gwalior. He secured M.Tech. CSE form Uttar Pradesh Technical University, Lakhnow. He earned his Ph.D. in Image Processing from NIU, Grater Noida. He is in teaching profession for more than 19 years. He has published 6 Scopus indexed, 2 SCI, and 13 Google Scholer papers in National and International Journals, Conference and Symposiums. He published 04 Patents, 03 Copyrights and 2 Books. His main area of interest includes noise detection and removal, image enhancement, and machine learning. He can be contacted at email: ashok79.shrivastava@gmail.com.