Codebook-Based Feature Engineering for Human Activity Recognition Using Multimodal Sensory Data

Seerat Fatima[1], Laiba Zahid[1], Tazeem Haider[2], Muhammad Hassan Khan[*2], and Muhammad Shahid Farid[2]
[1]Department of Software Engineering, University of the Punjab, Lahore, Pakistan.
[2]Department of Computer Science, University of the Punjab, Lahore, Pakistan.
[*]**Correspondence:** hassankhan@pucit.edu.pk

Citation | Fatima. S, Zahid. L, Haider. T, Khan. M. H, Farid. M. S, "Codebook-Based Feature Engineering for Human Activity Recognition Using Multimodal Sensory Data", IJIST, Special Issue pp 56-69, Oct 2024
Received | Oct 05, 2024 **Revised |** Oct 09, 2024 **Accepted |** Oct 13, 2024 **Published |** Oct 19, 2024.

Recently, Human Activity Recognition (HAR) using sensory data from various devices has become increasingly vital in fields like healthcare, elderly care, and smart home systems. However, many existing HAR systems face challenges such as high computational demands or the need for large datasets. This paper introduces a codebook-based approach designed to overcome these challenges by offering a more efficient method for HAR with reduced computational costs. Initially, the raw time series data is segmented into smaller subsequences, and codebooks are constructed using the Bag of Features (BOF) approach. Each subsequence is then assigned softly based on the center of each cluster (codeword), resulting in a histogram-based feature vector. These encoded feature vectors are subsequently classified using a Support Vector Machine (SVM). The proposed method was evaluated using the OPPORTUNITY dataset, comprising data from 72 sensors, achieving a classification accuracy of 90.7%. In comparison to other advanced techniques, our approach not only demonstrated superior accuracy in recognizing human activities but also significantly reduced computational costs. The use of soft assignments for mapping codewords to subsequences efficiently captured the key patterns within the activity data. The findings validate that the proposed codebook-based method provides substantial improvements in both accuracy and efficiency for HAR systems.

Keywords: Multimodal Sensory Data; Codebook; Bag of Features; Mini Batch K-Means; Soft Assignment.

Introduction:

HAR is a rapidly evolving field within machine learning and pervasive computing, focused on recognizing and interpreting human actions and behaviors using sensory data from wearable devices like smartphones, smart glasses, and smartwatches [1,2]. HAR systems collect data from various sources to identify simple activities such as sitting, walking, and running, as well as more complex ones like cooking, driving, and cleaning [1,2]. As modern societies increasingly incorporate smart devices into their daily routines to monitor and enhance activities, the demand for accurately recognizing daily human actions has grown significantly.

Existing HAR techniques typically rely on visual and time-series sensing modalities to record and analyze human movements, as illustrated in Figure 1. Visual data is captured using different types of cameras, including standard color cameras, multi-view setups, and Time-of-Flight (TOF) cameras. However, camera-based approaches often raise privacy concerns, as continuous video recording can be intrusive, and individuals may prefer not to be under constant surveillance. Additionally, the installation and maintenance of such systems involve considerable costs [3].

Conversely, time-series sensory data, collected through Inertial Measurement Units (IMUs) and pressure sensors, addresses privacy concerns and overcomes the limitations of camera-based methods. These sensors are integrated into devices like smartphones, smartwatches, and smart glasses, offering enhanced privacy since they do not capture visual information. Furthermore, they are less reliant on external factors such as placement and lighting conditions, making them easier to integrate into wearable devices [4].

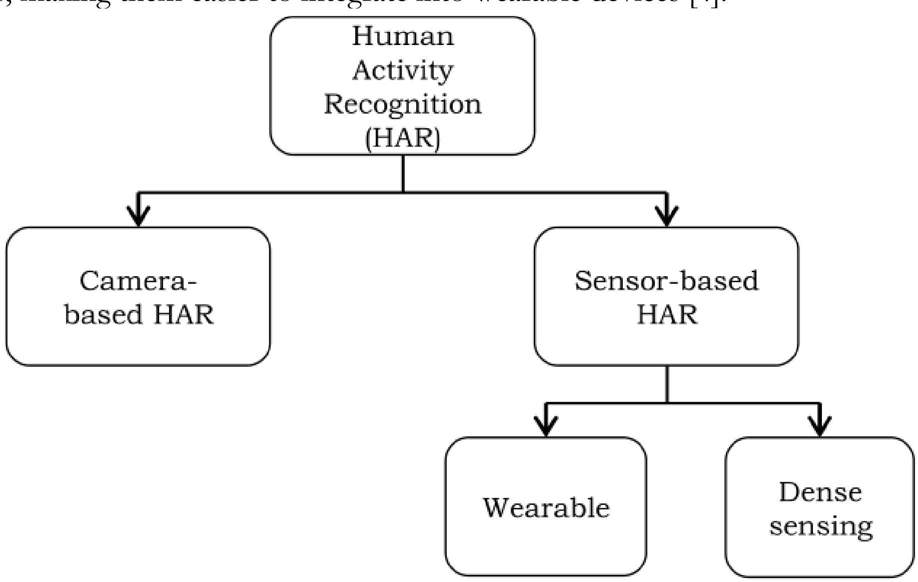

Figure 1. Categorization of different approaches of Human Activity Recognition based on different modalities

Researchers have explored various approaches to achieve precise and effective HAR, focusing on techniques for feature extraction, representation, and classification. Existing HAR systems can be broadly categorized into three groups based on their implementation methodologies: handcrafted feature-based techniques, deep learning approaches, and codebook-based methods. Each of these has distinct advantages and limitations.

Handcrafted feature-based techniques for HAR involve manually extracting features from preprocessed sensory data, such as domain-specific features, time-domain features (e.g., signal magnitude area), frequency-domain features (e.g., Fourier coefficients), and statistical measures (e.g., mean, variance, standard deviation) that are deemed relevant for distinguishing different activities. These features are then used as inputs for machine learning models to classify activities. However, this method is time-consuming and requires considerable domain expertise [5][6].

Codebook-based techniques refer to methods that represent sensory data as compact, high-level histogram-based feature vectors, created through unsupervised clustering algorithms. These encoding algorithms have demonstrated effective, state-of-the-art recognition performance across various time-series datasets for classification and activity recognition [7]. Deep learning-based techniques for HAR use neural networks to automatically learn features from raw sensory data, eliminating the need for manual feature engineering [8][9]. The key advantages of these methods include high accuracy, scalability to large datasets, and the ability to capture complex patterns in data. However, they also have significant drawbacks, such as higher computational demands, susceptibility to overfitting with limited data, and the need for large labeled datasets for effective training. A detailed summary of the literature is presented in the following subsections.

Handcrafted Feature-Based Techniques:

In the domain of machine learning and human activity recognition (HAR), various techniques focus on extracting and selecting features that effectively capture the characteristics of the activities to be recognized. Examples of these features include statistical features (e.g., mean, variance) [10], temporal features (e.g., zero-crossing rate) [11], frequency features (e.g., Fourier transform) [12], and spatial features (e.g., Local Binary Patterns, Histogram of Oriented Gradients (HOG)) [13]. Once selected for their relevance, these extracted features are input into machine learning algorithms such as support vector machines (SVM) or k-nearest neighbors (KNN) for classification.

Numerous studies have demonstrated the reliability and interpretability of this approach for HAR [14]. In [15], the authors combined handcrafted features with deep features, using a fusion of conventional handcrafted features like the Histogram of Oriented Gradients (HOG) and deep learning-based representations for human activity recognition. HOG is particularly effective in encoding spatial gradient information, making it suitable for distinguishing different types of human activity, as discussed in [16]. Additionally, the authors in [14] utilized Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) to extract key points and descriptors from activity data.

Handcrafted feature-based techniques play a significant role in HAR by offering a balance between computational efficiency and performance. However, despite their relative simplicity, these methods heavily rely on domain-specific expertise. They also tend to be less adaptable and are often tailored to specific problems. Furthermore, if new classes are introduced into the training process, the features must be recalculated, which can be a limitation in dynamic environments [17].

Codebook-Based Approaches:

Codebook-based approaches employ the bag-of-feature (BOF) method to represent high-dimensional, complex data using a discrete set of representative patterns known as codewords [18]. The process begins with extracting local features from raw sensory data, typically obtained from devices like accelerometers and gyroscopes. Using clustering algorithms such as K-means, these features are then grouped to form a codebook, with each cluster center serving as a codeword that represents a specific pattern within the data. Once the codebook is established, each new local feature is matched to its closest codeword, effectively quantizing the continuous feature space into discrete, manageable components [19]. This transformation results in a histogram of codewords, offering a compact and discriminative representation of the activity [20].

The authors in [21] provided a comprehensive overview of the applications and challenges of various HAR strategies, including codebook-based approaches. Codebook-based methods have demonstrated adaptability to different sensor data types and activity categories, striking a balance between computational efficiency and accuracy [22]. Furthermore, these techniques capture hidden patterns in human movement through clustering, which contributes to their superior performance in activity recognition [23].

Deep Learning-Based Approaches:

Deep learning techniques have significantly advanced the field of Human Activity Recognition (HAR) by enabling the automatic extraction of high-level features directly from raw sensory data. These methods utilize a variety of neural network architectures, including recurrent neural networks (RNNs) and convolutional neural networks (CNNs), as well as their combinations, to capture both the temporal and spatial dimensions of human activities. For example, CNNs are adept at identifying patterns in visual inputs like photos and videos. In contrast, RNNs, particularly Long Short-Term Memory (LSTM) networks, excel at processing sequential data, making them well-suited for capturing temporal relationships crucial for HAR [24]. Recent advancements in this field have further improved the accuracy and efficiency of HAR systems through deep learning. Incorporating attention mechanisms into CNN architectures has been shown to focus on the most relevant parts of input data, thereby enhancing feature extraction. This technique has significantly boosted HAR model performance across various datasets [25]. Additionally, hybrid models that integrate LSTM for temporal feature analysis and CNN for spatial feature extraction have gained popularity, offering robust methods for identifying complex behaviors from multimodal sensor data [26].

Despite their strong performance in HAR, deep learning-based approaches require large amounts of labeled data and substantial computational power to achieve high accuracy. The involvement of numerous hyperparameters can further complicate the architecture, requiring additional computation for fine-tuning. In contrast, codebook-based approaches are less demanding in terms of computational resources and data volume, achieving efficient performance with fewer resources and less computation time. Given these practical considerations, we have selected the codebook-based approach for our study, aiming to balance efficiency and performance without the need for extensive resources.

Initially, the data was cleaned and filtered to ensure accurate analysis. Next, it was preprocessed to calibrate and normalize the raw sensor values, ensuring readiness for further analysis. After preprocessing, time-series segmentation techniques were applied to extract subsets of activity within overlapping time windows [27]. These segmented subsets constituted the feature space. Subsequently, these extracted features served as inputs for the codeword generation model, which utilized segmented data to produce codewords. Specifically, we employed a codebook-based approach using the Mini Batch k-means clustering model to generate codewords, identifying patterns and relationships within the data. The model was then fine-tuned and optimized to enhance accuracy, ultimately enabling reliable classification of human activities.

Objectives of study:

Raw sensory data from wearable devices cannot be directly utilized by machine learning algorithms, necessitating an appropriate feature representation to accurately recognize daily human activities. The primary objective of this research is to develop a computationally efficient method for Human Activity Recognition (HAR) using sensory data through a codebook-based approach. This approach aims to improve the accuracy of activity recognition while reducing computational requirements. By leveraging the Bag of Features (BOF) technique for feature encoding and utilizing a Support Vector Machine (SVM) for classification, this method seeks to achieve enhancements in both performance and accuracy. **Novelty Statement:**

This study presents a codebook-based approach to overcome the challenges of high computational costs and the need for large datasets in Human Activity Recognition (HAR). The proposed method offers an efficient solution for HAR, designed to perform well with smaller datasets and reduced computational demands. The major contributions of this paper include:

- A computationally efficient codebook-based HAR method, utilizing the mini-batch k-means clustering algorithm, has been proposed.
- The performance of the proposed method was evaluated using the OPPORTUNITY dataset, demonstrating higher accuracy compared to many existing techniques that are more computationally intensive and resource-demanding.

Material and Methods:

Existing research demonstrates that deep learning algorithms often deliver superior performance for HAR. However, a significant drawback of deep learning architectures is the large number of hyperparameters, which necessitate a thorough and often time-consuming optimization process. In contrast, codebook-based approaches are effective in uncovering hidden patterns in human movements through the use of clustering algorithms. Thus, the proposed method leverages a codebook-based approach for feature engineering.

The codebook approach consists of three main steps: codebook construction, feature encoding, and classification. In this study, we employed the Bag of Features (BOF) approach, utilizing a minibatch k-means clustering algorithm for both codebook construction and codeword generation. Figure 2 illustrates the steps involved in the proposed method, while Algorithm 1 outlines the detailed steps of the study. The following subsections provide the implementation details of the proposed method.

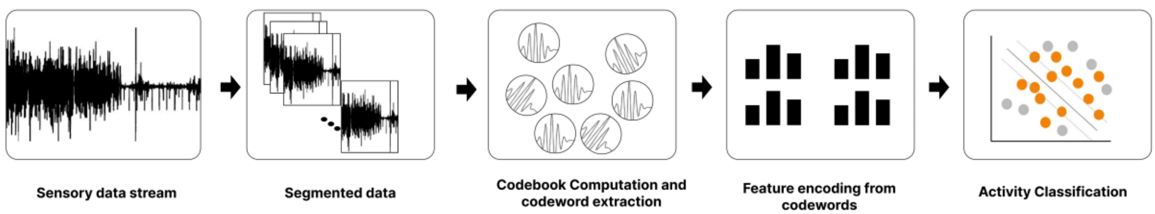

Figure 2. An overview of the proposed approach.

The process consists of several key steps: First, the raw data is segmented into smaller subsequences, followed by the generation of codebooks, where each codebook center represents a codeword. Next, feature encoding is performed using these codewords. Finally, the encoded features are used as input for a classifier to recognize the activities.

Algorithm 1: Human Activity Recognition using Codebook-Based Feature Engineering

Input: Dataset D with raw sensor data and labels
Parameters: Window size w = 24, Overlap o = 1, Soft assignment factor β = 0.001, Number of clusters k = 64.
Output: Predicted labels for test data
1: Load dataset D containing raw data and corresponding labels.
2: Extract subsequences from each data sample using window size w and overlap o.
3: **Apply Mini Batch K-Means on subsequences:**
4: Initialize K-means with k clusters.
5: Fit model to the extracted subsequences.
6: Obtain k cluster centers (codewords).
7: Write codewords to output file for later use.
8: Split dataset D into training set D_{train} and test set D_{test}.
9: **for** each unique label y_i in D_{train} **do**
10: **for** each data sample with label y_i **do**
11: Extract subsequences.
12: Compute distances between subsequences and codewords.
13: Select the top k closest codewords.
14: Calculate soft assignments based on distances using β.
15: Accumulate soft assignments into histogram for label y_i.

```
16:         end for
17:         Normalize the histogram for label y_i.
18: end for
19: Repeat step 9 for test set D_test.
20: Train an SVM classifier with a linear kernel using training histograms.
21: Predict labels for D_test using the trained SVM model.
22: Evaluate model accuracy by comparing predicted and true labels.
23: Return accuracy score as the final performance metric.
```

Raw data cannot be used directly due to the presence of noise, null value issues, and variability in the data [28]. Additionally, sensory datasets often exhibit high dimensionality, which can result in slow computations [29]. These challenges pose significant obstacles to effective activity recognition and must be addressed to ensure efficient classification. Therefore, preprocessing of the raw sensory data is essential before proceeding with codebook generation and classification.

To achieve this, rows containing null values were first removed, and the data was then normalized using techniques such as robust scaling [30] and z-score normalization [30][31]. By applying these techniques, the data becomes suitable for feature extraction, enabling us to reduce dimensionality and select the most relevant feature set.

Data Segmentation:

Given the challenges of high dimensionality and sparsity in sensory data, we employed dimensionality reduction techniques [32] to extract the most relevant feature set. To effectively manage continuous time-series data, it is more efficient to segment it into shorter subsequences that represent specific activities within designated time frames. For this purpose, we utilized a sliding window approach, defined by a window size \ (w \) and a sliding stride \ (l \), applied to each data sequence to obtain subsequences of \ (w \)-dimensions, as described in the study [27]. Figure 3 illustrates the segmentation of the data into various subsequences.

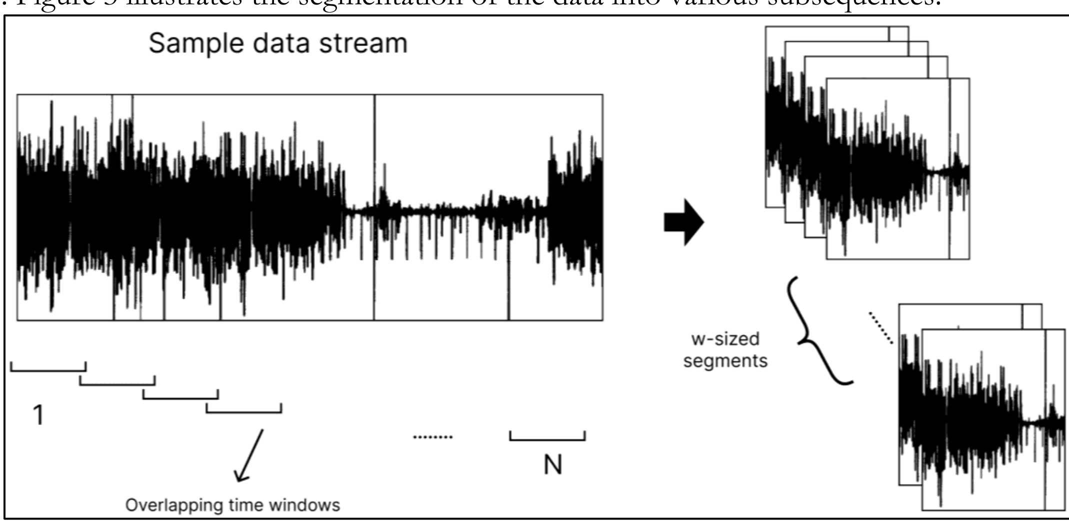

Figure 3. Segmentation of continuous stream of time-series data using sliding window approach

There are two primary methods for segmenting the data: randomly selecting subsequences or using fixed-size subsequences [33][28]. We opted for the latter, segmenting our data with a window size of 24 and an overlap size of 1. Selecting an appropriate window size is critical and should reflect the nature of the activities. Shorter activities warrant a smaller window size, while longer and more complex activities require a larger window size to accurately capture the essence of the activity. The overlapping technique ensures that no data is lost during the dimensionality reduction process while preserving essential features. Consequently, both window size and overlap can be adjusted to optimize performance for different activities and datasets.

To generate the subsequences, the time-series data is first concatenated into a single vector, which is then segmented according to the predetermined window size. The resulting vectors, each of size \ (w \), facilitate dimensionality reduction and make computations less resource-intensive. These resultant vectors are subsequently utilized for codebook construction.

Codebook Construction and Codeword Extraction:

For codebook construction, we employed the Mini Batch K-means algorithm, which serves as an alternative to the standard K-means algorithm for clustering large datasets. One of the key advantages of this algorithm is its ability to reduce computational costs by utilizing a fixed-size sub-sample from the dataset. This approach minimizes the number of distance computations per iteration, albeit at the expense of slightly lower cluster quality [34]. The Mini Batch K-means method is particularly suitable for large datasets, as it significantly decreases both computation time and memory usage compared to traditional K-means. The steps involved in the clustering algorithm are as follows:

- **Initialization:** The algorithm begins by randomly selecting initial cluster centroids.
- **Cluster Assignment:** Euclidean distance is employed to assign each subsequence to its nearest centroid, effectively grouping subsequences with similar patterns.
- **Centroid Update:** Through an iterative process, the centroids are updated by calculating the mean of the assigned subsequences. This cycle continues until there is no further change in the centroids.
- **Saving the Codebook:** Each cluster constitutes a codebook, with the updated centroids serving as the codewords. These codewords represent the primary patterns within our sensor data.

Feature Encoding:

After generating the codebook, the next step is to quantize the feature vectors using this codebook. During this process, each subsequence is probabilistically assigned to the nearest codewords, transforming the feature vectors into discrete symbols. The similarity between each subsequence and the codewords is determined by calculating the Euclidean distance, with the closest codewords representing the subsequences.

Rather than using a hard assignment approach, where each subsequence is assigned to a single codeword, we employed a Soft Assignment method. This technique calculates the weighted contribution of the nearest codewords to each subsequence [27]. The soft assignment is computed using a SoftMax function with a temperature parameter (beta). This function normalizes the distances to a range of 0 to 1, effectively converting them into probabilities, thereby assigning higher weights to closer codewords and lower weights to those that are farther away. The soft assignment for a given distance \ (d \), denoted as soft assignment(d), can be expressed as follows:

$$\text{Soft assignment}(d) = \frac{e^{-\beta . d}}{\sum_{j=1}^{k} e^{-\beta . d_j}}, \quad (1)$$

Here, \ (d \) represents the specific distance value for which the soft assignment is being calculated, while \ (d_j \) denotes the distance corresponding to the \ (j \)-th sorted distance index among the first \ (k \) sorted distances. The parameter \ (\beta \) is a smoothing factor that controls the softness of the assignment and can be tuned using cross-validation. In our implementation, we set \ (\beta \) to 0.001, with \ (k \) representing the number of nearest neighbors considered. The numerator, which is the exponential of the negative distance scaled by \ (\beta \), regulates the sharpness of the assignment and is divided by the sum of the exponentials of the distances for the \ (k \) nearest codewords. This process results in a probability distribution over the codewords for each subsequence. Finally, the soft assignments are normalized to ensure that the total contribution equals one, providing a robust representation of the subsequences.

We utilized vector quantization to represent the frequency distribution of codewords within a set of subsequences. This method efficiently and accurately captures the overall distribution of codewords in the data. For each activity label in the training data, we constructed a histogram-based feature vector by aggregating the soft assignments of all subsequences corresponding to that label. A feature vector of length \(k \) (where \(k \) is the size of the clusters in the codebook) is initialized with zeros, and the soft assignments for each subsequence of the current activity are added to the corresponding positions in that vector [27]. Subsequently, the histograms are normalized so that their sum equals one, ensuring they represent relative frequencies and are thus comparable across various activities. Similarly, the encoded feature vector for the test data is constructed to maintain consistent feature representation in both the training and testing phases. These \ (k \)-dimensional, histogram-based encoded feature vectors serve as inputs to our classifier.

Activity Recognition:

Support Vector Machines (SVM) is a state-of-the-art algorithm based on the robust theoretical framework of Vapnik-Chervonenkis theory [35]. SVM is particularly known for its strong regularization capabilities, which enable the model to generalize effectively to new data. Designed specifically for supervised learning classification tasks, it excels at finding the optimal hyperplane that separates different classes by maximizing the margin between data points of varying classes. We have chosen SVM for our study due to its proven ability to achieve high accuracy with time-series segmented data using the Bag of Features (BoF) approach [36]. Research findings suggest that the SVM model can attain the highest precision and recall for Human Activity Recognition (HAR) [37][38]. The steps involved in implementing SVM are:

Data Preparation:

To prepare the data, the histograms and their corresponding labels are divided into training and testing sets.

SVM Training:

We utilized the linear kernel due to its simplicity and effectiveness in high-dimensional spaces. The linear kernel calculates the dot product between feature vectors, making it well-suited for linearly separable data. Moreover, SVM with a linear kernel effectively handles nonlinear classification by applying a linear discriminant function in a high-dimensional space [39]. The equation for linear kernel SVM is:

$$f(x)=\mathbf{w}^T\mathbf{x}+b, \qquad (2)$$

Here, \ (x \) represents the input feature vector, \ (w \) is the weight vector, and \ (b \) is the bias term. The values of \ (w \) and \ (b \) are determined by solving the following optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \qquad (3)$$

Subject to:

$$y_i(\mathbf{w}^T\mathbf{x}_i+b)\geq 1-\xi_i \qquad (4)$$

$$\xi_i \geq 0,$$

In this context, \(||w||^2 \) represents the norm of the weight vector (which needs to be minimized), \(C \) is the regularization parameter that controls the trade-off between maximizing the margin and minimizing classification error, \(\xi_i \) are the slack variables that permit some misclassification of the training samples, \(y_i \) are the labels of the training samples, and \(x_i \) are the training samples.

The SVM model is then trained using the training set of histogram-based feature vectors. The model learns to classify the activity histograms by identifying the decision boundary that maximizes the margin between different classes. SVM solves this convex optimization problem to find the hyperplane that maximizes the margin, which is defined as the distance between the hyperplane and the nearest data points from each class.

Model Evaluation:

The trained SVM model is evaluated using the testing set. Performance metrics, including accuracy, are calculated to assess the model's effectiveness.

Table 1. The summary of OPPORTUNITY dataset. It contains the data of 72 sensors and 10 modalities.

Attributes	Values
No. of participants	12
No. of locomotion activities	5
No. of gesture activities	10
Data dimensions	3
Time given to each subject for activities	25-30 min
No. of modalities	10
Total Sensors	72

Result and Discussion:

Dataset:

We used the OPPORTUNITY dataset to evaluate the performance of the proposed method. Collected in the study [40], this publicly available dataset contains raw data from 72 sensors across 10 modalities. The sensors were deployed on the bodies of the subjects as well as on objects in the environment. We assessed the proposed model for the recognition of five locomotion activities within the dataset: standing, walking, sitting, lying, and null activities. Table 1 provides a comprehensive summary of the data contained in this dataset, while Table 2 lists the activities included.

Table 2. Activities included in the OPPORTUNITY dataset where standing, walking, sitting, lying, and null activity are locomotive and rest of these are gesture activities.

Standing	Walking	Sitting
Lying	null activity	Open door 1
Open door 2	Close door 1	Close door 2
Clean Table	Drink from cup	Toggle switch
Pour water	Stir pot	Chop with knife

Performance Metrics:

To evaluate the effectiveness of our activity classification model, we utilized the following performance metrics:

Accuracy:

This performance metric quantifies the ratio of correctly classified instances to the total number of instances, providing a comprehensive evaluation of the model's performance across all classes.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

Precision is defined as the ratio of true positive predictions to the total number of positive predictions made by the model. This metric assesses the accuracy and quality of the model's positive predictions.

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

Recall: Also known as sensitivity, calculates the ratio of true positive predictions to the total number of actual positive instances. It evaluates the model's effectiveness in identifying all relevant cases within a given class.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F_1 Score:

This measure represents the harmonic mean of precision and recall, providing a unified metric that balances the trade-off between the two. It is especially useful in scenarios with imbalanced class distributions.

$$\textbf{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where TP is True Positives, TN is True Negatives, FP is False Positive and FN is False Negatives.

Table 3. Summary of the results obtained by proposed method. The proposed method was evaluated on OPPORTUNITY dataset on the basis of four metrics: Accuracy, Precision, Recall and F_1 score.

Metrics	Results (%)
Accuracy	90.7
Precision	89.1
Recall	89.3
F_1 score	89.19

Table 4. Comparison of proposed method with existing state-of-the-art methods for HAR.

Methods	Year	Accuracy (%)
Hierarchical multi-task leaning [41]	2023	82.31
Personalized ML model with hybrid data split [42]	2023	84.87
Deep Belief Network (DBN) [43]	2021	85.7
Recurrent Neural Network [44]	2022	86.9
LSTM (Long Short-Term Memory [44]	2022	90.1
Convolutional Neural Network [45]	2022	90.19
Proposed Codebook-based method	**2024**	**90.7**

Results:

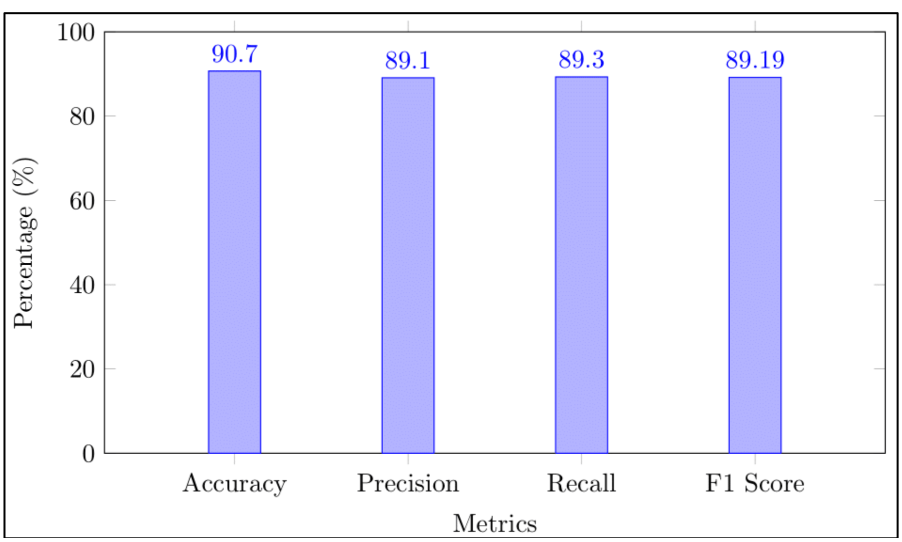

Figure 4. Visual representation of the performance of our classification model based on four key metrics: Accuracy, Precision, Recall, and F_1 score. These metrics show how well the model recognizes activities, managing precision and recall, and performing overall in identifying activities.

The results of the proposed method demonstrate strong performance across all activities in the dataset, achieving high accuracy rates along with balanced precision and recall. The F1 score further validates the model's effectiveness in classification tasks. As shown in Table 3, we attained an accuracy of 90.7%, significantly surpassing other methods on this dataset. The high precision and recall indicate a minimal occurrence of false positive and false negative errors generated by our approach. Figure 4 illustrates the results of the proposed method on the OPPORTUNITY dataset.

The proposed method surpasses many resource- and computation-intensive approaches for Human Activity Recognition (HAR). Table 4 presents a comparison of our codebook-based approach with existing state-of-the-art HAR methods, illustrating that our method achieves superior accuracy. Notably, our model demonstrates significantly better performance than various machine learning and deep learning techniques. While the performance of other models may be comparable, they tend to be far more computationally expensive and require substantially larger datasets compared to our proposed codebook-based approach.

Discussion:

The proposed codebook-based method for Human Activity Recognition (HAR) demonstrated a significant enhancement in both accuracy and efficiency when evaluated on the OPPORTUNITY dataset, achieving a classification accuracy of 90.7%. This level of accuracy surpasses that of many state-of-the-art HAR approaches. The soft assignment method for mapping codewords to subsequences effectively captures subtle variations in human activities, contributing to the method's high accuracy. While this approach offers a more lightweight and scalable solution compared to deep learning models, further testing on additional datasets is necessary to evaluate its broader applicability. Moreover, refining the clustering process could yield even better results. Future work may investigate the method's performance across different environments and its potential for real-time applications.

Conclusion:

This paper presents a framework for recognizing human activities using sensory data. Unlike machine learning and deep learning approaches, which require extensive training datasets, our codebook-based approach operates effectively without large datasets or significant computational resources. The process begins with the preprocessing and normalization of sensory data, followed by segmentation into smaller subsequences to enhance activity recognition efficiency. Next, codebooks are constructed using the mini-batch k-means clustering algorithm, from which codewords are extracted. These codewords are then represented as compact feature vectors using a soft assignment approach, which serve as inputs to the Support Vector Machine (SVM) classifier. The proposed method was evaluated on the OPPORTUNITY dataset, achieving an impressive accuracy of 90.7%. In addition to accuracy, we calculated other performance metrics, including precision, recall, and the F1 score, which reached 89.19%. This indicates a strong balance between precision and recall, essential for minimizing both false positives and false negatives. Furthermore, the results of our method were compared with existing state-of-the-art techniques for HAR, demonstrating superior performance.

Recommendations for Future Work:

We have determined that our combination of a codebook-based approach, Mini Batch K-means clustering, and Support Vector Classification is significantly beneficial in categorizing activities by interpreting high-performance metrics. Future work could explore several avenues to further enhance the system's performance and applicability:

- **Integrating Additional Sensor Data:** Incorporating data from additional types of sensors, such as environmental sensors or wearable ECG monitors, could provide more comprehensive insights into the user's activities and health status.
- **Advanced Machine Learning Algorithms:** Experimenting with more advanced machine learning algorithms, such as deep learning models, could potentially improve the accuracy and robustness of the classification system.
- **Real-World Testing and Deployment:** Conducting extensive real-world testing and

deployment will provide valuable feedback on the system's performance in practical scenarios, helping to identify areas for refinement and optimization.

- **Personalization and Adaptability:** Developing methods to personalize the system for individual users, adapting to their specific activity patterns and physiological responses, could enhance its effectiveness and user acceptance. The proposed method provides a thorough and efficient method of classifying various human activities, which shows a better performance and resilience for HAR.

Acknowledgement: Th is research was partially supported by the Higher Education Pakistan within the project "Hum Care: Human Activity Analysis in Healthcare" (Grant Number: 15041).

Author's Contribution: Conceptualization, S.F., L.Z., T.H., M.H.K. and M.S.F.; Methodology, S.F. and M.H.K.; Software, L.Z. and M.S.F.; validation, S.F.; L.Z.; investigation, T.H., M.H.K. and M.S.F.; writing-original draft preparation, T.H.; writing-review and editing, M.H.K. and M.S.F.; supervision, M.H.K. and M.S.F.; project administration, M.H.K. and M.S.F. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest: The authors declare no conflict of interest.

Project Details: This research was conducted under the project titled "Hum Care: Human Activity Analysis in Healthcare" under grant number: 15041 (Duration: 2022 - 2025).

References:

[1] Nazish Ashfaq, Muhammad Hassan Khan, and Muhammad Adeel Nisar. Identification of optimal data augmentation techniques for multimodal time-series sensory data: A framework. Information, 15(6):343, 2024.

[2] Muhammad Hassan Khan. Human activity analysis in visual surveillance and healthcare, volume 45. Logos Verlag Berlin GmbH, 2018.

[3] Lukas Köping, Kimiaki Shirahama, and Marcin Grzegorzek. A general framework for sensor-based human activity recognition. Computers in biology and medicine, 95:248–260, 2018.

[4] L Minh Dang, Kyungbok Min, Hanxiang Wang, Md Jalil Piran, Cheol Hee Lee, and Hyeonjoon Moon. Sensor- based and vision-based human activity recognition: A comprehensive survey. Pattern Recognition, 108:107561, 2020.

[5] Rimsha Fatima, Muhammad Hassan Khan, Muhammad Adeel Nisar, Rafał Doniec, Muhammad Shahid Farid, and Marcin Grzegorzek. A systematic evaluation of feature encoding techniques for gait analysis using multimodal sensory data. Sensors, 24(1), 2024.

[6] Hong Liang, Xiao Sun, Yunlei Sun, and Yuan Gao. Text feature extraction based on deep learning: a review. EURASIP journal on wireless communications and networking, 2017:1–12, 2017.

[7] Muhammad Hassan Khan, Muhammad Shahid Farid, and Marcin Grzegorzek. A comprehensive study on codebook-based feature fusion for gait recognition. Information Fusion, 92:216–230, 2023.

[8] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. Convolutional neural networks for time series classification. Journal of systems engineering and electronics, 28(1):162–169, 2017.

[9] Lam Tran, Thang Hoang, Thuc Nguyen, Hyunil Kim, and Deokjai Choi. Multi-model long short-term memory network for gait recognition using window-based data segment. IEEE Access, 9:23826–23839, 2021.

[10] Fatima Amjad, Muhammad Hassan Khan, Muhammad Adeel Nisar, Muhammad Shahid Farid, and Marcin Grze- gorzek. A comparative study of feature selection approaches for human activity recognition using multimodal sensory data. Sensors, 21(7):2368, 2021.

[11] Zhenghua Chen, Le Zhang, Zhiguang Cao, and Jing Guo. Distilling the knowledge from handcrafted features for human activity recognition. IEEE Transactions on Industrial Informatics, 14(10):4334–4342, 2018.

[12] Ervin Sejdić, Kristin A Lowry, Jennica Bellanca, Mark S Redfern, and Jennifer S Brach. A comprehensive assessment of gait accelerometry signals in time, frequency and time-frequency domains. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22(3):603–612,

2013.
[13] Mahmudul Hasan Abid and Abdullah-Al Nahid. Two unorthodox aspects in handcrafted-feature extraction for human activity recognition datasets. In 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), pages 1–4. IEEE, 2021.
[14] Allah Bux Sargano, Plamen Angelov, and Zulfiqar Habib. A comprehensive review on handcrafted and learning- based action representation approaches for human activity recognition. Applied Sciences, 7(1), 2017.
[15] Muhammad Attique Khan, Muhammad Sharif, Tallha Akram, Mudassar Raza, Tanzila Saba, and Amjad Rehman. Hand-crafted and deep convolutional neural network features fusion and selection strategy: An application to intelligent human action recognition. Applied Soft Computing, 87:105986, 2020.
[16] Mingtao Dong, Jindong Han, Yuan He, and Xiaojun Jing. Har-net: Fusing deep representation and hand-crafted features for human activity recognition. In Signal and Information Processing, Networking and Computers: Proceedings of the 5th International Conference on Signal and Information Processing, Networking and Computers (ICSINC), pages 32–40. Springer, 2019.
[17] Tazeem Haider, Muhammad Hassan Khan, and Muhammad Shahid Farid. An optimal feature selection method for human activity recognition using multimodal sensory data. Information, 15(10), 2024.
[18] Muhammad Hassan Khan, Muhammad Shahid Farid, and Marcin Grzegorzek. A comprehensive study on codebook-based feature fusion for gait recognition. Information Fusion, 92:216–230, 2023.
[19] Muhammad Hassan Khan, Muhammad Shahid Farid, and Marcin Grzegorzek. Using a generic model for codebook-based gait recognition algorithms. In 2018 International Workshop on Biometrics and Forensics (IWBF), pages 1–7. IEEE, 2018.
[20] Kimiaki Shirahama and Marcin Grzegorzek. On the generality of codebook approach for sensor-based human activity recognition. Electronics, 6(2), 2017.
[21] Muhammad Haseeb Arshad, Muhammad Bilal, and Abdullah Gani. Human activity recognition: Review, taxon- omy and open challenges. Sensors, 22(17), 2022.
[22] Frédéric Li, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek. Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors, 18(2), 2018.
[23] Sheeza Batool, Muhammah Hassan Khan, and Muhammah Shahid Farid. An ensemble deep learning model for human activity analysis using wearable sensory data. Applied Soft Computing, page 111599, 2024.
[24] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances. Sensors, 22(4), 2022.
[25] Morsheda Akter, Shafew Ansary, Md. Al-Masrur Khan, and Dongwan Kim. Human activity recognition using attention-mechanism-based deep learning feature combination. Sensors, 23(12), 2023.
[26] Md Milon Islam, Sheikh Nooruddin, Fakhri Karray, and Ghulam Muhammad. Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects. Computers in Biology and Medicine, 149:106060, 2022.
[27] Frédéric Li, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek. Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors, 18(2):679, 2018.
[28] Andreas Bulling, Ulf Blanke, and Bernt Schiele. A tutorial on human activity recognition using body-worn inertial sensors. 46(3), 2014.
[29] Chuanlin Zhang, Kai Cao, Limeng Lu, and Tao Deng. A multi-scale feature extraction fusion model for human activity recognition. Scientific Reports, 12(1):20620, 2022.
[30] Ivan Miguel Pires, Faisal Hussain, Nuno M. Garcia, Petre Lameski, and Eftim Zdravevski. Homogeneous data normalization and deep learning: A case study in human activity classification. Future Internet, 12(11), 2020.

[31] Tahmina Zebin, Matthew Sperrin, Niels Peek, and Alexander J Casson. Human activity recognition from inertial sensor time-series using batch normalized deep lstm recurrent networks. In 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pages 1–4. IEEE, 2018.

[32] Sujan Ray, Khaldoon Alshouiliy, and Dharma P Agrawal. Dimensionality reduction for human activity recognition using google colab. Information, 12(1):6, 2020.

[33] Mustafa Gokce Baydogan, George Runger, and Eugene Tuv. A bag-of-features framework to classify time series. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(11):2796–2802, 2013.

[34] Sanjay Hanji and Savita Hanji. Towards performance overview of mini batch k-means and k-means: case of four-wheeler market segmentation. In International Conference on Smart Trends in Computing and Communications, pages 801–813. Springer, 2023.

[35] Nan Xu. Applications of support vector machines in electromagnetic problems. PhD thesis, The University of New Mexico, 2011.

[36] Lyudmyla Kirichenko, Tamara Radivilova, and Vitalii Bulakh. Machine learning in classification time series with fractal properties. Data, 4(1):5, 2018.

[37] Himani Bhavsar and Mahesh H Panchal. A review on support vector machine for data classification. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(10):185–189, 2012.

[38] Isna Alfi Bustoni, I Hidayatulloh, AM Ningtyas, A Purwaningsih, and SN Azhari. Classification methods per- formance on human activity recognition. In Journal of Physics: Conference Series, volume 1456, page 012027. IOP Publishing, 2020.

[39] Lalita Mishra, Shekhar Verma, and Shirshu Varma. Brain tumor classification using feature extraction and non- linear svm hybrid model. In International Conference on Communication, Networks and Computing, pages 3–14. Springer, 2022.

[40] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Förster, Gerhard Tröster, Paul Lukow- icz, David Bannach, Gerald Pirkl, Alois Ferscha, et al. Collecting complex activity datasets in highly rich networked sensor environments. In 2010 Seventh international conference on networked sensing systems (INSS), pages 233–240. IEEE, 2010.

[41] Muhammad Adeel Nisar, Kimiaki Shirahama, Muhammad Tausif Irshad, Xinyu Huang, and Marcin Grzegorzek. A hierarchical multitask learning approach for the recognition of activities of daily living using data from wearable sensors. Sensors, 23(19):8234, 2023.

[42] Anna Ferrari, Daniela Micucci, Marco Mobilio, and Paolo Napoletano. Deep learning and model personalization in sensor-based human activity recognition. Journal of Reliable Intelligent Environments, 9(1):27–39, 2023.

[43] Azar Mahmoodzadeh. Human activity recognition based on deep belief network classifier and combination of local and global features. J. Inf. Syst. Telecommun, 9:33, 2021.

[44] Farhad Nazari, Darius Nahavandi, Navid Mohajer, and Abbas Khosravi. Comparison of deep learning techniques on human activity recognition using ankle inertial signals. In 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 2251–2256. IEEE, 2022.

[45] Chaolei Han, Lei Zhang, Yin Tang, Wenbo Huang, Fuhong Min, and Jun He. Human activity recognition using wearable sensors by heterogeneous convolutional neural networks. Expert Systems with Applications, 198:116764, 2022.