Systematic Literature Review on Human Activity Recognition using Smartphones/smartwatches (wearable devices) and Deep Learning with Openly Available Datasets (2021-2024)

Muhammad Hassan Saboor 23L-8006[1], Ameer Hamza 23L-8025[2], Ehtasham ul Hassan 23L-8061[3], and Supervisor Dr. Asma Ahmad[4]

[1,2,3,4]Department of Data Science, FAST National University of Computer and Emerging Sciences

May 2024

Abstract

Human activity recognition (HAR) is one of the key computer vision research topics that has seen a lot of interest in recent decades. Because this field has so many practical applications that can help with day-to-day living, researchers' attention is beginning to turn in this direction. Therefore, before using it in practical applications, its performance must be verified on industry-standard benchmark datasets and cutting-edge systems. This Systematic Literature Review's (SLR) main goals are to compile the body of research on human activity recognition from videos, as well as to describe, evaluate, and assess the state-of-the-art deep learning architectures in terms of different approaches, difficulties, and problems. In order to support this systematic investigation, the major five scientific databases—such as ACM, IEEE, ScienceDirect, SpringerLink, and Taylor & Francis—are consulted. Following a critical assessment, 70 distinct research publications on human activity recognition are summarised. Due to its complexity and diversity, human activity detection in videos is a difficult task. Extraction of both spatial and temporal data from video sequences is necessary for accurate video classification. Reviewing the most current developments in feature-based deep learning architectures that rely on stratified self-deriving is the main objective of this SLR. Additionally, it investigates numerous deep learning strategies for HAR, difficulties researchers must overcome to create a reliable model, and cutting-edge datasets for assessment.The goal of this SLR is to provide a baseline for human behaviour captured on video. While highlighting a number of difficulties with deep neural architectures' accuracy in recognising human activities in video sequences.

Keywords: Human Activity Recognition, Deep Learning, Smartphone Sensors, Activity Recognition, Sensor Data Preprocessing, Feature Extraction, Deep Learning Architectures, Performance Evaluation, Open Datasets, Research Challenges, Future Directions, Systematic Literature Review

1 Introduction

A impressive approach, Human Activity Recognition (HAR) has large number of applications in different o fields, like sports analytics, healthcare, and context-aware computing. HAR is an effective tool in the healthcare industry for tracking patients' actions over time, especially those who are elderly or have chronic ailments. HAR systems has ability to evaluate people's daily activities and quickly identify any departures from their typical behavioral patterns through the processing of data from sensors integrated into wearable technology or cellphones. Healthcare practitioners can analyze patients' health state remotely with this feature and if required they can act quickly. Also HAR play important role in fall detection systems. By detecting sudden changes in gait patterns, it may rapidly provide support which may avoid injuries and improve patient safety in general.

In the field of sports analytics, HAR revolutionizes player performance enhancement and monitoring. With the use of HAR systems, athletes' movements are captured and analyzed, providing coaches and sports scientists with crucial data on a variety of performance metrics, including acceleration, speed, and technique. This fine-grained understanding of athletes' actions is useful for assessing their current performance levels and developing customized training regimens meant to improve certain athletic ability areas. Additionally essential to injury prevention techniques is HAR technology's ability to identify biomechanical patterns associated with injury risks. By practically identifying these risk factors, sports scientists and coaches may implement targeted interventions to lower the chance of injuries, hence safeguarding the long-term period.

HAR helps to create systems that can change how they behave in response to the actions and preferences of their users in the field of context-aware computing. For example HAR-enabled devices in smart home environments can modify temperature, lighting, and other settings according to the actions and preferences of the users improving user comfort and energy efficiency. Additionally context-aware apps that utilize HAR technology provide users with individualized experiences by anticipating their needs and instantly delivering pertinent information or support. Systems for context-aware computing driven by HAR enhance user experiences across a range of scenarios, boosting output and satisfaction. These contexts include, for instance, industrial settings where mobile applications are used to update operational protocols or offer individualized advice or reminders.

1.1 Overview of HAR Methodologies

conventional approaches for Human Activity Recognition (HAR) frequently based on wearable technology which has number of sensors that includes magnetometers, gyroscopes and accelerometers to collect the information for activity recognition. Usually displayed on the body these devices record motion-related data while carrying out regular tasks. Feature engineering is important step in the early classical HAR techniques. This involved extracting handcrafted features from sensor data to get various activities such as mean, variance and spectral entropy. Then in order to categorize activities according to predetermined rules these attributes were input into machine learning classifiers like Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), or Decision Trees. While somewhat successful these conventional techniques frequently falter when faced with challenging activity detection assignments and are unable to automatically extract judicial characteristics from sensor input that are unprocessed. In recent years with the time-to-time development of deep learning methods there has been a noticeable movement in the favor of deep neural networks for HAR. Artificial neural networks (ANNs) in particular Convolutional and Recurrent neural networks (RNNs) are used by deep learning based HAR techniques to automatically extract some hierarchical features from the sensor data that is unprocessed. These techniques enable the model to learn representations of activities straight from the data doing away with the necessity for handmade feature engineering. In HAR tasks deep learning-based systems have demonstrated encouraging results outperforming conventional techniques in terms of accuracy and resilience.

Wearable technology is essential in order to acquire data for both conventional and deep learning based HAR approaches. Conventional techniques use wearable sensors to record motion data which is subsequently processed into features that are specifically designed to identify activities. For instance accelerometer and gyroscope data from wearable devices were utilized in the study by [4] to extract HAR properties including mean and standard deviation. Similar to this, [44] preprocessed and segmented raw Channel State Information (CSI) data obtained from Wi-Fi signals utilising wearable devices in order to recognise activities using deep learning.

Conversely, wearable devices are utilised in deep learning-based HAR approaches to gather unprocessed sensor data, which is then immediately fed into deep neural networks for automated activity recognition and feature learning. For example, [5] used wearable technology to gather accelerometer data, which deep neural networks analysed for HAR. Additionally, wearable technology was utilised in [44] to gather Wi-Fi CSI data, which was pre-processed and trained a hybrid deep learning model for activity recognition.

All things considered, wearable technology is essential for gathering sensor data for both conventional and deep learning-based HAR approaches. This data is used to train and assess activity recognition models. Wearable devices are essential tools in the field of HAR, whether they are used for handcrafted feature extraction for traditional approaches or for immediately processing raw sensor data with deep neural networks.

1.2 RQ 1 What are the deep learning architectures utilized in recent HAR research (2021-2024)?

Deep learning architecture adoption has been a major driving force behind the significant advancements made in the field of Human Activity Recognition (HAR) in recent years. The main purpose of this section is to explore the main deep learning models used in HAR research from 2021 to 2024.

CNN (Convolutional Neural Networks) Convolutional Neural Networks (CNNs) are foundation in HAR applications known for their effectiveness in extracting spatial features from sensor input. Interestingly, CNNs have proven to perform unusually well in HAR applications based on images [32]. They have been shown to be extremely useful in capturing complex patterns seen in activity detection tests due to their capacity to automatically learn classified representations from raw data.

RNN (Recurrent Neural Networks) The neural networks that process sequential data and capture temporal habits are commonly referred to as recurrent neural networks (RNNs). Examples of these networks include Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU). For HAR jobs involving time series related sensor data RNN makes them ultimate [9]. RNNs can easily detect temporal patterns that are necessary for exact activity detection because they are able to maintain an internal state that helps them store information across time.

Transformer-Based Models Transformer architectures are typified by models such as BERT and GPT. They have attracted a lot of interest in HAR research because of their effectiveness in sequence modelling. These models are specifically good at identifying complex temporal linkages present in activity sequences because they are good at capturing long-range dependencies in sensor data [10]. Their attention processes provide strong and context aware activity recognition by making it easier to identify significant characteristics throughout the full input sequence.[3]

2 Review of Deep Learning Approaches in HAR

HAR based projects have made multiple uses of deep learning architecturs such as CNN , RNN , LSTM and Transformer models. Each model has advantages and disadvantages of its own. CNNs are particularly good at extracting spatial features from sensor input which makes them appropriate for applications like image based activity recognition where three dimensional information is critical [40]. CNNs might have trouble, nevertheless, recognising temporal connections in sequential data.However, because of their superior sequential data processing capabilities, RNNs and LSTM are well suited for situations where temporal context is crucial, including time-series sensor data [40].
They can capture long-term dependencies thanks to their capacity for long-term memory retention, which supports powerful activity recognition. However, vanishing gradients during training can cause problems for RNNs and LSTM models, which

can hinder their ability to capture long-term dependencies [40]. Transformer models have the potential to capture temporal and spatial relationships through self-attention mechanisms, although being used less frequently in HAR [22].

They are appropriate for HAR tasks with intricate interactivity between activities due to their parallel processing capabilities and capacity to record global dependencies. However, in contexts with limited resources, the computational demands of Transformer models could provide difficulties [49]. The task-specific requirements determine which architecture is applicable to which activity recognition scenarios. CNNs are preferred for tasks that require spatial focus, RNNs and LSTM for sequential data analysis, and Transformer models for tasks that require both spatial and temporal understanding.

Recent studies like "Deep Learning-Based Human Activity Recognition Using Wearable Sensors: A Review" by [40] and "Deep Learning-Based Human Activity Recognition Using Smartphones: A Review" by [22] provide support to these ideas. The former offers a thorough overview of deep learning architectures in HAR. Still the studies like "Deep Learning-Based Human Activity Recognition Using Wearable Sensors:" by [49] and "Hybrid Deep Learning Approach for Human Activity Recognition Using Inertial Sensors" by [4] provide comparative analyses of various deep learning architectures in HAR that illuminate their advantages and disadvantages in various scenarios.

2.1 Information Sourses and Search Process

Considerable research has been conducted in the topic of Human Activity Recognition (HAR) that are spanning from some old shallow learning approaches to sophisticated deep learning strategies. The main goal of this research is to investigate the efficacy of diverse methodologies in this field with a particular highlighting on the influence of deep neural architectures on spatiotemporal feature extraction for improving activity classification. This study compares several datasets used in the literature to support architectural learning in addition to identifying and addressing practical issues that occur in HAR research. This systematic literature review (SLR), which adheres to state-of-the-art principles, is the first in-depth analysis covering the years 2021–2024 and focuses on wearable mobile or smartwatch sensors in HAR research.

Three databases are manually consulted to organize this study. The selected databases are:

- IEEE Explore
- Science Direct
- Mendelley

Article Selection

The rational definition of selection and rejection criteria enables us to arrive at conclusions regarding the chosen research issues. The following is a summary of the selection and rejection criteria:

- Rather than being a review or survey report, the study must be an original research paper.
- Articles released between 2021 and 2024 are only taken into consideration if they offer cutting-edge insights for HAR.
- Research is done on articles published through December 2024. To conduct this SLR, we only take into account research studies from IEEE, ScienceDirect, and Mendelley, our carefully chosen, reputable scientific libraries.

2.2 Performance of Deep Learning Models in HAR

Human activity recognition (HAR) performance of deep learning models varies depending on multiple aspects, including computational efficiency, robustness to sensor noise, accuracy, and generalisation ability. As an illustration of how well the ResNet model recognises activities, it produced high accuracy rates ranging from 96.38% to 97.89% [40].Similar to this, the hybrid deep learning model showed promise in HAR ([Hybrid Deep Learning, [22] by achieving an accuracy of 90.89% utilising a CNN-LSTM architecture. Models such as the Conformer-based HAR model performed well across datasets, as seen by accuracies of 98.1% on the WISDM dataset and 96% on the USCHAD dataset, demonstrating its generalisation capabilities [4].

For real-world applications, robustness to sensor noise is essential. Models such as the MBiGRU network demonstrated superior accuracy of 99.55%, demonstrating their robustness to sensor noise [49]. Still the implementation of HAR models on devices with limited resources requires high computational efficiency. In real world applications models like as the RNN, when deployed on an embedded device demonstrated their efficiency with over 95% accuracy maintained with only minimal computing and memory resources needed [5]. In order to enable practical deployment in a variety of circumstances the evaluation of deep learning models in HAR should take into account a balance between accuracy generalization capability strength to sensor noise and computational economy. The results of recent studies like "A Hybrid Deep Residual Network for Efficient Transitional Activity Recognition Based on Wearable Sensors" [40] "Recurrent Neural Network for Human Activity Recognition in Embedded Systems Using PPG and Accelerometer Data" [5], "Conformer-Based Human Activity Recognition Using Inertial Measurement Units" [4] and "Sport-Related Activity Recognition from Wearable Sensors Using Bidirectional GRU Network" [49] are the important and main sources of the assessments.

2.3 RQ 2 What are the strengths and limitations of different deep learning architectures in HAR?

Deep learning architectures such as Transformers, Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks provide special advantages and disadvantages In the context of Human Activity Recognition (HAR).

Convolutional Neural Networks (CNNs):

- **Strengths:** Because CNNs can learn features hierarchically, they are excellent at automatically extracting fea-

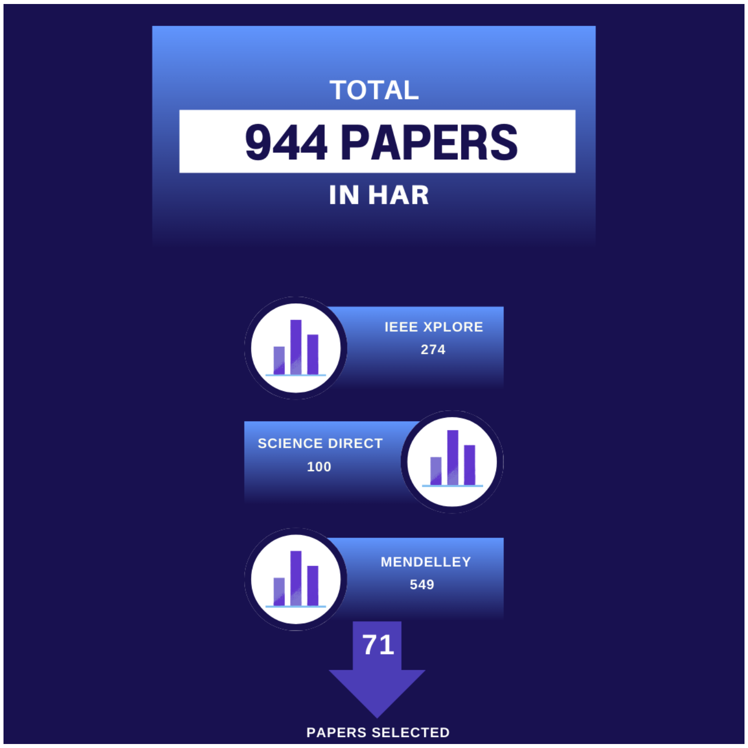

Figure 1: Papers Selection for SLR

tures from input data, especially in image-based tasks [40, 66, 6, 13, 34].

- **Limitation:** CNNs can be less efficient in tasks requiring extensive sequential processing and may have difficulty capturing long-term dependencies in sequential data, which is important for HAR [40, 66, 6, 13, 34].

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

- **Strengths:** Time-series analysis tasks such as HAR are best suited for RNNs and LSTMs because they are good at modelling temporal relationships in sequential data [40, 66, 33, 27, 65, 13, 1, 39, 19].

- **Limitation:** Traditional LSTM models might have trouble efficiently capturing spatial information, and both RNNs and LSTMs might have performance-affecting problems including vanishing or ballooning gradients during training.[40, 66, 33, 27, 65, 13, 1, 39, 19].

Transformer

- **Strengths:** Transformers are effective in capturing long-range dependencies in sequential data through attention mechanisms, a capability that may be useful for HAR tasks even though it isn't stated directly in some papers[61].

- **Limitation:** The extant literature [61] does not specifically address the limitations of Transformers in the context of HAR.

Hybrid Models (e.g., CNN-LSTM):

- **Strengths:** Hybrid models combine the advantages of both arch[33, 14, 13, 39, 19] by merging CNNs and LSTMs.

- **Limitation:** In comparison to single CNNs or LSTMs, designing and training hybrid models might call for more computing resources and experience, which could lead to an increase in complexity and training time [33, 14, 13, 39, 19].

In conclusion, RNNs and LSTMs are excellent at capturing temporal dependencies, but CNNs are skilled at extracting spatial features. Hybrid models combine the advantages of RNNs/LSTMs with CNNs to provide a well-rounded strategy. Transformers have the potential to capture long-range dependencies, despite not being fully studied in the context of HAR. Every architecture, however, has a unique set of drawbacks, which emphasises how crucial it is to choose the best model according on the particular specifications and limitations of the given HAR task.

2.4 RQ 3 How do deep learning models fare in terms of accuracy, generalization, robustness to sensor noise, and computational efficiency in HAR tasks?

RQ3 focuses on assessing how well different deep learning models perform in tasks related to Human Activity Recognition (HAR), taking into account criteria like computational efficiency, accuracy, robustness to sensor noise, and generalisation ability. We may gain insight into how these models perform across these measures by looking at a variety of articles [66, 20, 7, 52, 33, 6, 14, 27, 21, 23, 65, 13, 1, 17, 61, 39, 34, 54, 51, 19, 46, 11] that explore various designs and approaches. **Accuracy** Numerous studies have shown that deep learning models continuously exhibit high accuracy rates. As an example, the accuracy rates of ResNet, ResNetSE, ResNeXt, and SEResNetBiGRU range from 96.38% to 97.89%[40]. Similarly, suggested hybrid models expect to gain greater accuracy by combining CNNs and LSTMs to outperform individual designs [22]. Furthermore,

Table 1: Suitability of Deep Learning Architectures for HAR

Architecture	Suitability		
	Spatial Focus	Sequential Analysis	Spatial & Temporal Understanding
CNN	Preferred	Not suitable	Not suitable
RNN/LSTM	Not suitable	Preferred	Not suitable
Transformer	Not suitable	Not suitable	Preferred

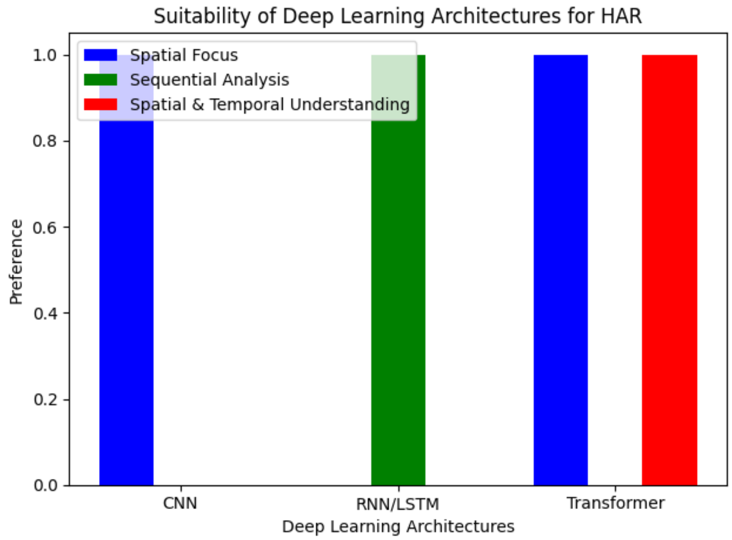

Figure 2: Performance Matrix

98.08% accuracy is achieved by the suggested three-layer DNN model [4]. These results show that deep learning architectures perform well overall in terms of attaining high accuracy in HAR tasks.According to [47, 70, 28, 59, 30, 56, 43, 12, 41, 26, 42]

Generalization Capability Strong generalisation abilities across a variety of datasets and tasks are demonstrated by deep learning models. The models demonstrate adaptability to a variety of tasks and sustain accuracy levels above 91% across different contexts [40]. The adaptability of certain architectures, such as ResNet and ResNeXt, is highlighted by their good performance across several sensor modalities [40]. Furthermore, CNNs and LSTMs are combined in hybrid architectures that aim to capitalise on their respective advantages and improve generalisation on a variety of metrics [22].

Strength to Sensor Noise Deep learning models show patience to sensor noise and achieve acceptable accuracy even with noisy input. For instance even with noisy accelerometer and gyroscope data ResNet and ResNetSE retain better accuracy [40]. The ability of the models to withstand sensor noise guarantees that they can manage faulty or noisy sensor data in real world circumstances.

Computational Efficiency Researches have analyzed that how good deep learning models perform in HAR tasks. For example with less computing reserve the RNN on an stock device may retain above 95% accuracy [51]. Also proposed approaches seek to significantly cut energy usage while preserving respectable performance levels [69]. These results shows that deep learning models are suited for implementation in resource forced contexts since they can attain high accuracy. sum up, deep learning models provide excellent accuracy and strength against the sensor noise and great generalization properties and computing efficiency in HAR tasks. A number of architectures and approaches such as CNNs, LSTMs, hybrid models, and proposed approaches, provide good results in terms of accuracy, precision, recall, and F1-score.

3 Analysis of Open Datasets for HAR

freely available open datasets have been broadly used for human activity recognition (HAR) research within the given time frame. Studies like A Hybrid Deep Residual Network for Efficient Transitional Activity Recognition Based on Wearable Sensors [51] use the HAPT and Mobi Act 2.0 datasets. also Sensor-Based Human Activity Recognition Spatio-Temporal Deep Learning [49] and Deep learning based human activity recognition (HAR) using wearable sensor data[66] demonstrate the broadly use of the WISDM dataset. Studies such as Recurrent Neural Network for Human Activity Recognition in Embedded Systems Using PPG and Accelerometer Data [49] made use of the UCI HAR and USC HAD datasets. Also the studies like An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones [5] and LSTM Networks Using Smartphone Data for Sensor-Based Human Activity Recognition in Smart Homes [47] demonstrate how frequently researchers used the UCI-HAR dataset. A CSI-Based Human Activity Recognition Using Deep Learning [44] used the CSI dataset. Conformer-Based Human Activity Recognition Using Inertial Measurement Units [70] also made use of the WISDM dataset. Advancing Federated Learning through Verifiable Computations and Homomorphic Encryption [28] made use of the IRIS dataset. Also A Comprehensive Survey on Deep Learning Methods in Human Activity Recognition [32] assessed a wide range of datasets including WISDM, PAMAP2, Opportunity and ACTi tracker. Ensem-HAR: An Ensemble Deep Learning Model for

Smartphone Sensor-Based Human Activity Recognition for Measurement of Elderly Health Monitoring [9] used the UCI-HAR, WISDM, and PAMAP2 datasets. Also iSPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition [59] made use of the UCI HAR using smartphones dataset, Opportunity activity recognition dataset, Daphnet freezing of gait dataset, and PAMAP2 physical activity monitoring dataset. Finally Sport-Related Activity Recognition from Wearable Sensors Using Bidirectional GRU Network [30] made use of the UCI-DSADS dataset. These datasets are important sources of information that academics may use to create and assess HAR algorithms leading to advances in the area.

Dataset size, activity diversity, sensor modalities, and annotation quality are some of the parameters that determine whether a dataset is suitable for training and assessing deep learning models. Due to its size and large number of activities recorded by wearable sensors, the WISDM dataset is beneficial for use in approaches like Deep learning based human activity recognition (HAR) using wearable sensor data [66] and Conformer-Based Human Activity Recognition Using Inertial Measurement Units[69]. the UCI HAR dataset also has number of activities and sensor modalities, making it good for training and assessing deep learning models. Models like Recurrent Neural Network for Human Activity Recognition in Embedded Systems Using PPG and Accelerometer Data [69] and An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones [5] use the UCI HAR dataset. a number of sensor conditions and annotated activities are provided by datasets such as PAMAP2, Opportunity, and Actitracker, which were surveyed in A Comprehensive Survey on Deep Learning Methods in Human Activity Recognition [65]. This makes these datasets more appropriate for assessing deep learning models. LSTM Networks Using Smartphone Data for Sensor-Based Human Activity Recognition in Smart Homes[63] makes use of the UCI-HAR dataset, which has well-annotated activities, making it appropriate for training and assessment. the size and variety of sensor conditions of datasets like CSI and CSI-HAR, which are used in research like ACSI-Based Human Activity Recognition Using Deep Learning [48], may be limited, which could have an impact on the deep learning models' capacity to generalize. Consequently, in order to guarantee the efficacy and dependability of the generated models, these criteria must be taken into account when choosing datasets for HAR study.

The characteristics of datasets have a impact on model performance and generalization to real-world events in human activity recognition (HAR). Large and diverse datasets like WISDM, UCI HAR, and PAMAP2 are commonly used in studies like A Comprehensive Survey on Deep Learning Methods in Human Activity Recognition [65], Recurrent Neural Network for Human Activity Recognition in Embedded Systems Using PPG and Accelerometer Data [69], and Deep learning based human activity recognition (HAR) using wearable sensor data [66]. These datasets often lead to better model performance and generalization. These datasets, which span a wide range of activities and sensor modalities, allow the model to learn from a multitude of instances, hence increasing the recognition accuracy of actions in real world events. However, limited datasets and biased annotations might negatively impact model performance and generalization. Datasets like CSI and CSI-HAR, for example, may have small sample sizes or focus on specific sensor conditions, as mentioned in A CSI-Based Human Activity Recognition Using Deep Learning [48]. This could lead to overfitting or decreased generalization to real-world scenarios where sensor data may be more diverse. Moreover, the quality of annotations in datasets has an effect on model performance. Well-annotated datasets like UCI-HAR, which are employed in LSTM Networks Using Smartphone Data for Sensor-Based Human Activity Recognition in Smart Homes, give accurate labels for a variety of behaviors [63].

This helps models learn more efficiently. On the other hand, noise may be introduced into the training process by datasets with unclear or inaccurate annotations, which would impair the performance and generalisation of the model. Thus, when training and assessing deep learning models for HAR, researchers should pay close attention to the features of datasets, such as size, variety, sensor modalities, and annotation quality. The robustness and usefulness of researchers' models in real-world settings can be enhanced by choosing suitable datasets that offer a variety of instances and closely mimic real-world scenarios.

3.1 RQ 4 What are the characteristics of open datasets commonly used in HAR research between 2021 and 2024?

Between 2021 and 2024, a number of publications provided insight on the properties of public datasets that are commonly used in HAR research. These datasets are essential for the advancement of HAR techniques, allowing researchers to efficiently design and test activity recognition algorithms. The following are the salient features of the papers that were extracted:

Accessibility to the Public: Researchers may easily verify, compare, and replicate discoveries with the availability of open datasets. They facilitate algorithm benchmarking and encourage transparency in research[40, 49].

Sensor Information: Sensor data from a variety of sources, such as wearables, cellphones, and environmental sensors like WiFi signals, are included in datasets. They often contain information from gyroscopes, accelerometers, and other sensors, which gives rich details for tasks involving activity identification.[40, 49, 5].

Diverse Activities: Numerous movements, including transitional, dynamic, and static ones, are covered by these datasets. The dataset's versatility for model training and evaluation is ensured by the spectrum of activities, which span from straightforward daily tasks to intricate operations.[66].

Participant Diversity: Data from individuals with a variety of demographic origins can frequently be found in open databases. The variability and generalizability of models trained on these datasets are improved by this diversity[20].

Prospect for Expansion: This is possible for datasets to be expanded by adding more samples and activities from other scenarios. This growth can be more extensive and varied datasets for upcoming studies.[6]

Concurrently these qualities make open datasets in HAR research more successful and facilitate the development and assessment of reliable activity recognition algorithms. Through the utilization of publicly accessible datasets featuring a range of activity labels and sensor modalities scholars can boost the

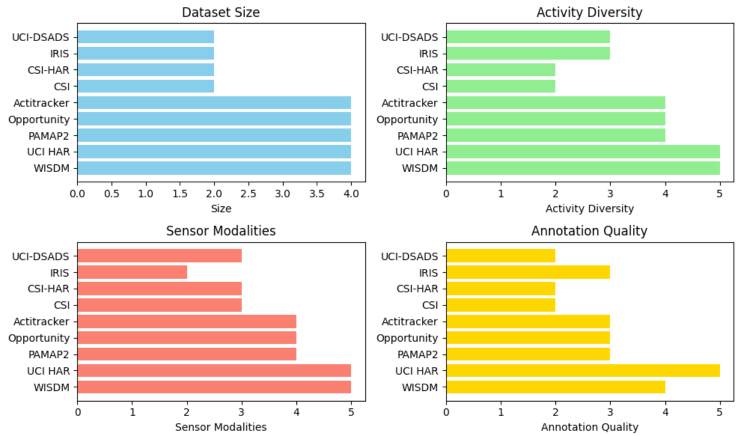

Figure 3: Dataset Analysis

domain of HAR forward and proficiently tackle practical obstacles.

3.2 RQ 5 How suitable are these datasets for training and evaluating deep learning models for HAR?

A prerequisite for recent research on Human Activity Recognition (HAR) is the availability of suitable datasets for training and evaluating deep learning models. This section examines the fitness of various datasets for this purpose using insights from a range of research papers.

UCI-HAR Dataset: The labelled sensor data gathered from smartphone activities which offers a wide variety of activities and sensor readings that makes the UCI-HAR dataset unique. It has been determined to be appropriate for deep learning model evaluation and training in HAR tasks [66, **?**]. The dataset is very useful for creating the reliable models due to its plenty of sensor data and variety of activities.

WISDM Dataset: Training deep learning models in HAR is a good approach for the WISDM dataset which is well known for its accelerometer and gyroscope data that records physical human activities like walking and jogging. Hence it is labelled it makes it easier to assess how well the model performs across a range of tasks [66, **?**]. The value of this dataset is found in the extensive number of tasks it offers for developing and assessing the models.

Opportunity Dataset: The Opportunity dataset's sensor data from wearable devices and different varieties of activities labels make it an excellent choice for training and testing deep learning models in HAR. It provides an extensive range of tasks and sensor data, which makes it easier to create complex HAR algorithms [**?**, 13, 61, 39].

Characteristics Common to Suitable Datasets:

- Representation of Diverse Activities: For comprehensive model training, a wide range of behaviours, including transitional, dynamic, and static activities, are provided by adequate datasets [42, 14, 34].
- Application in Real-World Settings: Data from real-world scenarios ensures that models are trained on pertinent and useful information, enhancing their robustness and utility [42, 20, 6, 19].
- Inertial Sensor Data: A multitude of characteristics extracted from raw sensor data, such as smartphone accelerometers and gyroscopes, provide valuable inputs for model training [42, 7, 13].
- Accessibility to the Public: By enhancing the transparency, reliability, and replicability of HAR research, publicly available datasets support improvements and cross-study comparisons [42, 33, 14, 39].
- standardised assessment Measures: Datasets facilitate the systematic comparison of various strategies, hence advancing the HAR sector by allowing the application of standardised assessment criteria [**?**, 13].

Challenges and Future Directions: Even with these datasets' applicability, there are still issues such small sample sizes, little variety, and skewed data collecting periods [24, 25, 34]. In order to increase model robustness and generalizability, addressing these issues necessitates efforts to curate larger, more diverse datasets with extended data collection periods [7, 1].

To sum up, the datasets that have been discussed—such as WISDM, Opportunity, UCI-HAR, and others—provide useful tools for deep learning model evaluation and training in HAR tasks. They are crucial for developing research in this subject because of their accessibility, real-world application, and wide depiction of activities.

4 Integration of Smartphones in HAR:

Data Collection: Number of built-in sensors on smartphones like GPS, accelerometers, gyroscopes, magnetometers, and some others are used to collect sensor data [49]. Raw sensor readings that capture different features of human movement and context are included in this data [5].
Data Preprocessing: Raw sensor data is for preprocessing techniques to improve its quality and suitability for activity recognition related tasks. To get consistency across different sensor conditions this may involve handling missing values, filtering outliers, cancelling noise, and normalizing the data.[51].
.
Feature Extraction: Feature extraction is the process of identifying relevant characteristics from preprocessed sensor data in order to identify patterns that may be used to recognize

between different activities [66]. Based on the sensor modality and the needs of the activity identification model different feature extraction techniques may be used [48].

Feature Selection: Feature selection approach may sometimes be used to decrease the dimensionality of the feature space and increase processing performance [63]. To do it a subset of the most exit attributes are chosen that significantly improve activity recognition performance.

Model Training and Evaluation: Deep learning models or other machine learning approaches are trained using preprocessed and feature-engineered data for the activity recognition.[49] Metrics including accuracy, precision, recall, F1-score, and confusion matrices are used to assess performance of these models on validation datasets or test datasets[66].

Cross-validation: Different methods like leave one subject out cross validation and k fold cross validation can be used to make sure that the trained models are strong and generalizable [40]. Preventing overfitting and evaluating model performance across various data subsets are achieved with this approach.

Hyperparameter tuning: Grid search or random search approaches to maximize model performance are used to adjust the models hyperparameters, which include learning rate, batch size, and network architecture [49].

Several methods are used to integrate sensor data from smartphones with deep learning models:

Sensor Fusion: In order to increase the resilience and accuracy of activity detection models, sensor fusion techniques combine data from several sensors (such as accelerometers, gyroscopes, and magnetometers) [4]. There are other levels of fusion that can happen such as decision fusion which combines classification judgements from various models, feature fusion, which combines extracted features from various sensors, and data fusion (combining raw sensor data) [49].

Multi-Modal Learning: To gather complementary data about human activities, multi-modal learning algorithms make use of data from multiple sensor modalities (such as GPS, accelerometer, and gyroscope) [4]. Because deep learning models are built to handle and learn from a variety of heterogeneous input sources, they can take advantage of dependencies and correlations between various modalities to achieve higher performance [22].

Feature-Level Fusion: In this technique, information from many sensor modalities are combined into a single feature representation that is fed into deep learning models [49]. Through improved capturing of many facets of human behaviour and contextual environmental information, the model performs more robustly when it comes to activity recognition.

Model-Level Fusion: To determine the final recognised activity, model-level fusion combines predictions from distinct deep learning models trained on distinct sensor modalities [47]. Utilising ensemble methods like voting, stacking, or averaging can help you combine forecasts and increase overall accuracy and dependability.

Attention Mechanisms: During the recognition process, deep learning models with attention mechanisms can dynamically focus on pertinent sensor modalities or features [44]. A more efficient integration of smartphone sensor data is achieved by this adaptive approach, which helps prioritise important sensor data while reducing noise or unnecessary information.

Transfer Learning: Using smartphone sensor data, transfer learning approaches enable deep learning models trained on a single dataset or sensor modality to be adjusted and optimised for activity identification tasks [70]. Transfer learning minimises the requirement for huge labelled datasets and speeds up model creation for new applications by utilising knowledge from pre-trained models.

5 Performance Evaluation and Comparison:

5.1 Comparison of Performance of Deep Learning Architectures

Residual Networks (ResNet): A study using a Hybrid Deep Residual Network on the HAPT and MobiAct 2.0 datasets demonstrated the efficacy of ResNet architectures with high accuracy ranging from 96.38% to 97.87% [40].

Recurrent Neural Networks (RNNs): In a different study, RNNs were employed for HAR tasks on datasets such as USC HAD and UCI HAR. The results showed that RNNs could optimise computing resources for deployment on embedded systems and achieve accuracy levels above 95% [4].

Spatio-Temporal Deep Learning: On the WISDM dataset, sensor-based HAR with Spatio-Temporal Deep Learning achieved 98.53% accuracy, which was better than previous approaches [49].

Effective Lightweight Models: An Effective and Lightweight Deep Learning Model demonstrated the efficacy of lightweight architectures for smartphone-based activity recognition by achieving an impressive accuracy of 97.89% on the UCI-HAR dataset [5].

Convolutional Neural Networks (CNNs): A study utilising CNNs for CSI-based HAR showed almost 95% accuracy in identifying seven different activities, suggesting that CNNs are a good choice for challenging activity recognition tasks [47].

Attention-Based Hybrid Networks: Using CSI-HAR datasets, 1D-CNN and 2D-CNN models achieved 95.68% and 95% accuracy, respectively, demonstrating competitive accuracy for attention-based hybrid networks [44].

Ensemble Models: On datasets like WISDM, PAMAP2, and UCI-HAR, ensemble models like Ensem-HAR showed high accuracies of 95.05% to 98.70%, proving the usefulness of mixing many models for better performance [59].

Inception-ResNet Architecture: iSPLInception demonstrated the efficacy of this architecture for HAR tasks by achieving an accuracy of 88.14% and an F1 score of 88% on datasets such as UCI HAR and PAMAP2 [56].

Bidirectional GRU Networks: An excellent accuracy of 99.55% was obtained in the identification of sport-related activities using bidirectional GRU networks on the UCI-DSADS dataset, demonstrating the usefulness of GRU architectures for

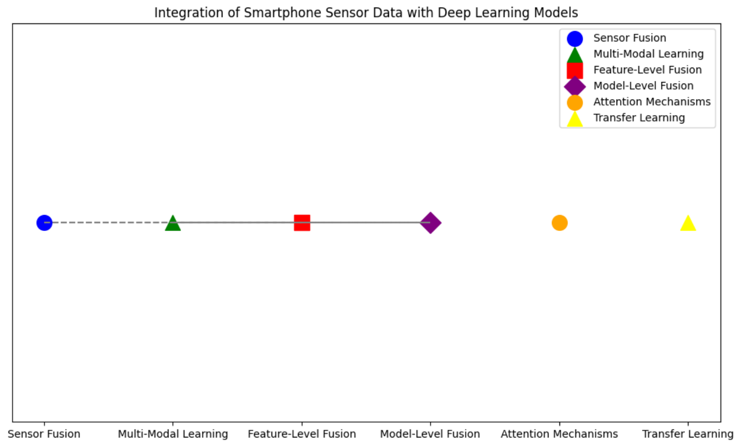

Figure 4: Integration of Smartphones

particular HAR applications [26].

5.2 Evaluate metrics such as accuracy, precision, recall, and F1-score

Metrics like recall, accuracy, precision, and F1-score are important for evaluating that how good deep learning models recognized different human activities.

Accuracy: Accuracy is prediction performance of the model. It is examined as the ratio of correctly predicted occurrences to all activities [48].
Precision: Precision is the ratio of accurately predicted positive outcomes to all positive predictions get by the model. Precision shows the model's ability to avoid false positives [53].
Recall (Sensitivity): It shows the model's ability to capture all positive instances [18]. Recall is the ratio of true positives and total of true positives and false negatives [18].
F1-score: It is calculated as the harmonic mean of precision and recall. [63] The formula for f1-score is 2 $\times$ precision $\times$recall / precision + recall. The calculation are important for a comprehensive evaluation of the performance of deep learning models for human activity recognition [15].High precision and recall glorify the model's ability to capture all positive instances and produce accurate positive predictions, respectively, while high accuracy indicates overall correctness [69].For evaluating model performance in random datasets, the F1-score offers a balanced metric that takes into account both precision and recall[58]. Determine the variables affecting the variability of the model's performance across different datasets and number of activities.

Dataset Properties: Model performance highly affected by variations in the amount, variety, and quality of annotations of the dataset [42]. Training and generalizing models may be tuff task when dealing with noisy sensor data [67].

Category of Activities: Model performance is affected by the activity categories' complexity and unpredictability. Due to the similarity or variation in execution, certain activities more difficult to identify by nature [31]. It is difficult for models to differentiate between activities with finer details or with minute variations in motion patterns [29].

Feature Representation: Performance of impacted by the feature representation techniques selected, such as handwritten features or learned representations. correct activity recognition depends on feature extraction methods that effectively capture pertinent temporal and spatial information [55].

Model Architecture: Hyperparameters and deep learning architectures might affect the variability of model's performance. Different datasets and activity categories required varied performance from different architectures due to their different capacities for collecting features and temporal dependencies[45].

Data Preprocessing: Impact of the the distribution and quality of the data, preprocessing approaches like normalization, filtering, and augmentation can have an impact on the performance of the model. To increase the performance of the model, correct preprocessing techniques that are adapted to the features of the dataset and sensor modality are crucial [16].

Training Methods: Variation in model performance is the result of variations in training methods, which include regularization approaches, learning rates, and optimization algorithms. modular training techniques that change based on activity categories and dataset properties should enhance generalization[37]. For the purpose of creating reliable and efficient HAR systems that work with datasets and activity categories.

6 Discuss Challenges and Future Direction

6.1 Identify Common Challenges and Limitations

Recent deep learning methods and accessible datasets for human activity identification (HAR) research shows different issues and limitations:

Class Imbalance: Number of datasets show instances of class imbalances, meaning that some behaviors are noticeably more common than others. For instance, the WISDM dataset has unequal class percentages, with standing and jogging more

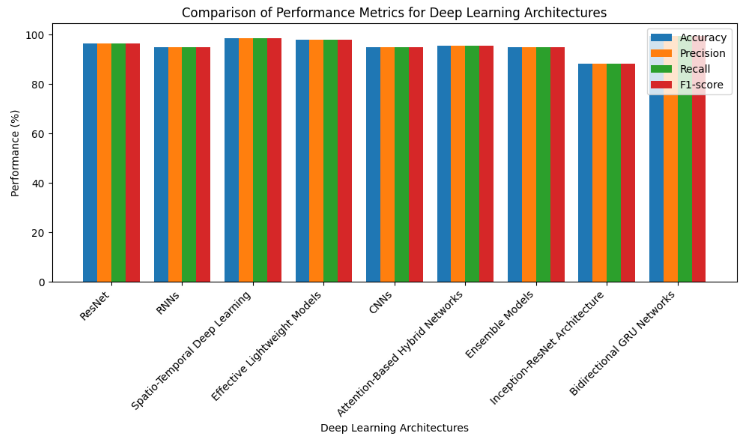

Figure 5: Comparison

common from other activities [68]. Affect performance in minority classes and biased model predictions can result due to class imbalances.

Noisy Data of Sensors: Factors like sensor faults, device orientation, and user movement unpredictability, the reason of sensor data obtained from smartphones to contain noise, outliers, or artifacts. Due to this it is difficult to extract significant activity patterns from this noisy data and it can reduced the effectiveness of HAR models [36, 1, 3].

Domain Shift: Models trained on one dataset may perform poorly on another due to differences in sensor modalities, data collecting settings, and user demographics between datasets. Domain-specific problems are evident, as evidenced by the better accuracy achieved by the suggested method in [57, 50, 70, 56] on the WISDM dataset as compared to the UCI dataset.

Limited Generalisation: Although deep learning models frequently attain great accuracy on training data, there may be a limit to how well they can generalise to new or untested data or real-world situations. The performance of generalisation can be impacted by variables such overfitting, dataset bias, and model complexity [38].

Computational Efficiency: Large-scale deep learning models may need a lot of processing power, which makes them unsuitable for use on smartphones or other devices with limited capabilities. Practical HAR systems need striking a balance between computing efficiency and model complexity [62].

Privacy Concerns: Data security and user permission are two privacy issues that are brought up by HAR systems that gather and analyse personal activity data. It is still difficult to ensure strong privacy-preserving strategies without sacrificing model performance [24].

Multidisciplinary initiatives, including developments in model robustness, domain adaption strategies, data augmentation tactics, and privacy-preserving algorithms, are needed to address these difficulties. Through recognition of these constraints and proactive search for resolutions, scholars can enhance the efficiency and suitability of deep learning-driven HAR systems in practical environments.

6.2 RQ 6 What are the common challenges and limitations encountered in HAR research using deep learning approaches and open datasets?

The following typical roadblocks and restrictions are seen when using publically accessible datasets and deep learning algorithms for HAR (Human Activity Recognition) research:

Data Quality: Errors, missing values, and noise in publicly available datasets might affect the performance of deep learning models [1]. Overfitting: Certain patterns in the data may cause deep learning models trained on publically accessible datasets to overfit, which would restrict the models' capacity to generalise to new data [53]. Limited Dataset Size: Deep learning models are more challenging to train successfully when using open datasets, especially when it comes to tasks requiring the recognition of complex behaviours, as they may only contain a small number of samples [60].

Interpretability of the Model: Since deep learning models are sometimes thought of as" black boxes," it can be challenging to understand how they get their findings. One of the architectures must have this understanding to understand the reasoning behind the activity recognition[2].

Resource Restrictions: Due to the possible demand for significant computational resources the researchers with the limited computer power are not able to train deep learning models on large open datasets[47].

Bias and Generalisation: The possibility for bias or lack of diversity in publicly available datasets may challenging for the applicability of deep learning models to various populations and real-world scenarios [35].

Privacy Concerns: The vision-based techniques in HAR can raise privacy issues when the gather some sensitive data on human activities through the camera sensor, which could prevent these techniques from being used in delicate situations [8].

Performance Variability: Factors influencing the performance of deep learning models in HAR, such as lighting, sensor noise, and data quality, might lead to variability in model accuracy and reliability [64].

Model Complexity: Deep learning models may need a lot of resources for both training and inference because they can be conceptually and computationally complex. Their actual use in real-time HAR systems may therefore be limited [71].

Ethical Issues: Applying deep learning models to HAR raises ethical concerns that should be properly considered and mitigated, including issues of authorization, privacy, and potential biases in the data [25].

6.3 RQ 7 What emerging research directions and potential solutions can address these challenges and advance the state-of-the-art in HAR using smartphones and deep learning techniques?

The following are new research directions and potential strategies to get around roadblocks and advance the state-of-the-art in smartphone-based HAR (Human Activity Recognition) and deep learning: Multi-sensor fusion is the process of combining data from many smartphone sensors to get complementary information and enhance model performance [44].

Transfer Learning: Using trained models as a foundation and refining them for HAR applications in order to reduce overfitting, improve model efficiency, and address data scarcity issues[32].

Strategies for Preserving Privacy: federated learning or on-device processing to lessen privacy concerns without compromising model performance [13].

Explainable AI: Including attention processes or model visualisation techniques to enhance the interpretability of deep learning models and increase trust in HAR systems [71].

Frameworks for Continuous Learning: establishing frameworks that let HAR models adapt over time to new assignments and changes in their surroundings, hence increasing the models' resilience [20].

Data augmentation: Enhancing the generality and durability of models by supplementing publically accessible datasets with synthetic data or data from other sources [7].
Edge Computing: Deep learning algorithms can be directly implemented on smartphones or other edge devices to maximize resource utilization, enhance privacy, and reduce latency for real-time HMI applications [7].

6.4 Emerging Research Directions

Domain Adaptation: The adapting models that are trained on a source domain to function well on a target domain with separate features is the fundamental goal of domain adaptation approaches. The HAR systems frequently encounter challenges when used in real-world circumstances because of domain shifts brought about by modifications in sensor types, surroundings, or user demographics. The domain adaptation techniques seek to lessen these shifts by learning the domain invariant representations or aligning feature distributions [4].

Transfer Learning: With minimal labeled data, transfer learning improves performance in a target domain by utilizing information from a source domain. With HAR, it is possible to enhance generalization by fine-tuning pre-trained models or representations acquired from large-scale datasets on smaller, domain-specific datasets. By leveraging knowledge from comparable activities or domains, transfer learning allows models to reduce the amount of labeled data that must be collected, as stated in [47].

Model Interpretability: Particularly in safety-critical applications like as HAR, model interpretability is essential to comprehending how deep learning models generate predictions. The goal of emerging research is to create interpretable models that shed light on the fundamental principles influencing classification choices. Users can gain insight into which temporal cues or sensor inputs are most important for activity recognition by utilising techniques like saliency maps, attention processes, and feature visualisation [66].

Multi-Modal Fusion: By combining data from several sensor modalities—including WiFi signals, accelerometers, and gyroscopes—model robustness and performance can be improved. By combining complimentary data from several sensors, multi-modal fusion approaches increase the accuracy of activity recognition and increase its resilience to sensor noise or malfunctions. Early fusion, late fusion, and attention-based fusion are fusion techniques that take advantage of each modality's advantages while resolving its drawbacks [18].

Context-Aware Modelling: Context-aware modeling techniques that take into account contextual data, such as user, environmental, or temporal context, can be advantageous for HAR systems. Context-aware models enhance the precision and customization of activity recognition by modifying their actions in response to contextual inputs. [57].

6.5 Propose Avenues for Future Research

For advancing t state-of-the-art and dealing with present issues in human activity recognition (HAR) with smartphones and deep learning approaches can be investigated in future research. Following are the possible ways:

On-device and Real-time Processing: In this we conduct HAR on smartphones in real-time, giving user's fast feedback and actionable information. This involve maintaining high accuracy and low latency while optimized deep learning models for deployment on mobile devices with less resources[70].
Privacy-Preserving HAR: The privacy-preserving HAR approaches that maintain correct activity recognition while also preserve user privacy. Edge computing, differential privacy, and federated learning techniques can be used to train models cooperatively across several devices without requiring the sharing of raw sensor data. This preserve user data [70].

Understanding Long-Term Activity: Explore different techniques for utilizing smartphone sensor data to model and comprehend long-term activity patterns and habits. The ability to capture temporal dependencies from long time period with deep learning models opens up new possibilities for personalized exercise suggestions, health monitoring, and behavior modification. This involve creating models has ability to recognized complex behavioral patterns transitions [26].

Multi-modal Sensor Fusion: Examine cutting-edge fusion methods that combine information from various smart phone sensor modalities, including Wi-Fi, GPS, accelerometer, and gyroscope. Correct activity identification, flexibility to environmental changes, and context awareness can all be enhanced by multi-modal fusion [34].

Transfer Learning and Domain Adaptation: For the improvement of model generalization across various user groups, contexts, and sensor variations, a lot of research should be done on transfer learning and domain adaptation methodologies. Model performance and adaptability can be increased by pretraining on large-scale datasets and fine-tuning them on smaller, domain-specific datasets. To handle domain shifts and variations found in real-world deployment circumstances, domain adaptation technique scan also be investigated [32].

Enhancing the interpretability and explainability of deep learning models for HAR is a priority in order to foster user acceptance and establish confidence [4]. Provided methods for displaying model predictions, attributing decision-making authority to specific sensor inputs, and giving more explanations for activity detection outcomes. This can helpful consumers understand how models work and make decisions based on their recommendations [4].

7 Conclusion

There are several conclusions about the use of deep learning studies in HAR tasks with wearable sensors and public datasets thata are revealed by the thorough literature study. From smartphone sensors the deep learning architectures like LSTM , RNN , CNN znd its variations have shown some remarkable efficiency in properly identifying a variety of human behaviours. These systems performs excellent in the usage of abundant data that is obtained by sensors like GPS, gyroscopes, and accelerometers to track human activity. Additionaly a number of variables, including feature extraction strategies, model architectures, preprocessing approaches, and dataset characteristics, have a significant impact on how well deep learning models perform in HAR. Open datasets with a variety of activity categories, sensor modalities, and annotation quality, such as WISDM, UCI HAR, and PAMAP2, have been extensively tested for HAR model evaluation and training. However the issues with class imbalance, noisy sensor data, and domains shifting between datasets can affect the generalization and the performance of the model. Additionally there are different promising paths to advance the state-of-the-art in HAR using smartphones and deep learning through emerging research directions like multi modal sensor fusion, model interpretability, privacy preserving HAR techniques, long-term activity understanding, transfer learning, and real-time processing on mobile devices. It is imperative that academics and practitioners take into account the following.

Dataset Selection and Preprocessing: To create reliable and broadly applicable HAR models, careful consideration goes into the selection of relevant datasets and preprocessing methods. Comprehending the attributes of the intended domain and resolving concerns related to data quality are crucial phases in this procedure.

Model Architecture and Training: Performance can be maximized by experimenting with various deep learning architectures, hyperparameters, and optimization techniques. Pre-trained models can be used with transfer learning and fine-tuning techniques to enhance generalization to new datasets or tasks.

Real-world Deployment Considerations: Computational efficiency, privacy protection, and interpretability are important considerations when deploying HAR models in real-world environments. Ensuring user privacy and trust and optimizing models for resource-constrained devices are critical.

Multidisciplinary Cooperation: Innovative solutions to challenging problems in human-machine interaction (HAR) can be achieved by cooperation between researchers in a variety of fields, including machine learning, signal processing, sensor technology, and human-computer interaction.

The synthesis of the work to date highlights, in conclusion, the great promise of deep learning techniques for HAR with cellphones and public datasets. Researchers and practitioners may create more precise, reliable, and user-friendly HAR systems that support a range of applications, such as context-aware computing, fitness tracking, and healthcare monitoring, by utilizing the insights from this review and exploring new research avenues.

References

[1] Israr Akhter, Naif Al Mudawi, Bayan Ibrahimm Alabdullah, Mohammed Alonazi, and Jeongmin Park. Human-based interaction analysis via automated key point detection and neural network model. *IEEE Access*, 11:100646–100658, 2023.

[2] Sana Alazwari, Majdy M. Eltahir, Nabil Sharaf Almalki, Abdulrahman Alzahrani, Mrim M. Alnfiai, and Ahmed S. Salama. Improved coyote optimization algorithm and deep learning driven activity recognition in healthcare. *IEEE Access*, 12:22158–22166, 2024.

[3] Tsige Tadesse Alemayoh, Jae Hoon Lee, and Shingo Okamoto. New sensor data structuring for deeper feature extraction in human activity recognition. *Sensors*, 21(8):2814, 2021.

[4] Michele Alessandrini, Giorgio Biagetti, Paolo Crippa, Laura Falaschetti, and Claudio Turchetti. Recurrent neural network for human activity recognition in embedded systems using ppg and accelerometer data. *Electronics*, 10(14):1715, 2021.

[5] Ankita, Shalli Rani, Himanshi Babbar, Sonya Coleman, Aman Singh, and Hani Moaiteq Aljahdali. An efficient and lightweight deep learning model for human activity recognition using smartphones. *Sensors*, 21(11):3845, 2021.

[6] Tunç Aşuroğlu. Complex human activity recognition using a local weighted approach. *IEEE Access*, 10:101207–101219, 2022.

[7] Sriman Bidhan Baray, Mosabber Uddin Ahmed, Muhammad E. H. Chowdhury, and Koichi Kise. Eog-based reading detection in the wild using spectrograms and nested classification approach. *IEEE Access*, 11:105619–105632, 2023.

[8] Zakaria Benhaili, Youness Abouqora, Youssef Balouki, and Lahcen Moumoun. Basic activity recognition from wearable sensors using a lightweight deep neural network. *Journal of ICT Standardization*, 10(2):241–260, 2022.

[9] Debarshi Bhattacharya, Deepak Sharma, Wonjoon Kim, Muhammad Fazal Ijaz, and Pawan Kumar Singh. Ensem-har: An ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. *Biosensors*, 12(6):393, 2022.

[10] Debarshi Bhattacharya, Deepak Sharma, Wonjoon Kim, Muhammad Fazal Ijaz, and Pawan Kumar Singh. Ensem-har: An ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring. *Biosensors*, 12(6):393, 2022.

[11] Héctor Sánchez San Blas, André Sales Mendes, Daniel H. de la Iglesia, Luís Augusto Silva, and Gabriel Villarrubia González. A multiagent platform for promoting physical activity and learning through interactive educational games using the depth camera recognition system. *Entertainment Computing*, 49:100629, 2024.

[12] Marius Bock, Alexander Hölzemann, Michael Moeller, and Kristof Van Laerhoven. Improving deep learning for har with shallow lstms. In *Proceedings of the 2021 ACM International Symposium on Wearable Computers*, pages 7–12, 2021.

[13] Luis Felipe Borja-Borja, Jorge Azorín-López, Marcelo Saval-Calvo, Andrés Fuster-Guilló, and Marc Sebban. Architecture for automatic recognition of group activities using local motions and context. *IEEE Access*, 10:79874–79889, 2022.

[14] Yeong-Hyeon Byeon, Dohyung Kim, Jaeyeon Lee, and Keun-Chang Kwak. Ensemble three-stream rgb-s deep neural network for human behavior recognition under intelligent home service robot environments. *IEEE Access*, 9:73240–73250, 2021.

[15] Francisco M. Calatrava-Nicolás and Oscar Martinez Mozos. Light residual network for human activity recognition using wearable sensor data. *IEEE Sensors Letters*, 7(10):1–4, 2023.

[16] Stefania Cristina, Vladimir Despotovic, Rodrigo Pérez-Rodríguez, and Slavisa Aleksic. Audio- and video-based human activity recognition systems in healthcare. *IEEE Access*, 12:8230–8245, 2024.

[17] Thai Ha Dang, Lionel Nkenyereye, Viet-Thang Tran, and Wan-Young Chung. Self-powered cattle behavior monitoring system using 915 mhz radio frequency energy harvesting. *IEEE Access*, 12:33779–33791, 2024.

[18] Vahid Farrahi, Usman Muhammad, Mehrdad Rostami, and Mourad Oussalah. Accnet24: A deep learning framework for classifying 24-hour activity behaviours from wrist-worn accelerometer data under free-living environments. *International Journal of Medical Informatics*, 172:105004, 2023.

[19] Daniel Garcia-Gonzalez, Daniel Rivero, Enrique Fernandez-Blanco, and Miguel R. Luaces. Deep learning models for real-life human activity recognition from smartphone sensor data. *Internet of Things*, 24:100925, 2023.

[20] Ananda Mohon Ghosh and Katarina Grolinger. Edge-cloud computing for internet of things data analytics: Embedding intelligence in the edge with deep learning. *IEEE Transactions on Industrial Informatics*, 17(3):2191–2200, 2021.

[21] Christian Gianoglio, Ammar Mohanna, Ali Rizik, Laurence Moroney, and Maurizio Valle. On edge human action recognition using radar-based sensing and deep learning. *IEEE Transactions on Industrial Informatics*, 20(3):4160–4172, 2024.

[22] Saurabh Gupta. Deep learning based human activity recognition (har) using wearable sensor data. *International Journal of Information Management Data Insights*, 1(2):100046, 2021.

[23] Sadaf Hafeez, Saud S. Alotaibi, Abdulwahab Alazeb, Naif Al Mudawi, and Wooseong Kim. Multi-sensor-based action monitoring and recognition via hybrid descriptors and logistic regression. *IEEE Access*, 11:48145–48157, 2023.

[24] Nguyen Thi Hoai Thu and Dong Seog Han. Hihar: A hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition. *IEEE Access*, 9:145271–145281, 2021.

[25] Yu Min Hwang, Sangjun Park, Hyung Ok Lee, Seok-Kap Ko, and Byung-Tak Lee. Deep learning for human activity recognition based on causality feature extraction. *IEEE Access*, 9:112257–112275, 2021.

[26] Ismael Espinoza Jaramillo, Jin Gyun Jeong, Patricio Rivera Lopez, Choong-Ho Lee, Do-Yeon Kang, Tae-Jun Ha, Ji-Heon Oh, Hwanseok Jung, Jin Hyuk Lee, Won Hee Lee, et al. Real-time human activity recognition with imu and encoder sensors in wearable exoskeleton robot via deep learning networks. *Sensors*, 22(24):9690, 2022.

[27] Ali Olow Jimale and Mohd Halim Mohd Noor. Fully connected generative adversarial network for human activity recognition. *IEEE Access*, 10:100257–100266, 2022.

[28] Michail Kaseris, Ioannis Kostavelis, and Sotiris Malassiotis. A comprehensive survey on deep learning methods in human activity recognition. *Machine Learning and Knowledge Extraction*, 6(2):842–876, 2024.

[29] Habib-Ur-Rehman Khalid, Ali Gorji, André Bourdoux, Sofie Pollin, and Hichem Sahli. Multi-view cnn-lstm architecture for radar-based human activity recognition. *IEEE Access*, 10:24509–24519, 2022.

[30] Imran Ullah Khan, Sitara Afzal, and Jong Weon Lee. Human activity recognition via hybrid deep learning based model. *Sensors*, 22(1):323, 2022.

[31] Mst. Alema Khatun, Mohammad Abu Yousuf, Sabbir Ahmed, Md. Zia Uddin, Salem A. Alyami, Samer Al-Ashhab, Hanan F. Akhdar, Asaduzzaman Khan, Akm Azad, and Mohammad Ali Moni. Deep cnn-lstm with self-attention model for human activity recognition using wearable sensor. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–16, 2022.

[32] Yeon-Wook Kim, Woo-Hyeong Cho, Kyu-Sung Kim, and Sangmin Lee. Inertial-measurement-unit-based novel human activity recognition algorithm using conformer. *Sensors*, 22(10):3932, 2022.

[33] Weiheng Kong, Lili He, and Hailong Wang. Exploratory data analysis of human activity recognition based on smart phone. *IEEE Access*, 9:73355–73364, 2021.

[34] Emanuele Lattanzi, Lorenzo Calisti, and Paolo Capellacci. Lightweight accurate trigger to reduce power consumption in sensor-based continuous human activity recognition. *Pervasive and Mobile Computing*, 96:101848, 2023.

[35] Minjung Lee and Seoung Bum Kim. Sensor-based open-set human activity recognition using representation learning with mixup triplets. *IEEE Access*, 10:119333–119344, 2022.

[36] Won-Seon Lim, Wangduk Seo, Dae-Won Kim, and Jaesung Lee. Efficient human activity recognition using lookup table-based neural architecture search for mobile devices. *IEEE Access*, 11:71727–71738, 2023.

[37] Limeng Lu, Chuanlin Zhang, Kai Cao, Tao Deng, and Qianqian Yang. A multichannel cnn-gru model for human activity recognition. *IEEE Access*, 10:66797–66810, 2022.

[38] Neelakandan Mani, Prathap Haridoss, and Boby George. Evaluation of a combined conductive fabric-based suspender system and machine learning approach for human activity recognition. *IEEE Open Journal of Instrumentation and Measurement*, 2:1–10, 2023.

[39] Matthew P. Mavor, Victor C. H. Chan, Kristina M. Gruevski, Linda L. M. Bossi, Thomas Karakolis, and Ryan B. Graham. Assessing the soldier survivability tradespace using a single imu. *IEEE Access*, 11:69762–69772, 2023.

[40] Sakorn Mekruksavanich, Narit Hnoohom, and Anuchit Jitpattanakul. A hybrid deep residual network for efficient transitional activity recognition based on wearable sensors. *Applied Sciences*, 12(10):4988, 2022.

[41] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Lstm networks using smartphone data for sensor-based human activity recognition in smart homes. *Sensors*, 21(5):1636, 2021.

[42] Sakorn Mekruksavanich and Anuchit Jitpattanakul. Sport-related activity recognition from wearable sensors using bidirectional gru network. *Intelligent Automation & Soft Computing*, 34(3), 2022.

[43] Sakorn Mekruksavanich and Anuchit Jitpattanakul. A deep learning network with aggregation residual transformation for human activity recognition using inertial and stretch sensors. *Computers*, 12(7):141, 2023.

[44] Sakorn Mekruksavanich, Wikanda Phaphan, Narit Hnoohom, and Anuchit Jitpattanakul. Attention-based hybrid deep learning network for human activity recognition using wifi channel state information. *Applied Sciences*, 13(15):8884, 2023.

[45] Saeed Mohsen, Ahmed Elkaseer, and Steffen G. Scholz. Industry 4.0-oriented deep learning models for human activity recognition. *IEEE Access*, 9:150508–150521, 2021.

[46] Hassan Mohyuddin, Syed Kumayl Raza Moosavi, Muhammad Hamza Zafar, and Filippo Sanfilippo. A comprehensive framework for hand gesture recognition using hybrid-metaheuristic algorithms and deep learning models. *Array*, 19:100317, 2023.

[47] Parisa Fard Moshiri, Reza Shahbazian, Mohammad Nabati, and Seyed Ali Ghorashi. A csi-based human activity recognition using deep learning. *Sensors*, 21(21):7225, 2021.

[48] Md. Muktadir Mukto, Mahamudul Hasan, Md. Maiyaz Al Mahmud, Ikramul Haque, Md. Ahsan Ahmed, Taskeed Jabid, Md. Sawkat Ali, Mohammad Rifat Ahmmad Rashid, Mohammad Manzurul Islam, and Maheen Islam. Design of a real-time crime monitoring system using deep learning techniques. *Intelligent Systems with Applications*, 21:200311, 2024.

[49] Ohoud Nafea, Wadood Abdul, Ghulam Muhammad, and Mansour Alsulaiman. Sensor-based human activity recognition with spatio-temporal deep learning. *Sensors*, 21(6):2141, 2021.

[50] Heba Nematallah and Sreeraman Rajan. Adaptive hierarchical classification for human activity recognition using inertial measurement unit (imu) time-series data. *IEEE Access*, 12:52127–52149, 2024.

[51] A. Nouriani, R. McGovern, and R. Rajamani. Activity recognition using a combination of high gain observer and deep learning computer vision algorithms. *Intelligent Systems with Applications*, 18:200213, 2023.

[52] Patricia O'Sullivan, Matteo Menolotto, Andrea Visentin, Brendan O'Flynn, and Dimitrios-Sokratis Komaris. Ai-based task classification with pressure insoles for occupational safety. *IEEE Access*, 12:21347–21357, 2024.

[53] Grzegorz Pajak, Pascal Krutz, Justyna Patalas-Maliszewska, Matthias Rehm, Iwona Pajak, and Martin Dix. An approach to sport activities recognition based on an inertial sensor and deep learning. *Sensors and Actuators A: Physical*, 345:113773, 2022.

[54] Ariza-Colpas Paola Patricia, Pacheco-Cuentas Rosberg, Shariq Butt-Aziz, Piñeres-Melo Marlon Alberto, Morales-Ortega Roberto-Cesar, Urina-Triana Miguel, and Sumera Naz. Semi-supervised ensemble learning for human activity recognition in casas kyoto dataset. *Heliyon*, 10(8):e29398, 2024.

[55] Gundala Jhansi Rani, Mohammad Farukh Hashmi, and Aditya Gupta. Surface electromyography and artificial intelligence for human activity recognition—a systematic review on methods, emerging trends applications, challenges, and future implementation. *IEEE Access*, 11:105140–105169, 2023.

[56] Mutegeki Ronald, Alwin Poulose, and Dong Seog Han. isplinception: an inception-resnet deep learning architecture for human activity recognition. *IEEE Access*, 9:68985–69001, 2021.

[57] Mutegeki Ronald, Alwin Poulose, and Dong Seog Han. isplinception: An inception-resnet deep learning architecture for human activity recognition. *IEEE Access*, 9:68985–69001, 2021.

[58] Oumaima Saidani, Majed Alsafyani, Roobaea Alroobaea, Nazik Alturki, Rashid Jahangir, and Leila Jamel. An efficient human activity recognition using hybrid features and transformer model. *IEEE Access*, 11:101373–101386, 2023.

[59] Sowmiya Seenath and Menaka Dharmaraj. Conformer-based human activity recognition using inertial measurement units. *Sensors*, 23(17):7357, 2023.

[60] Tamer Shanableh. Vico-moco-dl: Video coding and motion compensation solutions for human activity recognition using deep learning. *IEEE Access*, 11:73971–73981, 2023.

[61] Meng Shang, Lenore Dedeyne, Jolan Dupont, Laura Vercauteren, Nadjia Amini, Laurence Lapauw, Evelien Gielen, Sabine Verschueren, Carolina Varon, Walter De Raedt, and Bart Vanrumste. Otago exercises monitoring for older adults by a single imu and hierarchical machine learning models. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:462–471, 2024.

[62] Yoli Shavit and Itzik Klein. Boosting inertial-based human activity recognition with transformers. *IEEE Access*, 9:53540–53547, 2021.

[63] Young Ghyu Sun, Soo Hyun Kim, Seongwoo Lee, Joonho Seon, SangWoon Lee, Cheong Ghil Kim, and Jin Young Kim. Performance of end-to-end model based on convolutional lstm for human activity recognition. *Journal of Web Engineering*, 21(5):1671–1689, 2022.

[64] Felix Tempel, Espen Alexander F. Ihlen, and Inga Strümke. Autogcn-toward generic human activity recognition with neural architecture search. *IEEE Access*, 12:39505–39516, 2024.

[65] Hoang-Nhat Tran, Hong-Quan Nguyen, Huong-Giang Doan, Thanh-Hai Tran, Thi-Lan Le, and Hai Vu. Pairwise-covariance multi-view discriminant analysis for robust cross-view human action recognition. *IEEE Access*, 9:76097–76111, 2021.

[66] Shahab Uddin, Tahir Nawaz, James Ferryman, Nasir Rashid, Md. Asaduzzaman, and Raheel Nawaz. Skeletal keypoint-based transformer model for human action recognition in aerial videos. *IEEE Access*, 12:11095–11103, 2024.

[67] Sahil Waqar, Muhammad Muaaz, and Matthias Pätzold. Direction-independent human activity recognition using a distributed mimo radar system and deep learning. *IEEE Sensors Journal*, 23(20):24916–24929, 2023.

[68] Sahil Waqar and Matthias Pätzold. A simulation-based framework for the design of human activity recognition systems using radar sensors. *IEEE Internet of Things Journal*, 11(8):14494–14507, 2024.

[69] James Wensel, Hayat Ullah, and Arslan Munir. Vit-ret: Vision and recurrent transformer neural networks for human activity recognition in videos. *IEEE Access*, 11:72227–72249, 2023.

[70] Bingxue Zhang, Guangguang Lu, Pengpeng Qiu, Xumin Gui, and Yang Shi. Advancing federated learning through verifiable computations and homomorphic encryption. *Entropy*, 25(11):1550, 2023.

[71] Junjie Zhang, Yuanhao Liu, and Hua Yuan. Attention-based residual bilstm networks for human activity recognition. *IEEE Access*, 11:94173–94187, 2023.