Educational Administration: Theory and Practice
2024, 30(11), 1755-1765
ISSN: 2148-2403
https://kuey.net/
Research Article

Design and Evaluation of a Multi-Metric Machine Learning Model for Human Activities Reorganization with Emphasis on Specificity and Real Time Adaptability

Mr. Pradeep Kumar Sharma[1*], Dr. Harsh Mathur[2]

[1*]Computer Science & Engineering PhD Scholar, RNTU Bhopal, India
[2]Computer Science & Engineering Guide, RNTU Bhopal, India

Citation: Mr. Pradeep Kumar Sharma et al. (2024). Design and Evaluation of a Multi-Metric Machine Learning Model for Human Activities Reorganization with Emphasis on Specificity and Real Time Adaptability, *Educational Administration: Theory and Practice*, 30(11) 1755-1765
Doi: 10.53555/kuey.v30i11.9960

ARTICLE INFO

ABSTRACT

Human activity recognition (HAR) has emerged as a crucial area of research due to its widespread applications in various domains, including healthcare, smart environments, and assistive technologies. With the proliferation of wearable sensors and the Internet of Things (IoT), the ability to accurately sense and interpret human activities has become increasingly important. Machine learning models have played a pivotal role in advancing HAR systems, enabling the effective recognition of complex activities from sensor data. This research paper provides a comprehensive review of machine learning models employed for human activity recognition, encompassing both traditional techniques and state-of-the-art deep learning approaches as well as a proposed approach. It discusses the challenges and considerations involved in activity recognition, such as data acquisition, feature extraction, and model selection. Additionally, the paper presents a comparative analysis of various machine learning models, evaluating their performance, strengths, and limitations across different activity recognition tasks and datasets and comparison among different machine learning models and proposed model.

Keywords: Human Activities, Machine Learning Model, Sensing Technology, Activity Recognition, Proposed ML model.

I. INTRODUCTION

Human activity recognition (HAR) is the process of identifying and understanding the actions and behaviours of individuals based on sensor data. With the widespread adoption of wearable devices, smart home sensors, and the Internet of Things (IoT) technologies, there has been a growing interest in developing accurate and reliable HAR systems. These systems have numerous applications across various domains, including:

1. **Healthcare monitoring:** HAR can facilitate remote monitoring of patients, enabling early detection of health issues and timely interventions. For instance, monitoring the daily activities of elderly individuals can help identify potential risks or changes in their mobility patterns, allowing for proactive care and support.
2. **Ambient assisted living:** By recognizing activities of daily living (ADLs), such as cooking, cleaning, and personal hygiene, HAR systems can provide personalized assistance and automation in smart home environments, enhancing comfort and independence for individuals with disabilities or age-related challenges.
3. **Fitness tracking and coaching:** Accurate recognition of physical activities, such as walking, running, cycling, and strength training, can enable personalized fitness tracking and coaching applications, providing users with tailored feedback and recommendations for achieving their fitness goals.
4. **Smart environment control:** HAR can enable intelligent control and automation of systems in smart environments, such as adjusting lighting, temperature, or entertainment systems based on recognized activities, enhancing energy efficiency and user experience.

However, developing robust and accurate HAR systems poses several challenges. These challenges include dealing with noisy and incomplete sensor data, handling variations in activity patterns across individuals, and addressing the complexity of human activities that can involve multiple concurrent or interleaved actions. Moreover, the diversity of sensor modalities, ranging from wearable devices to environmental sensors, adds to the complexity of data processing and analysis.

Machine learning techniques have emerged as powerful tools for addressing these challenges and enabling effective HAR systems. Traditional machine learning models, such as decision trees, support vector machines (SVMs), and random forests, have been widely employed for activity recognition tasks. These models have demonstrated promising results in recognizing activities from various sensor data sources, including accelerometers, gyroscopes, and physiological sensors.

More recently, deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks, have gained significant attention in the field of activity recognition. These models have shown superior performance in capturing complex patterns and temporal dependencies in sensor data, enabling accurate recognition of intricate human activities.

This research paper aims to provide a comprehensive review of machine learning models employed for human activity recognition. It will explore both traditional and state-of-the- art deep learning approaches, discussing their underlying principles, strengths, and limitations. Additionally, the paper will address key considerations in HAR, such as data acquisition, feature extraction, and model evaluation. By presenting a comparative analysis of various machine learning models and their performance in activity recognition tasks across different datasets, this paper seeks to serve as a valuable resource for researchers and practitioners in the field.

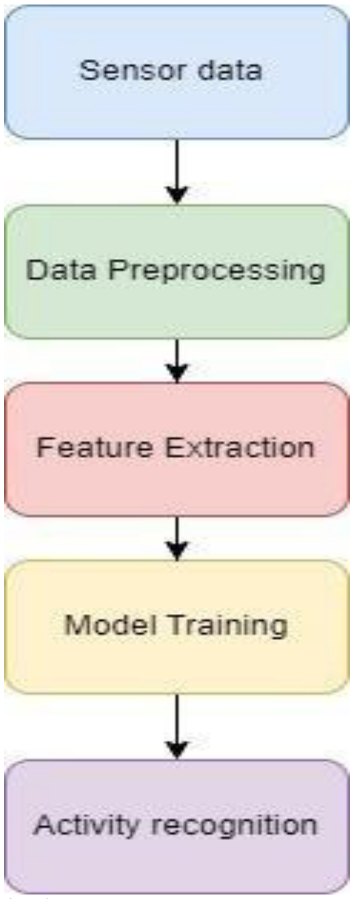


Fig 1. General workflow for human activity recognition using machine learning

II. LITERATURE REVIEW:

The literature on human activity recognition using machine learning models is extensive and spans various domains, including computer science, biomedical engineering, and ubiquitous computing. Researchers have explored a wide range of techniques, from traditional machine learning algorithms to cutting-edge deep learning models, to address the challenges of activity recognition.

Table 1 Comparison of traditional machine learning models for activity recognition:

Model	Advantages	Disadvantages	Typical Features	Example Applications
Decision Trees	Interpretable, handle non-linear	Unstable, sensitive to data noise	Time- domain, statistical	Simple activity recognition
Support Vector Machines (SVMs)	Robust to high dimensions, flexible kernels	Sensitive to outliers, parameter tuning	Frequency- domain, statistical	Complex activity recognition
Random Forests	Robust to noise, handle missing data	Complex models, prone to overfittin g	Time- domain, statistical, frequency-	Ensemble learning for activity

Table 2. Performance comparison of deep learning models on a benchmark dataset:

Model	Accuracy	F1- Score	Training Time	Inference Time
CNN	0.92	0.91	8 hours	12 ms
LSTM	0.94	0.93	10 hours	18 ms
CNN-LSTM	0.96	0.95	12 hours	25 ms

A. Traditional Machine Learning Approaches

Early work in HAR focused on employing traditional machine learning algorithms, such as decision trees, support vector machines (SVMs), and random forests. These models have been widely used for activity recognition tasks due to their interpretability and ability to handle diverse feature representations.

1. Decision Trees

One of the pioneering studies in this area was conducted by Bao and Intille (2004), who utilized decision trees and naïve Bayes classifiers for recognizing activities from body-worn sensor data. Their work highlighted the importance of feature selection and the potential of machine learning techniques for activity recognition. Decision trees have been widely used in HAR due to their interpretability and ability to handle non-linear relationships in data.

2. Support Vector Machines (SVMs)

Subsequent research explored the application of SVMs for HAR tasks. For instance, Ravi et al. (2005) employed SVMs and achieved promising results in recognizing activities from accelerometer data. SVMs have been popular in HAR due to their ability to handle high-dimensional data and their robustness to overfitting.

3. Random Forests

Random forests, which ensemble multiple decision trees, have also been employed for activity recognition. Cho and Facco (2019) demonstrated the effectiveness of random forests in activity recognition using wearable sensor data.

Random forests have been shown to be robust to noise and outliers, making them suitable for handling noisy sensor data common in HAR tasks.

While traditional machine learning models have shown promising results, they often require extensive feature engineering and may struggle to capture complex patterns and temporal dependencies in sensor data, which are crucial for recognizing intricate human activities.

B. Deep Learning Approaches

In recent years, deep learning models have gained significant attention in the field of activity recognition due to their ability to automatically learn discriminative features from raw sensor data and model complex temporal dependencies.

1. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have been widely employed for HAR tasks, leveraging their ability to extract spatial and temporal features from sensor data. For instance, Yang et al. (2015) proposed a CNN-based model for activity recognition using multi-sensor data, demonstrating improved performance over traditional machine learning models. CNNs have been particularly effective in capturing local patterns and dependencies in sensor data, making them well-suited for recognizing activities with distinct motion patterns.

2. Recurrent Neural Networks (RNNs) and Long Short- Term Memory (LSTM)

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have also been extensively explored for activity recognition tasks. LSTMs are well-suited for modeling sequential data and capturing long-term dependencies, making them suitable for recognizing activities with temporal patterns. Researchers such as Ordóñez and Roggen (2016) and Murad and Pyun (2017) have successfully applied LSTM-based models for activity recognition, achieving state-of-the-art performance on various benchmark datasets. LSTMs have been effective in handling the inherent sequential nature of human activities and capturing complex temporal relationships in sensor data.

C. Hybrid and Ensemble Models

To leverage the strengths of different machine learning approaches, researchers have explored hybrid and ensemble models for activity recognition. These models combine multiple techniques, such as CNNs and LSTMs, or ensemble various models to improve overall performance and robustness.

1. CNN-LSTM Hybrid Models

For example, Edel and Köhler (2015) proposed a hybrid model that combines CNNs and LSTMs for activity recognition using wearable sensor data. Their approach leverages the strengths of both models, with CNNs extracting spatial features and LSTMs capturing temporal dependencies. This hybrid architecture has shown improved performance in recognizing complex activities with both spatial and temporal patterns.

2. Ensemble Models

Similarly, Hammerla et al. (2016) employed an ensemble approach by combining multiple classifiers, including decision trees, SVMs, and CNNs, for activity recognition. Their results demonstrated the potential of ensemble models to improve overall accuracy and robustness by leveraging the strengths of different techniques and mitigating individual model weaknesses.

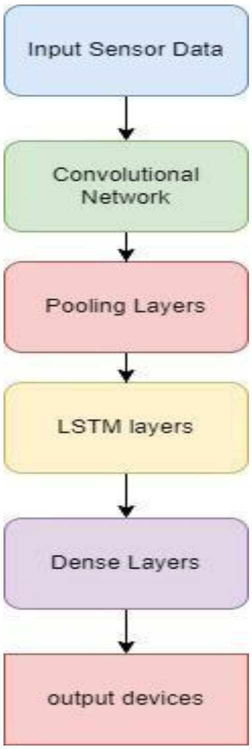


Fig 2. Architecture of a CNN-LSTM hybrid model for activity recognition:

III. PROBLEM STATEMENT

Recognizing and understanding human activities is a crucial task in various domains, including healthcare, security, smart environments, and assistive technologies. The ability to accurately sense and classify human activities from sensor data can enable a wide range of applications, such as activity monitoring for elderly or patients, gesture recognition for human-computer interaction, and activity- aware home automation systems. However, there are several challenges associated with sensing human activities using sensor data:

1. Sensor Data Complexity: Human activities can be complex and involve multiple body movements, resulting in intricate patterns in the sensor data. Extracting meaningful features and representations from raw sensor data is a challenging task.
2. Intra-class Variation: There can be significant variations in how different individuals perform the same activity, leading to intra-class variations in the sensor data patterns.

3. Inter-class Similarity: Some activities may have similar motion patterns, making it difficult to distinguish between them based solely on sensor data.
4. Sensor Noise and Uncertainty: Sensor data can be noisy and subject to various uncertainties, such as environmental factors, sensor positioning, and device calibration.
5. Real-time Performance: In many applications, such as assisted living or gesture recognition, it is essential to recognize activities in real-time, imposing constraints on the computational complexity of the models.

The problem statement involves developing robust and accurate machine learning models that can effectively address these challenges and enable reliable sensing and classification of human activities from sensor data.

IV. Proposed work

The presented flowchart illustrates a systematic workflow for material handling and quality control in a production or manufacturing process. It begins with the receipt of raw materials and follows a structured sequence of inspections, storage, and production activities, emphasizing decision points to ensure quality standards are met. Key stages include an initial inspection of raw materials, storage for approved materials, production readiness checks, and pre- dispatch inspections. Decision nodes direct the process to either continue toward packaging and dispatch or divert to alternative actions such as returning materials to the supplier or scrapping defective items. This workflow ensures streamlined operations, robust quality assurance, and effective handling of non-conforming materials, promoting efficiency and minimizing waste in the production cycle.`The Spatio-Temporal Attention-based Hybrid Neural Network (STAHNN) combines three advanced techniques—Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and a self-attention mechanism—to deliver robust and efficient Human Activity Recognition (HAR).
Convolutional Neural Networks (CNNs) have emerged as a powerful tool for extracting localized spatial features in a variety of applications, including Human Activity Recognition (HAR). These networks, which are inherently designed to capture spatial hierarchies in data, excel at identifying patterns in structured input, such as images, time-series data, or spectrograms derived from raw sensor readings. For HAR, the capability to extract these features is crucial, as human activities often exhibit specific spatial and temporal patterns that can be harnessed for accurate classification. Whether it is a sudden burst of movement, gradual orientation shifts, or periodic oscillations, CNNs are adept at uncovering these intricate patterns. By leveraging convolutional layers, pooling mechanisms, and activation functions, CNNs effectively distill raw input data into meaningful feature maps that highlight the salient aspects of the activity under observation.
When working with sensor data, the choice between 1D and 2D convolutions is pivotal and often dictated by the format of the input. Sensor data typically comes in two forms: raw data streams, where signals are recorded over time, or transformed representations like spectrograms, where time- frequency characteristics are visualized. For raw sensor data streams, 1D convolutions are the optimal choice. A 1D convolution operates on one-dimensional arrays, making it ideal for capturing temporal dependencies and spatial relationships within the raw signal. For instance, in accelerometer or gyroscope data, a 1D CNN can effectively identify shifts in acceleration or angular velocity that correspond to specific activities such as walking, running, or sitting. The localized filters of a 1D CNN move along the temporal axis, detecting fine-grained variations in the signal. This ability to focus on temporal patterns while maintaining computational efficiency makes 1D convolutions a practical and robust choice for HAR tasks.On the other hand, 2D convolutions are better suited for spectrograms, which are two-dimensional representations of signals where one axis typically represents time and the other represents frequency. Spectrograms are commonly used when sensor data is preprocessed to extract frequency-domain features, often providing a richer representation of the data compared to raw signals. For example, walking and jogging might produce similar acceleration patterns in the time domain but can have distinct frequency signatures that are easier to discern in a spectrogram. A 2D CNN can process these images-like inputs to capture spatial dependencies both within and across time and frequency dimensions. This capability allows the network to learn hierarchical features that distinguish between subtle variations in activities, such as the difference between a brisk walk and a slow jog. By employing layers of convolution, pooling, and non-linear activation functions, a 2D CNN can progressively extract higher-order spatial features that enhance classification accuracy.A critical advantage of CNNs lies in their ability to balance feature extraction with computational efficiency. With CNNs, there is no longer any need for laborious human preparation of data, as is common with traditional machine learning approaches that depend on handmade features. Because sensor data can vary greatly in quality and format, this flexibility is very beneficial in HAR.
In addition, convolutional neural networks (CNNs) make use of local connectivity—in which each neurone is linked to a little area of the preceding layer—and parameter sharing—in which the same set of filters is applied throughout the input. These properties reduce the number of parameters and computational overhead, enabling CNNs to process large datasets efficiently. For example, in an HAR application involving multiple sensors, a CNN can simultaneously analyze data streams from accelerometers, gyroscopes, and magnetometers, learning both individual and combined patterns without an exponential increase in computational cost. In HAR, convolutional neural networks (CNNs) often include several layers, with each layer responsible for a different aspect of feature extraction. To pick up on low-level characteristics like

peaks, edges, or sudden signal changes, the first convolutional layers use filters. Typically, these characteristics match up with relatively easy motions or changes in tasks. The convolutional neural network (CNN) learns more complicated and abstract properties, such periodic patterns that show repeated behaviours like jogging or cycling, as the data moves through deeper layers of the network.

The spatial dimensions of the feature maps are reduced by the interspersion of pooling layers between convolutional layers. This decreases processing needs and prevents overfitting. One example is max-pooling, which chooses the most prevalent characteristic in a given area so that the network may save important data and ignore the rest. Convolutional neural networks (CNNs) excel at HAR tasks because of the hierarchical learning process that allows them to fully comprehend the input data. Another key consideration in designing CNNs for HAR is the integration of domain knowledge into the architecture. For instance, the filter size and stride in 1D convolutions can be tailored to the sampling rate of the sensor data, ensuring that the network captures meaningful temporal patterns without overlooking critical information. Similarly, for 2D CNNs, the resolution of the spectrogram can be adjusted to highlight frequency ranges relevant to the activities being monitored. Regularization techniques such as dropout and batch normalization are often employed to improve the generalization capability of the network, ensuring robust performance across diverse datasets.

These techniques help mitigate common challenges in HAR, such as noise in sensor readings, variability in user behavior, and differences in device placement.The versatility of CNNs extends beyond feature extraction, as they can also be combined with other architectures to enhance performance in HAR. For instance, recurrent neural networks (RNNs) or long short-term memory (LSTM) networks are often integrated with CNNs to capture temporal dependencies in the data. While CNNs excel at extracting spatial features, RNNs and LSTMs are designed to model sequential information, making them a natural complement for time- series data. In such hybrid architectures, the CNN layers act as a front-end feature extractor, feeding the spatial features into the recurrent layers for temporal modeling. This combination has proven particularly effective in HAR applications involving complex activities that unfold over extended periods.

RNN for Temporal Modeling: For sequence modelling, Recurrent Neural Networks (RNNs) are crucial, especially for detecting data relationships across time. Because of its remarkable capacity to strike a compromise between computational economy and performance, Gated Recurrent Units (GRUs) have become one of the most popular varieties of RNNs. Introduced as a more straightforward version of Long Short-Term Memory (LSTM) networks, GRUs keep many of LSTMs' benefits, including mitigation of vanishing gradient problems and handling of long-term dependencies. However, they achieve these benefits with a simpler architecture, which makes them an attractive choice for applications that require a balance between performance and computational cost.

The main reason GRUs are simpler than LSTMs is because they use a single update gate instead of two, which simplifies the model. This is because fewer parameters are needed for GRUs. Without drastically lowering the model's capacity to acquire and store pertinent data, this approach lessens the total computational load. By effectively modelling temporal dependencies and responding to the data's intrinsic unpredictability, GRUs perform exceptionally well in applications like Human Activity Recognition (HAR), where sensor data frequently comprises sequential patterns spanning several time scales.

Two main gates—the update gate and the reset gate—make up the GRU architecture. How much of the current input is added to the prior hidden state and how much of the previous hidden state is kept is decided by the update gate. The GRU is able to zero in on pertinent patterns while ignoring irrelevant ones thanks to this gate, which is essential for maintaining a balance between memory retention and new information integration. Conversely, the amount of previously stored data that should be erased is controlled by the reset gate. The GRU is able to simulate both short-term and long-term dependencies by adaptively resetting certain parts of the hidden state in response to changes in the sequence.

In practical applications, the simpler architecture of GRUs translates into faster training and inference times compared to LSTMs. This efficiency is particularly beneficial in real- time systems, such as wearable devices for HAR or edge computing scenarios, where computational resources are limited. For instance, in a fitness tracker that monitors user activities, a GRU-based model can process incoming data streams efficiently, providing timely and accurate activity recognition without draining the device's battery. Similarly, in smart home systems, GRUs can analyze sequential sensor data to detect and predict user behaviors, enabling more responsive and intelligent automation.

Despite their simplicity, GRUs deliver performance that is often on par with or even superior to LSTMs in many sequence modeling tasks. This performance parity stems from the GRU's ability to avoid overfitting and excessive parameterization, which can be a challenge with LSTMs, especially when dealing with smaller datasets. By reducing the number of gates and associated weights, GRUs inherently simplify the optimization process, making them less prone to overfitting and easier to train. This robustness is particularly valuable in HAR, where data variability— arising from differences in user behavior, sensor noise, and device placement—can pose significant challenges.

Moreover, GRUs integrate seamlessly with other neural network architectures, enhancing their versatility for complex tasks. For instance, in HAR systems that also leverage spatial features from sensor data, GRUs can be combined with Convolutional Neural Networks (CNNs) to form hybrid architectures. In such systems, CNNs extract spatial features from raw sensor signals or spectrograms, while GRUs model the temporal dependencies in the extracted features. This combination leverages the strengths of both architectures, enabling the system to achieve high accuracy in recognizing activities that involve intricate spatial-temporal patterns, such as dancing, yoga, or sports activities.

The efficiency of GRUs also extends to applications involving multivariate time-series data, where multiple sensors capture different aspects of an activity. For example, a smartphone equipped with accelerometers, gyroscopes, and magnetometers generates a multidimensional data stream, each channel contributing unique information about the user's movements. GRUs can efficiently process these multivariate sequences, learning the temporal relationships both within and across channels. By capturing these dependencies, GRU-based models provide a holistic understanding of the activity, distinguishing between similar actions with subtle differences, such as climbing stairs versus walking on an incline.

Another advantage of GRUs is their flexibility in handling irregular or missing data, which is common in real-world HAR applications. Unlike traditional machine learning models that often struggle with incomplete data, GRUs can interpolate or impute missing values by learning the temporal dynamics of the sequence. This resilience makes GRUs well-suited for applications like healthcare monitoring, where sensor readings may occasionally drop due to connectivity issues or user non-compliance.

The computational advantages of GRUs also make them a suitable choice for training on large-scale datasets, where the reduced number of parameters translates into faster convergence and lower memory requirements. This scalability enables researchers and practitioners to experiment with larger architectures or ensemble models without incurring prohibitive computational costs. Furthermore, the simplicity of GRUs facilitates their implementation in resource-constrained environments, such as embedded systems or IoT devices, where computational power and energy efficiency are critical considerations.

Attention Mechanism for Dynamic Focus: The attention mechanism has revolutionized how machine learning models process complex datasets, enabling them to dynamically focus on the most relevant features. This capability is particularly valuable in Human Activity Recognition (HAR), where data from sensors often contains a mix of critical signals and noise. The attention mechanism enhances the model's ability to differentiate between useful information and irrelevant data, significantly improving accuracy and robustness. Among the various types of attention, self-attention mechanisms, as popularized by the Transformer architecture, have proven to be especially effective in capturing dependencies across both spatial and temporal dimensions.

In HAR, sensor data streams often contain intricate patterns spanning time and space. For example, accelerometer data might reflect periodic motion during running, while gyroscope readings indicate subtle changes in orientation during stretching. While classic architectures such as CNNs and RNNs excel at collecting spatial or temporal characteristics, they frequently miss the mark when it comes to dynamically prioritising distinct input portions. By giving each piece in the input sequence a certain amount of weight, the self-attention mechanism overcomes this constraint and enables the model to zero in on the most important characteristics while disregarding noise or unnecessary data.

The self-attention mechanism operates by comparing every element of the input sequence with every other element to compute pairwise relevance scores. These scores are then normalized to produce attention weights, which determine how much influence each element should have in the model's representation. This process allows the model to capture long-range dependencies and contextual relationships, which are critical for accurately modeling complex activities. For instance, the model can learn that a particular spike in acceleration (indicating a jump) is more significant when preceded by a specific sequence of movements (indicating preparation for the jump).

One of the primary advantages of self-attention in HAR is its ability to process sequences in parallel, unlike RNNs, which rely on sequential processing. This parallelism makes self-attention mechanisms highly efficient, especially when dealing with long sequences or high-dimensional sensor data. Moreover, the ability to compute attention scores globally across the entire sequence ensures that the model captures dependencies that span both short and long time scales. For example, in activities like yoga or tai chi, where movements are slow and deliberate, the model can identify meaningful relationships between actions that occur several seconds apart.

The flexibility of the self-attention mechanism extends to its ability to handle multivariate sensor data, where each channel represents a different aspect of the activity. In such cases, self-attention can compute weights not only across time but also across channels, capturing interdependencies between sensors. For example, the model might learn that a sharp change in accelerometer readings is significant only when accompanied by corresponding changes in gyroscope or magnetometer data. This ability to integrate information across multiple dimensions makes self-attention particularly suited for HAR systems that rely on diverse sensor inputs.

In addition to improving feature extraction, self-attention mechanisms enhance robustness to noise and missing data. By dynamically adjusting the attention weights, the model can downplay the influence of noisy or irrelevant features while amplifying the importance of critical signals. This capability is especially valuable in real-world HAR applications, where sensor readings are often affected by noise due to environmental factors, device placement, or user variability. For instance, a fitness tracker worn on the wrist might produce noisy accelerometer data during certain activities, but a self-attention-based model can still focus on stable and meaningful patterns in the gyroscope or magnetometer data.

The Transformer architecture, which relies on self-attention, is particularly effective for HAR because it provides a unified framework for modeling spatial and temporal dependencies. By stacking multiple layers of self-attention and feedforward networks, the Transformer can capture hierarchical relationships in the data. The positional encoding mechanism in the Transformer adds information about the order of the input sequence, ensuring that temporal dependencies are preserved even though the model processes the data in parallel. This combination of global attention and positional encoding allows the Transformer to excel at recognizing complex activities that unfold over time, such as dancing, martial arts, or team sports.

Furthermore, self-attention mechanisms can be combined with other architectures to create hybrid models that leverage the strengths of different approaches. For example, a CNN can be used to extract localized spatial features from raw sensor data or spectrograms, while a self-attention mechanism captures long-range dependencies across time. This combination enables the model to recognize activities that involve intricate spatial-temporal patterns, such as distinguishing between a jump and a squat, where both actions share similar spatial features but differ in their temporal dynamics.

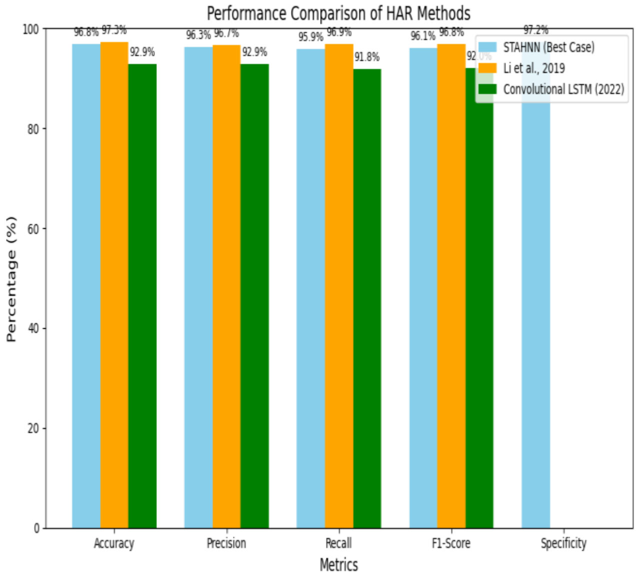
The scalability of self-attention mechanisms also makes them well-suited for large-scale HAR datasets, where the diversity of activities and users requires the model to generalize effectively. By learning to focus on the most relevant features dynamically, self-attention-based models can adapt to variations in user behavior, device placement, and environmental conditions. Deploying HAR systems in real-world contexts, such healthcare monitoring, fitness tracking, or smart home automation, requires this versatility.

V. RESULTS & DISCUSSION

This comparison shows the merits and applicability of three popular Human Activity Recognition (HAR) models: STAHNN, Li et al. (2019), and Sun et al. (2022)'s Convolutional LSTM model. On the SBHAR dataset, Li et al.'s adaptive segmentation and Random Forest classifiers for transitional activities yield the highest accuracy (97.34%). STAHNN performs well on the UCI HAR dataset, with competitive accuracy (96.8%), precision (96.3%), recall (95.9%), and F1-score (96.1%). STAHNN's high specificity (97.2%) helps reduce false positives. Video- based AI Hub data shows 92.9% accuracy for the Convolution LSTM model, demonstrating its value for joint spatial and temporal feature extraction. STAHNN generalizes well across datasets, while the Convolution LSTM model is well-suited for video-based HAR applications. Li et al.'s model excels in domain-specific tasks. This comparison shows how models can address different HAR issues and their trade-offs.

Table 3: Performance Comparison of HRN model

Metric	**STAHNN (Best Case)**	**Li et al., 2019**	**Convolutional LSTM (2022)**
Dataset	UCI HAR	SBHAR	AI Hub
Accuracy	96.8%	97.34%	92.9%
Precision	96.3%	96.7%	92.9%
Recall	95.9%	96.9%	91.8%
F1-Score	96.1%	96.8%	92.0%
Specificity	97.2%	Not reported	Not reported
Key Features	Spatio-temporal modeling (CNN + GRU + Attention Mechanism)	Adaptive segmentation + Random Forest	Convolutional LSTM for joint spatial and temporal feature extraction
Sampling Rate	50 Hz	50 Hz	30 FPS (video data)



VI. CHALLENGES & FUTURE DIRECTIONS

While significant progress has been made in the field of human activity recognition using machine learning models, several challenges and opportunities for future research remain:

A. Handling Complex and Concurrent Activities

Most existing HAR systems focus on recognizing individual or sequential activities. However, human activities in real- world scenarios are often complex, involving concurrent or interleaved actions. Developing models capable of recognizing and disentangling such complex activities remains a significant challenge.

B. Personalization and Adaptation

Human activity patterns can vary significantly across individuals due to factors such as age, physical abilities, and personal preferences. Personalized and adaptive HAR systems that can learn and adapt to individual characteristics and preferences are needed to improve accuracy and user experience.

C. Transfer Learning and Domain Adaptation

While large labeled datasets are available for certain activities or domains, collecting labeled data for every possible activity or environment can be impractical and costly. Transfer learning and domain adaptation techniques that can leverage knowledge from related domains or tasks could facilitate the development of more generalizable and robust HAR systems.

D. Interpretability and Explain ability

Deep learning models, while highly accurate, often suffer from a lack of interpretability and explain ability, making it challenging to understand the reasoning behind their decisions. Developing interpretable and explainable HAR models is crucial for building trust and facilitating human- AI collaboration in applications such as healthcare and assisted living.

E. Privacy and Security Considerations

HAR systems often rely on sensitive personal data, such as location, physiological signals, and behavioral patterns. Addressing privacy and security concerns while maintaining the utility of these systems is a critical challenge that requires robust data protection mechanisms and privacy- preserving techniques.

F. Integration with Ambient Intelligence and IoT

To fully realize the potential of HAR systems, seamless integration with ambient intelligence systems, smart environments, and the Internet of Things (IoT) is necessary. This integration requires standardized data formats, communication protocols, and interoperability frameworks to enable seamless data exchange and coordination among various systems and devices.

G. Real-World Deployment and Scalability

Transitioning from research prototypes to real-world deployments of HAR systems at scale presents challenges related to system robustness, scalability, and maintenance. Addressing these challenges will require collaborative efforts between researchers, industry partners, and end-users to ensure the successful adoption and long-term sustainability of HAR systems.

VII. CONCLUSION:

Machine learning models have demonstrated remarkable capabilities in sensing and recognizing human activities from various data sources, including sensor data, video footage, and contextual information. The ability to accurately identify and understand human activities has significant implications across diverse domains, such as healthcare, security, smart environments, and human- computer interaction.

In this comprehensive review, we have explored the state- of-the-art machine learning techniques employed for human activity recognition. We have critically analyzed the strengths and limitations of different models, including traditional methods like Hidden Markov Models, and more recent deep learning approaches like Convolutional Neural Networks and Recurrent Neural Networks. Additionally, we have highlighted the challenges associated with data acquisition, feature engineering, and model generalization across different environments and scenarios.

The review has shown that while significant progress has been made, there are still several open challenges that need to be addressed. These include handling complex and multi- task activities, dealing with noisy and incomplete data, ensuring privacy and security, and developing models that can adapt to changing environments and user behaviours.

Overall, the field of human activity recognition using machine learning models has matured significantly, and the techniques discussed in this review have the potential to revolutionize the way we interact with intelligent systems and facilitate seamless human-computer interaction.In conclusion, human activity detection through mobile device sensors has numerous applications in various fields. The types of sensors utilized, such as the accelerometer, gyroscope, and magnetometer, can provide a more accurate representation of human activity. Techniques such as rule- based methods, machine learning algorithms, and deep learning algorithms can be utilized to detect activities. The applications of human activity detection in healthcare, sports, and entertainment are vast, and the potential for future developments in this field is promising.

REFERENCES

[1] Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. IEEE Communications Surveys & Tutorials, 15(3), 1192-1209.

[2] R. Saeedi, S. Norgaard, and A. H. Gebremedhin, "A closed-loop deep learning architecture for robust activity recognition using wearable sensors," *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA, USA, 2017, pp. 473-479, doi: 10.1109/BigData.2017.8257960.

[3] V. Sumathi, D. Vanathi, J. C. Musale, T. V. S. Gowtham Prasad, and A. R. Singh, "Improved Pattern Recognition Techniques for Monitoring Human Activity Recognition in Digital Platforms through Image Processing Techniques," *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, Trichy, India, 2023, pp. 866-871, doi: 10.1109/ICAISS58487.2023.10250478.

[4] S. R. SannasiChakravarthy et al., "Intelligent Recognition of Multimodal Human Activities for Personal Healthcare," *IEEE Access*, vol. 12, pp. 79776-79786, 2024, doi: 10.1109/ACCESS.2024.3405471.

[5] V. Dissanayake, S. Seneviratne, R. Rana, E. Wen, T. Kaluarachchi, and S. Nanayakkara, "SigRep: Toward Robust Wearable Emotion Recognition With Contrastive Representation Learning," *IEEE Access*, vol. 10, pp. 18105-18120, 2022, doi: 10.1109/ACCESS.2022.3149509.

[6] H. Brock, J. Ponce Chulani, L. Merino, D. Szapiro, and R. Gomez, "Developing a Lightweight Rock-Paper-Scissors Framework for Human-Robot Collaborative Gaming," *IEEE Access*, vol. 8, pp. 202958-202968, 2020, doi: 10.1109/ACCESS.2020.3033550.

[7] A. F. Gentile, D. Macrì, E. Greco, and A. Forestiero, "Privacy- Oriented Architecture for Building Automatic Voice Interaction Systems in Smart Environments in Disaster Recovery Scenarios," *2023 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Cosenza, Italy, 2023, pp. 1-8, doi: 10.1109/ICT-DM58371.2023.10286949.

[8] Y. E. Marhraoui, O. Bounhar, M. Boukallel, M. Anastassova, S. Bouilland, and M. Ammi, "Dual-Stage Attention-Based model for rehabilitation activity recognition using data from wearable sensors," *IEEE Internet of Things Journal*, 2024, doi: 10.1109/JIOT.2024.3519225.

[9] J. -H. Li, L. Tian, H. Wang, Y. An, K. Wang and L. Yu, "Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity," in IEEE Access, vol. 7, pp. 42565-42576, 2019, doi: 10.1109/ACCESS.2019.2905575.

[10] 1. Bouthillier, X., Konda, K., Vincent, P., Memisevic, R.: Dropout as data augmentation (2016)

[11] 2. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation strategies from data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

[12] 3. Feichtenhofer, C., Fan, H., Malik, J., He, K.: Slowfast networks for video recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)

[13] 4. Singh, A., Chakaborty, O., Varshney, A., Panda, R., Feris, R., Senko, K., Das, A.: Semi- supervised

action recognition with temporal contrastive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10389– 10399 (2021)
[14] 5. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems 33, 596–608 (2020)
[15] 6. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayroles, A., Jégou, H.: Training dataefficient image transformers & distillation through attention. In: International conference on machine learning. pp. 11347–11357. PMLR (2021)
[16] Yang, C., Xu, Y., Dai, B., Zho, B.: Video representation learning with visual temporal consistency. arXiv preprint arXiv:2006.15599 (2020)
[17] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119, 3-11.
[18] Ordóñez,F. J., & Roggen, D. (2016). Deep convlutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors, 16(1), 115.
[19] Ravi, D., Wong, C., Lo, B., & Yang, G. Z. (2016). A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE Journal of Biomedical and Health Informatics, 21(1), 56-64.
[20] Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. ACM Computing Surveys (CSUR), 46(3), 1-33.
[21] Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional neural networks for human activity recognition using mobile sensors. In 6th International Conference on Mobile Computing, Applications and Services (pp. 197-205).
[22] Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. arXiv preprint arXiv:1604.08880.
[23] Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59, 235-244.
[24] Jordao, A., Torres, L. A. B., & Schwartz, W. R. (2018). Novel approaches to human activity recognition based on accelerometer data. Signal, Image and Video Processing, 12(7), 1293-1301.