HUMAN ACTIVITY RECOGNITION USING MACHINE LEARNING

Mrs SARANYA P AP/CSE [1], BHARATH KUMAR V[2], HARIPRATHAP G[3], HARISH K[4], ANTONY R [5].

Professor [1], Assistant Professor, Department of CSE, The Kavery Engineering College , Tamil Nadu, India
Student [2,3,4,5], Department of CSEE, The Kavery Engineering College , Tamil Nadu, India

---***---

ABSTRACT: *Human Activity Recognition (HAR) has emerged as a significant area of research due to its wide range of applications in healthcare, smart homes, fitness tracking, and surveillance systems. Despite substantial progress, recognizing complex human activities remains a challenging problem because of variations in motion patterns and environmental conditions. This study proposes a machine learning-based approach for classifying human activities using an optimized feature selection method. The system is trained and evaluated on the UCI Human Activity Recognition dataset, which includes six categories of activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying. Experimental results demonstrate that the proposed approach achieves high classification performance, with accuracy exceeding 99% on the original dataset and maintaining strong performance after feature optimization. Various evaluation metrics such as accuracy, precision, F1-score, and AUC are used to assess the effectiveness of the models.*

Keywords: **machine learning, accuracy, precision, F1-score, and AUC.**

I.INTRODUCTION

Human Activity Recognition (HAR) is an important and rapidly growing area in the field of machine learning and data science. It focuses on identifying and classifying physical activities performed by individuals using data collected from sensors such as accelerometers and gyroscopes, commonly embedded in smartphones and wearable devices. With the increasing use of smart devices in everyday life, HAR systems have gained significant attention for their ability to automatically monitor and analyze human behavior in real time. The application of HAR spans across various domains, including healthcare, fitness tracking, smart homes, security, and human-computer interaction. In healthcare, HAR systems are used to monitor patients, detect abnormal activities, and assist elderly individuals in maintaining independent living. In fitness and sports, these systems help track physical activity levels and provide insights into user performance. Additionally, HAR plays a crucial role in developing intelligent environments where systems can adapt their behavior based on user activities. Despite its wide range of applications, accurately recognizing human activities remains a challenging task. This is mainly due to the complexity and variability of human movements, differences in individual behavior, and the presence of noise in sensor data. Activities such as sitting and standing may exhibit similar patterns, making classification difficult. Furthermore, environmental factors and sensor placement can also affect the quality and consistency of the collected data. Machine learning techniques have proven to be highly effective in addressing these challenges. By learning patterns from labeled datasets, supervised learning models can classify activities with high accuracy. Traditional machine learning algorithms such as Decision Trees, Support Vector Machines, and Random Forests are widely used for HAR tasks. These models rely on extracted features from sensor data to make predictions. However, the presence of irrelevant or redundant features can negatively impact model performance, making feature selection an essential step in the process.In this research, we focus on improving the performance of human activity recognition systems by applying machine learning techniques along with an effective feature selection approach. The study utilizes the UCI Human Activity Recognition dataset, which includes six common activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying. Various machine learning models are implemented and evaluated using performance metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC).

II. EXISTING SYSTEM

Human Activity Recognition (HAR) systems have been extensively studied over the past decade, with many existing approaches focusing on classifying human movements using sensor data from smartphones, wearable devices, and external monitoring systems. Traditional HAR systems primarily rely on predefined rules or classical machine learning techniques to identify activities such as walking, sitting, standing, and lying. These systems typically follow a pipeline that includes data collection, preprocessing, feature extraction, and classification. In earlier approaches, activity recognition was often performed using rule-based methods, where specific thresholds and conditions were defined manually to distinguish between different activities. While these methods are simple and easy to implement, they lack flexibility and fail to adapt to variations in user behavior or environmental conditions. As a result, their accuracy is generally limited, especially when dealing with complex or overlapping activities.With the advancement of machine learning, many existing systems have shifted toward supervised learning techniques. Algorithms such as Decision Trees, k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Naïve Bayes have been widely used for HAR tasks. These models are trained on labeled datasets to learn patterns in sensor data and classify activities accordingly. Although these approaches have improved accuracy compared to rule-based systems, they still face several challenges.

2.1DRAWBACKS:

Existing Human Activity Recognition (HAR) systems, although widely used, have several limitations that affect their overall performance and reliability. One of the major drawbacks is the heavy dependence on handcrafted feature extraction. Designing features manually requires domain expertise and may not always capture all the important patterns present in the sensor data. As a result, the model may miss critical information, leading to reduced accuracy.. Another significant issue is the difficulty in distinguishing between similar activities. Activities such as sitting and standing or walking upstairs and downstairs often produce very similar sensor signals. This overlap can confuse the model and result in misclassification, especially when simple machine learning algorithms are used. Existing systems are also sensitive to noise and variations in data. Factors such as sensor placement, device orientation, and user behavior can introduce inconsistencies in the data. These variations make it challenging for the model to generalize well across different users and environments, reducing its robustness in real-world applications. High computational complexity is another drawback, particularly when dealing with large datasets or high-dimensional feature spaces. Many systems include redundant or irrelevant features, which increase processing time and memory usage. This makes them less suitable for real-time applications or devices with limited resources.

III. PROPOSED FRAMEWORK

To overcome the limitations of existing Human Activity Recognition (HAR) systems, the proposed system introduces an efficient and robust machine learning framework combined with an effective feature selection approach. The main objective of this system is to improve classification accuracy while reducing computational complexity, making it suitable for real-time and practical applications. The proposed system begins with data acquisition from the UCI Human Activity Recognition dataset, which contains sensor signals collected from smartphones equipped with accelerometers and gyroscopes. These sensors capture motion-related information that reflects different human activities such as walking, walking upstairs, walking downstairs, sitting, standing, and lying. The collected data is first subjected to preprocessing steps, including noise removal, normalization, and handling of missing values, to ensure data quality and consistency. A key component of the proposed system is the feature selection process. Instead of using all available features, the system identifies and selects only the most relevant and informative features that contribute significantly to activity classification. This reduces redundancy, lowers computational cost, and improves model performance. By eliminating unnecessary features, the system becomes more efficient without sacrificing accuracy.The proposed system also incorporates performance evaluation using metrics such as accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). A confusion matrix is used to analyze classification results in detail and identify any misclassifications between similar activities. This helps in understanding the strengths and weaknesses of the model.

3.1BENEFITS

The proposed Human Activity Recognition (HAR) system offers several advantages over traditional approaches, making it more efficient, accurate, and suitable for real-world applications. One of the key benefits is its high classification accuracy. By applying machine learning algorithms along with an effective feature selection technique, the system is able to correctly identify human activities with very high precision, even when dealing with complex datasets. Another important benefit is reduced computational complexity. The use of feature selection helps in eliminating irrelevant and redundant data, which decreases the number of features used for training the model. This not only speeds up the processing time but also reduces memory usage, making the system more suitable for real-time applications and devices with limited resources such as smartphones and wearable devices. The system also demonstrates improved efficiency and performance. By focusing only on the most relevant features, the models can learn faster and make predictions more quickly. This leads to better overall system performance without compromising accuracy. Additionally, it helps in reducing the risk of overfitting, ensuring that the model performs well on both training and unseen data. Another advantage is its robustness and adaptability. The proposed system can handle variations in sensor data caused by differences in user behavior, device orientation, and environmental conditions. This makes it more reliable in real-world scenarios compared to existing systems. The system is also scalable and flexible, allowing it to be extended for additional activities or integrated with other technologies. It can be applied in various domains such as healthcare monitoring, fitness tracking, smart homes, and security systems. For example, it can be used to monitor patient activities, detect falls, or track physical exercise routines. Furthermore, the use of multiple machine learning models allows for better comparison and selection of the best-performing algorithm. This ensures that the system achieves optimal results based on the given dataset and requirements.

IV. SYSTEM IMPLEMENTATION

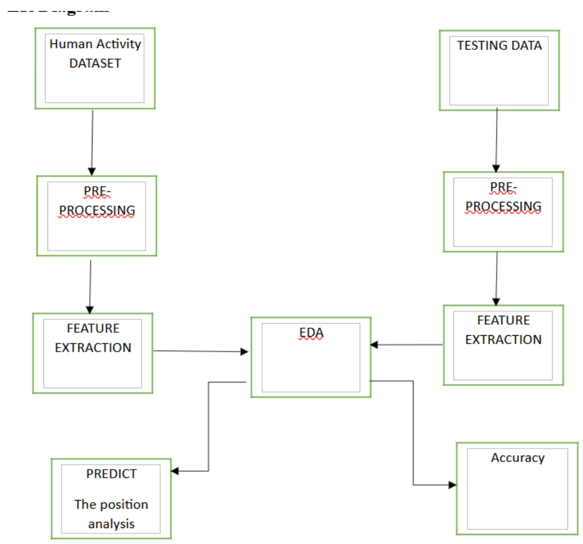

Figure 1: Block Diagram of the IOT BASED INDUSTRY PROTECTION SYSTEM

The diagram represents the workflow of the proposed Human Activity Recognition (HAR) system using machine learning. It shows how both training and testing data are processed to predict human activities and evaluate performance. The process begins with the Human Activity Dataset, which contains sensor data such as accelerometer and gyroscope readings. This dataset is passed through a pre-processing stage where noise, missing values, and inconsistencies are removed. Data normalization and cleaning are performed here to improve data quality. After preprocessing, the cleaned data moves to the feature extraction stage. In this step, important features are derived from raw sensor signals. These features may include statistical measures (mean, variance), frequency-based attributes, and other relevant patterns that help distinguish between activities.

The extracted features are then given to the EDA (Exploratory Data Analysis) block. This is a key component of the system where data is analyzed to understand patterns, relationships, and distributions. It also plays a role in selecting the most relevant features for improving model performance. On the right side of the diagram, testing data follows a similar path: preprocessing and feature extraction. This ensures that both training and testing data are treated consistently before being evaluated by the model.

From the EDA block, two outputs are generated:

- The first output goes to the prediction stage, where the trained model classifies the activity (such as walking, sitting, or standing). This is labeled as "position analysis," meaning the system identifies the type of human movement.
- The second output goes to the accuracy evaluation block, where the model's performance is measured. Metrics like accuracy are calculated to determine how well the system is performing.

Overall, the diagram illustrates a complete pipeline where data is processed, features are extracted and analyzed, and machine learning models are used to predict human activities while also evaluating their accuracy.

V.CONCLUSION

In this work, a machine learning-based approach for Human Activity Recognition (HAR) has been presented to accurately classify different human activities using sensor data. The system utilizes the UCI HAR dataset and follows a structured pipeline that includes data preprocessing, feature extraction, feature selection, and classification using multiple machine learning algorithms. The results demonstrate that the proposed system achieves high accuracy in recognizing activities such as walking, sitting, standing, and lying. The incorporation of an effective feature selection method significantly reduces computational complexity while maintaining strong performance. This highlights the importance of selecting relevant features to improve both efficiency and model reliability. Furthermore, the use of multiple machine learning models allows for better comparison and selection of the most suitable algorithm, with ensemble methods showing superior performance in handling complex patterns. The evaluation using metrics such as accuracy, precision, recall, F1-score, and confusion matrix confirms the robustness of the system.

Overall, the proposed approach successfully addresses the limitations of existing systems by improving accuracy, reducing processing time, and enhancing adaptability. It can be effectively applied in real-world applications such as healthcare monitoring, fitness tracking, and smart environments. Future improvements can focus on integrating deep learning techniques and real-time implementation to further enhance system performance..

VI.FUTURE WORK

Although the proposed Human Activity Recognition (HAR) system achieves high accuracy and efficiency, there are several areas where further improvements can be made to enhance its performance and applicability. One important direction for future work is the integration of deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models can automatically learn complex patterns from raw sensor data without the need for manual feature extraction, potentially improving accuracy for more complex activities. Another area of improvement is the use of real-time activity recognition systems. Implementing the model in real-time environments using smartphones or wearable devices would make the system more practical for applications like healthcare monitoring, fitness tracking, and smart homes. Optimizing the model for low-power and resource-constrained devices will be essential for such implementations. Expanding the dataset is also a key aspect of future work. Including more diverse activities, different age groups, and varied environmental conditions can help improve the robustness and generalization of the system. This will allow the model to perform better across different users and real-world scenarios.

REFERENCES

[1] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition Using Smartphones," Proceedings of the 21st European Symposium on Artificial Neural Networks (ESANN), pp. 437–442, 2013.

[2] 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, Dec. 2010.

[3] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition Using Wearable Sensors," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192–1209, 2013.

[4] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data," International Conference on Pervasive Computing, pp. 1–17, 2004.

[5] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed., New York, NY, USA: Springer, 2009.

[6] C. M. Bishop, Pattern Recognition and Machine Learning, New York, NY, USA: Springer, 2006.

[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[8] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.

[9] S. R. Safavian and D. Landgrebe, "A Survey of Decision Tree Classifier Methodology," IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660–674, 1991.

[10] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.