Department of Computer Science
Texas State University

Time series classification using Activity Recognition system based on Multisensor data fusion (AReM)

Habib Irani
habibirani@txstate.edu

Arshia Kermani
arshia.kermani@txstate.edu

1 Introduction

The realm of human activity recognition is pivotal for advancing how technology interacts with human behavior. Sensors embedded in daily wearables like smartwatches and fitness bands are constantly collecting data about our physical activities. The challenge is not just collecting this vast amount of data, but also effectively interpreting it to provide useful feedback and insights. This task becomes complex due to the varied and dynamic nature of human activities.

Effective activity recognition requires not only identifying static poses but also dynamically changing activities and transitions between them. This requires sophisticated models that can process time-dependent data, recognize patterns, and learn from sequences of sensor readings. The objective is to move beyond mere data collection, transforming raw sensor outputs into intelligent, context-aware systems that enhance user experience in healthcare monitoring, personal fitness, and beyond.

By focusing on the AReM dataset and applying deep learning techniques, this project seeks to tackle these challenges head-on, developing solutions that could lead to more intuitive and responsive technology.

The project adopts deep learning techniques to improve the sophistication of activity recognition. These techniques are particularly suited to handling the sequence-dependent complexities of sensor data, which often involve nonlinearities and temporal dependencies that traditional machine learning models might struggle with. By leveraging architectures like CNNs for spatial pattern recognition, the project aims to push the boundaries of what can be achieved with current HAR technologies.

This project's primary objective is to refine deep learning models for more accurate, efficient, and robust activity recognition. Key goals include identifying the best neural network architectures for this application, optimizing models to enhance their learning efficacy with less computational cost, and testing the models' ability to generalize across different types of activities and environmental conditions. Moreover, the project seeks to contribute to the academic and practical understanding of sensor-based activity recognition systems, setting a benchmark for future research and application in non-tabular data handling in deep learning frameworks.

2 Related Work

Human Activity Recognition (HAR) using sensor data has become a critical area of research with extensive applications in healthcare, sports, security, and smart environments. By leveraging machine learning models and deep learning architectures, researchers aim to accurately classify and predict human activities based on data captured through various sensors embedded in wearable devices or environments. The UC Irvine Machine Learning Repository's AReM dataset, which consists of sensor data for different activities, has been instrumental in advancing research in this field. This dataset provides a valuable benchmark for evaluating the effectiveness of various machine learning techniques and deep learning models in recognizing complex patterns and activities from time-series sensor data.

Biswas and Samanta [1] employed the AReM dataset to develop an ensemble random forest model for anomaly detection in wireless sensor networks. Their methodology integrated Decision Trees, Naive Bayes, and K-Nearest Neighbors as base learners to enhance detection accuracy. The model demonstrated superior performance over individual base learners by effectively identifying anomalous data, indicating the robustness of ensemble methods in handling complex sensor data environments.

In another study [2], a multi-task self-supervised learning approach was employed to effectively utilize wearable sensor data for HAR. This method involved training on large-scale datasets to pre-train models that were then fine-tuned on specific HAR tasks. The approach proved beneficial for improving the generalization capabilities of HAR models across different datasets and conditions, showcasing the potential of self-supervised learning in developing versatile and effective HAR systems.

Tan et al. [4] utilized the AReM dataset to implement a many-to-one architecture with GRUs, focusing on elderly activity monitoring. This method achieved an impressive accuracy of 97.14%, demonstrating the effectiveness of GRUs in classifying complex time series data from wearable sensors.

Tehrani et al. [5] explored the AReM dataset alongside others for HAR using Bi-LSTM, which involves learning both forward and reverse sequences of sensor data. Their approach, which also included a novel postprocessing step through windowing and voting, resulted in high F1 scores, underscoring the Bi-LSTM's capability to enhance activity detection accuracy significantly.

Yang et al. [6]focused on enhancing human activity recognition through ensemble learning methods that combine random forests and neural networks. Using the AReM dataset, their approach aimed to improve recognition accuracy by leveraging the strengths of both learning models. This integration showcased significant improvements in predictive performance, highlighting the dataset's suitability for developing advanced HAR systems that require robust and accurate classification capabilities.

Yuan et al. [7] advanced the field of HAR by implementing a self-supervised learning framework that utilized a massive dataset from the UK Biobank. This approach allowed the construction of robust deep learning models capable of generalizing across various datasets, significantly enhancing activity recognition accuracy. Their study demonstrated how leveraging large, unlabeled datasets could overcome the limitations posed by smaller, less diverse training sets commonly used in HAR.

3 Dataset Description

The AReM dataset [3], central to this project, serves as a comprehensive data source tailored for developing and enhancing activity recognition systems using multi-sensor data fusion techniques. This dataset is particularly designed for testing and deploying sophisticated deep learning frameworks, due to its structured yet complex multi-dimensional sensory data. The data is obtained from both wearable sensors on the human body and environmental sensors, providing a rich multi-contextual view of human activities.

Activities recorded include bending, lying down, sitting, standing, walking, and cycling, each with distinct sensor signatures captured through accelerometers and environmental sensors like RSS (Received Signal Strength) indicators. The dataset's inclusion of different settings-controlled environments like labs and more dynamic, uncontrolled settings-enriches the data's variability and robustness, which is crucial for building adaptive and resilient activity recognition systems.

The dataset was compiled with the intent of supporting not just generic activity recognition but also complex scenarios that involve data fusion and real-time processing demands. It allows for the exploration of various machine learning strategies, particularly those that can manage sequential and time-series data effectively. The integration of multi-sensor data fusion in this dataset underscores the complexities and potential of real-world applications in smart homes and ambient assisted living environments, where accurate activity recognition can significantly impact user care and interaction.

The sitting activity plot shows a remarkably stable and consistent pattern in the RSS values over time. The mean RSS between beacons, indicating proximity, remains relatively unchanged, suggesting the participant maintains a steady position relative to the beacons. This is corroborated by the low variance in RSS readings, indicative of minimal movement, which is characteristic of a sitting posture.

Contrastingly, the bending activity plot reveals greater variability in RSS readings. This activity, characterized by dynamic movement and changes in orientation to the beacons, leads to more pronounced fluctuations in the mean and variance of RSS. Such variability is reflective of the complex nature of bending movements, where the participant's distance from the beacons is in constant flux.

Further insight is drawn from a tabular summary of the RSS features, which encapsulates the average readings and their variance, offering a snapshot of the dataset's central tendencies and dispersion. This summary aids in differentiating between the activities based on sensor data characteristics.

A detailed breakdown of the bending dataset in another table, segmented into time intervals, presents an even more granular perspective. It showcases how RSS values evolve over shorter periods, capturing the essence of movement patterns within these segmented windows.

The contrasting stability in sitting versus the variability in bending provides valuable information for feature extraction in machine learning models.

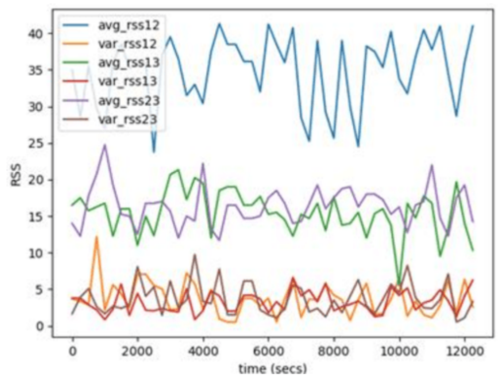

Figure 1: Plot of means and variances of RSS of sitting activity

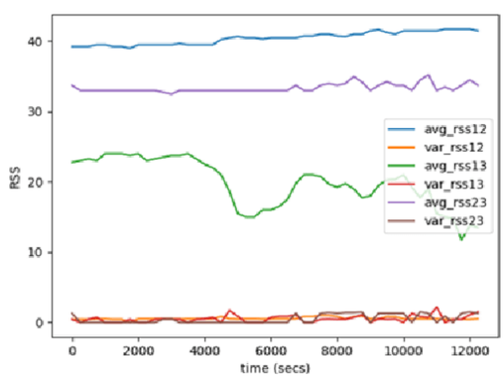

Figure 2: Plot of means and variances of RSS of bending 1 activity

Table 1: Features of dataset

	Mean	Variance
Beacon 1	avg_rss12	var_rss12
Beacon 2	avg_rss13	var_rss13
Beacon 3	avg_rss23	var_rss23

Table 2: Snippet of bending-1 dataset

Time interval (ms)	$\overline{x}_{rss12}$	var_{rss12}	$\overline{x}_{rss13}$	var_{rss13}	$\overline{x}_{rss23}$	var_{rss23}
0-250	43.67	0.47	24.75	0.43	30	0
250-500	43.33	0.47	25.33	0.43	30	0
500-750	42.75	0.83	25.25	0.83	30.5	0.5

4 Methodology

4.1 Data Preprocessing

The preprocessing of the AReM dataset was meticulously structured to optimize the data for use in deep learning models aimed at human activity recognition. This process is critical as it prepares the data for efficient and effective analysis by the models, ensuring that the data's inherent qualities are maintained and enhanced to support accurate activity classification.

Data loading was the initial step in the preprocessing pipeline. Here, data was methodically extracted from directories corresponding to specific activities, such as bending and walking. Each directory contained CSV files that were concatenated into a single DataFrame. This phase included rigorous error handling to manage and mitigate issues arising from corrupted files or load failures, ensuring that the subsequent stages of the data processing pipeline had a solid foundation of clean and reliable data.

Following data loading, the next crucial step was data cleaning and preparation. This stage involved a detailed examination of the data to remove non-numeric values and strings that could interfere with the analysis. Each cell in the dataset was inspected to verify its data type, and any row containing inappropriate data types was excluded to maintain the integrity of the dataset. Furthermore, rows containing NaN values—often a result of incomplete data capture or errors in data entry—were removed. This cleaning process is essential to ensure the consistency and reliability of the input data, as the presence of such anomalies could skew the results and lead to inaccurate model predictions.

Feature standardization was then applied to the cleaned data, utilizing the StandardScaler to normalize the features related to sensor data, such as 'avg-rss12' and 'var-rss12'. Standardization is a critical step in preparing the data for neural network training as it normalizes feature values to have a mean of zero and a standard deviation of one. This normalization helps in mitigating the problem of varying scales among different features, which can significantly influence the performance of the model, particularly in algorithms that are sensitive to feature scaling.

Label encoding was another crucial step in the data preprocessing phase. Activity labels, which were initially in string format, were converted to integer codes using the LabelEncoder. This transformation is vital for the neural network to perform multi-class classification, as it allows the model to effectively interpret and process the labels during training and predictions.

Finally, the prepared data was converted into sequences of 50 time steps each, aligning with the input requirements of LSTM networks renowned for their ability to process sequential and temporal information. This transformation is crucial for models that rely on understanding the temporal dynamics in the data, such as LSTM networks, enabling them to capture the sequence-dependent characteristics of human activities effectively.

4.2 Deep Learning Model Architecture

4.2.1 LSTM-Based Model

In the pursuit of advancing human activity recognition (HAR), we opted for an LSTM (Long Short-Term Memory) architecture, renowned for its ability to effectively capture and learn from sequences, making it particularly suited to discerning the temporal patterns inherent in activity data. The LSTM's proficiency in recognizing long-term dependencies is crucial in contexts where the sequence and timing of events are predictive of the current state.

To enhance the robustness of our LSTM model and safeguard against overfitting, dropout layers were strategically incorporated within the LSTM layers at a rate of 0.5. This approach randomly deactivates a subset of neurons in each training iteration, which not only prevents the network from becoming overly dependent on any single neuron but also promotes a more generalized model that performs well on unseen data.

The architecture further refines the LSTM's outputs through a fully connected layer designed to compress the extracted features into a more manageable form that directly corresponds to the number of activity classes. This dense layer plays a pivotal role in translating the temporal features identified by the LSTM into specific activity classifications, effectively bridging the gap between raw data interpretation and practical application.

The dataset was divided into training, validation, and testing segments. This structured segmentation supports comprehensive training and robust validation, ensuring that the model's performance is consistently evaluated against unbiased, unseen data. Utilizing the Adam optimizer, the model benefits from an adaptive learning rate mechanism that optimizes the convergence speed towards the most effective weights for the task at hand.

The training process spanned epochs to provide the LSTM enough iterations to adequately learn from the data without the risk of overfitting. Throughout this phase, key metrics such as loss and accuracy were meticulously monitored. These metrics serve as critical indicators of the model's ability to correctly classify activities, offering insights that guide ongoing adjustments to improve training outcomes.

Cross-Entropy Loss, chosen for its efficacy in multi-class classification problems, was employed to quantify the difference between predicted probabilities and actual class labels, further refining the model's accuracy.

The validation phase, integral to the training process, was crucial for fine-tuning the model. By providing regular feedback on the model's performance on validation data, it enabled timely adjustments to hyperparameters, thus enhancing the model's accuracy and preventing overfitting.

4.2.2 CNN Model

The Convolutional Neural Network (CNN) model we have crafted for human activity recognition (HAR) is built with a sophisticated architecture specifically designed to detect and classify the intricate patterns of physical movement captured through time-series sensor data. The model incorporates three convolutional layers, each with an increasing number of filters—32, 64, and 128 respectively. This configuration allows the network to extract a broad range of features from simple to complex, enhancing its ability to interpret various activities. Each convolutional layer uses a kernel size of three with padding, which helps in capturing immediate temporal relationships without reducing the output dimensionality prematurely.

Following each convolutional layer, a max pooling layer with a stride of two reduces the spatial dimensions of the output by half. This downsampling is essential for decreasing the computational load, minimizing overfitting, and retaining only the most significant features detected by the convolutional filters. The pooled features are then flattened and processed through two dense layers. The first, consisting of 256 units, compresses the feature map into a form suitable for classification, while the second reduces the dimensions to match the number of activity classes.

The training of the CNN model employs the Cross-Entropy Loss function and the Adam optimizer, chosen for its adaptive learning rate capabilities, which facilitate faster and more effective convergence by adjusting the learning rate based on how training progresses. The model is trained over multiple epochs, with each epoch involving complete passes of the entire dataset through both forward and backward propagations to update the model weights effectively. DataLoader is used for batch processing, introducing stochasticity into the learning process that helps escape local minima during gradient descent.

At the end of each epoch, the model undergoes a validation phase to evaluate its performance on unseen data, which is crucial for monitoring the model's ability to generalize beyond the training data. This validation phase provides feedback on the model's performance, using metrics such as loss and accuracy, and informs any necessary adjustments to the training parameters. This could include learning rate adjustments or early stopping if validation performance begins to degrade.

The architecture and training strategy are closely aligned to ensure that each component contributes effectively to the final task of activity recognition. The integration of convolutional and pooling layers allows the network to efficiently extract and condense features from raw sensor data, which are then mapped to specific activity classes by the dense layers. This structured approach ensures that the CNN model not only learns the training data effectively but also generalizes well to new, unseen data.

4.2.3 Transformer Model

In advancing our Transformer model for human activity recognition, we have meticulously integrated a suite of methodologies particularly well-suited to the nuanced challenges of time-series data analysis. We began by establishing the TimeSeriesPatchEmbeddingLayer, which segments the data into patches. This technique reflects an intuitive approach to processing continuous information streams and is particularly useful in the context of time-series data, allowing the model to process segments of the data as discrete elements. This technique is advantageous for handling large volumes of time-series data efficiently in batch processing.

The embedding layer in our model is further enhanced with positional encodings. This crucial feature compensates for the Transformer's inherent lack of sequential processing, enabling the model to preserve the order and context of the activity data, which is vital for classifying sequences based on sensor readings. Our PositionalEncoding module introduces a unique marker to each time step, endowing the model with the capability to detect temporal patterns and dependencies within the activity data.

Our Transformer architecture includes multiple layers of TransformerEncoder, each featuring a self-attention mechanism and a feed-forward neural network. The self-attention mechanism's capacity to process all sequence positions concurrently allows the model to emphasize the significance of each segment of the input data, thereby concentrating more on the sections that are most relevant for activity recognition.

The TransformerEncoder's outputs are further processed by a feed-forward layer and culminate in a classifier that assigns the activity class. We've carefully selected the number of attention heads and set the dimensions of the feed-forward network to ensure a fine balance between the model's learning capacity and computational demands.

During the training process, our model undergoes a series of epochs, aiming to minimize Cross-Entropy Loss. Loss and accuracy metrics are closely observed, providing essential feedback for model optimization. A learning rate scheduler is employed to dynamically adjust the rate during training, enhancing the model's learning trajectory.

The incorporation of these methodological advancements, including batch embedding and the strategic setting of embedding dimensions and a fixed number of attention heads, ensures that our Transformer model is not only effective in learning from training data but also adept at generalizing to new, unseen data. This comprehensive approach has endowed our model with robust predictive capabilities, as evidenced by high accuracy rates on test data, highlighting its effectiveness in utilizing both spatial and temporal information inherent in sensor data, and securing its position as a formidable tool for HAR. The model's ability to adapt to data characteristics, unencumbered by the constraints of sequence order, offers a significant advantage over more traditional modeling approaches.

5 Experiments and Results

In the realm of human activity recognition (HAR), the robustness and generalizability of models are paramount due to the inherent variability and complexity of sensor data. During the experimental phase, it became apparent that the AReM dataset had undergone prior preprocessing, indicating that the data was not in its raw, original form. This dataset has been organized into several folders, each containing CSV files corresponding to different activity classes such as bending and walking. This organization of data suggests a level of initial data arrangement which could influence the performance and evaluation of machine learning models trained on this dataset. This discovery suggested the presence of data leakage, which could artificially inflate the performance metrics of predictive models, as observed in various papers that referenced this dataset, achieving exceptionally high accuracy rates.

In an effort to address this issue and aim for a genuine assessment of model performance, we employed data augmentation techniques. Specifically, we integrated controlled noise into the dataset to simulate the natural variability encountered in raw sensor data. This step was taken to enhance the generalizability of the models and mitigate the risk of overfitting, ensuring that the models developed would be robust and capable of performing accurately on truly unseen data, reflective of real-world scenarios. To ensure a thorough evaluation and effective training, the dataset was methodically divided into 60% training, 20% validation, and 20% testing segments. We used the Adam optimizer with a learning rate of 0.001. We introduced Gaussian noise with a standard deviation of 0.1 to mimic real-world sensor noise, serving as a regularization technique to prevent overfitting by injecting randomness and helping the model learn more generalizable patterns. Additionally, we employed time shifting by randomly adjusting the data by up to five time steps to simulate potential delays or early occurrences in sensor data recording, which enhances the model's temporal flexibility. To ensure uniform input dimensions, shorter sequences were padded with zeros, and longer sequences were trimmed to a standard length of 50 time steps.

The augmented data was then integrated with the original dataset, effectively doubling the volume of training data. This richer dataset provided the model with a broader array of examples, embodying a wider array of scenarios that might occur in practical applications and ensuring high performance even with real-world, imperfect data. The impact of augmentation was quantitatively significant, particularly noted in the LSTM model which initially displayed lower accuracy and higher validation losses, indicating potential overfitting to the training data. However, with augmented data, there was noticeable improvement in both training and validation metrics, showing quicker convergence and enhanced generalization, evidenced by reduced validation losses and increased accuracy in early epochs.

Comparing the LSTM, CNN, and Transformer models, the Transformer demonstrated superior performance, achieving approximately 98.4% accuracy with augmentation, compared to around 96.9% for the LSTM model. This superior performance can be attributed to the Transformer's advanced attention mechanisms and positional encoding, which effectively capture and interpret complex dependencies in data irrespective of their position within the sequence.

Post-augmentation improvements were particularly marked in precision, recall, and F1-score for each activity class, especially for activities like 'sitting' and 'standing' which previously showed lower accuracy. These improvements underscore the efficacy of incorporating realistic noise and temporal variations into the training datasets.

The use of data augmentation significantly bolstered the models' ability to handle diverse and challenging datasets, leading to more robust and adaptable systems. We utilized the tsai library to facilitate the comparison of various deep learning methods for human activity recognition. This comprehensive library, specifically tailored for time series analysis, provided us with advanced tools and pre-built models that enabled efficient implementation and evaluation of different architectures.

The results reported represent the average outcomes of ten independent runs of the experimental code. This approach ensures the reliability and stability of the reported performance metrics, mitigating the influence of random variations and providing a robust basis for evaluating the efficacy of the models.

Table 3: This table presents a comprehensive comparison of the average prediction accuracy percentages for the proposed method, CNN, LSTM, and Transformer-encoder models. It includes the standard deviation (Std. Dev.) for the proposed method to indicate result variability and 95% confidence intervals (95% CI) for all methods, providing a statistical perspective on the precision of the accuracy measurements.

Dataset	**Transformer-encoder**		**CNN**		**LSTM**	
	Acc. %	**95% CI**	**Acc. %**	**95% CI**	**Acc. %**	**95% CI**
AReM-original	97.3	[96.0, 98.1]	95.4	[94.3, 96.2]	96.0	[95.2, 96.8]
AReM-augmentation	98.4	[97.2, 99.1]	96.4	[95.4, 97.2]	96.9	[95.8, 97.6]

6 Discussion

It is crucial to address the comparative performance of the Transformer, LSTM, and CNN models used in our experiments. The Transformer model exhibited superior accuracy compared to LSTM and CNN models. This enhanced performance can be attributed to its distinctive use of positional encodings and a sophisticated attention mechanism. Positional encodings inject sequence information into the model, allowing the Transformer to consider the order of events in time-series data, which is absent in

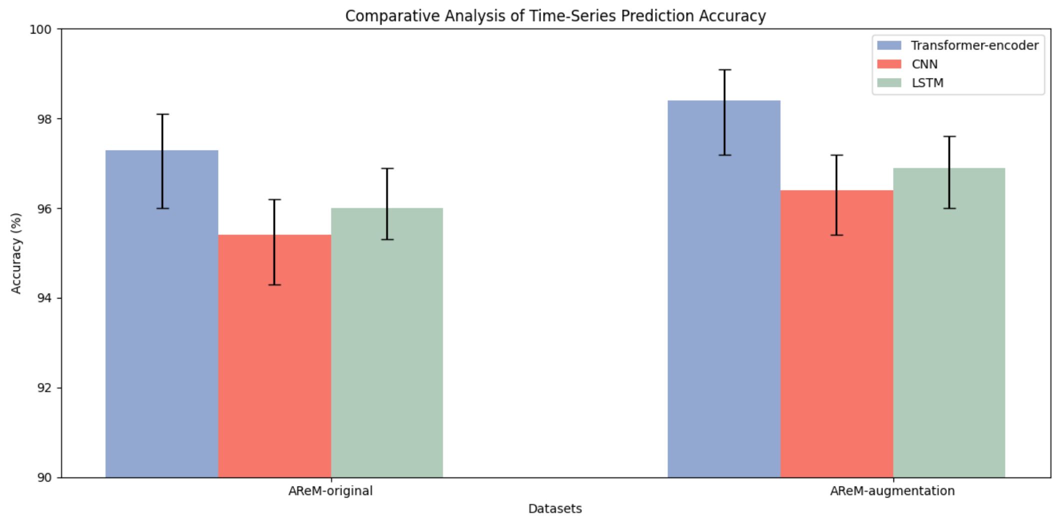

Figure 3: Bar chart visualization of the results from Table 3.

standard CNN architectures. Additionally, the Transformer's ability to focus on different parts of the sequence through its attention mechanism—utilizing a higher number of attention heads—provides a more nuanced understanding of dependencies and relationships within the data.

Despite the higher accuracy achieved by the Transformer, all three models—Transformer, LSTM, and CNN demonstrated high accuracy levels. This outcome likely resulted from the data preprocessing steps applied to the AReM dataset, which included organizing the data into structured folders for each activity class. This arrangement could potentially lead to data leakage, where information between the training and test sets is inadvertently shared, thus inflating the performance metrics. Such data leakage makes it challenging to definitively conclude which model performs best under general circumstances.

Comparatively, the LSTM model outperformed the CNN in recognizing and classifying sequences. This is because LSTMs are inherently designed to capture long-term dependencies in sequence data, making them more adept at processing time-series information where the order and context of events are critical. In contrast, CNNs, while effective in spatial feature extraction from images, do not naturally encode sequence information, which makes them less suited for detailed temporal analysis without substantial modifications or integrations with sequence-focused architectures like RNNs.

7 Conclusion

In conclusion, the exploration of human activity recognition through the AReM dataset and advanced machine learning techniques highlights significant insights and advancements in the field. Despite the high accuracy achieved across LSTM, CNN, and Transformer models, the experimental results must be interpreted with caution due to the preprocessing and potential data leakage within the dataset.

The Transformer model outperformed others due to its ability to effectively utilize positional encodings and an extensive attention mechanism, facilitating superior handling of complex dependencies within the time-series data. While all models demonstrated robust performance, the relative simplicity of LSTM in handling sequences and the CNN's capacity in feature extraction were also notable. The experimental use of data augmentation strategies like noise addition and time shifting effectively enhanced the models' generalization capabilities, addressing the initial challenges posed by the preprocessed nature of the dataset. This comprehensive approach not only advances our understanding of sensor-based activity recognition but also sets a precedent for future research to refine these technologies for real-world applications.

8 Future Work

For the future work, substantial potential exists in the exploration of hybrid models that synergize the strengths of LSTM, CNN, and Transformer architectures. These combined models have the potential to capitalize on the local feature extraction capabilities of CNNs, the sequential memory benefits inherent in LSTMs, and the global contextual insights provided by Transformers. Such an integrative approach holds the promise of achieving superior performance across a wider spectrum of activities by enhancing the model's accuracy and breadth of recognition.

In pursuit of further enhancements, more sophisticated data augmentation techniques that simulate a broader spectrum of real-world variations are considered. These techniques are anticipated to train models that are both more robust and less susceptible to overfitting. The exploration could extend to methods like synthetic minority over-sampling, refined simulation of sensor noise, or geometric transformations that are pertinent to time-series data.

The application of transfer learning, leveraging pre-trained models on analogous tasks, could also lead to substantial efficiencies. This strategy could considerably abbreviate training duration and enhance model efficacy, particularly in scenarios where labeled data are in limited supply. By utilizing the extensive knowledge encapsulated in large datasets, models can be fine-tuned to the subtle intricacies of activity recognition tasks.

Building upon the foundation of basic augmentation techniques, there is scope for developing more complex augmentations that introduce realistic variations and anomalies to sensor data. Such advanced strategies are poised to better equip models for handling unforeseen changes in data patterns, a frequent challenge in real-world settings.

Another avenue for advancement lies in the implementation of continuous learning mechanisms. By enabling models to update themselves with new information over time without losing previously acquired knowledge, thus mitigating the issue of catastrophic forgetting—these models can remain adaptable and viable over longer terms in practical deployments.

Finally, attention mechanisms, which have proven their value in Transformer models, can also be integrated into other architectures, such as CNNs or LSTMs. These mechanisms allow models to dynamically focus on various segments of input data, which could be particularly advantageous for processing extended sequences or isolating the most informative elements of the data. Integrating attention could provide a more nuanced approach to sequence analysis, opening up new possibilities for enhancing model responsiveness and accuracy in activity recognition.

References

[1] Priyajit Biswas and Tuhina Samanta. Anomaly detection using ensemble random forest in wireless sensor network. *International Journal of Information Technology*, 13(5):2043–2052, 2021.

[2] Najmul Hassan, Abu Saleh Musa Miah, and Jungpil Shin. A deep bidirectional lstm model enhanced by transfer-learning-based feature extraction for dynamic human activity recognition. *Applied Sciences*, 14(2):603, 2024.

[3] Gallicchio Claudio Pucci Rita Palumbo, Filippo and Alessio Micheli. Activity Recognition system based on Multisensor data fusion (AReM). UCI Machine Learning Repository, 2016. DOI: https://doi.org/10.24432/C5SS33.

[4] Yi Fei Tan, Xiaoning Guo, and Soon Chang Poh. Time series activity classification using gated recurrent units. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(4):3551–3558, 2021.

[5] Amir Tehrani, Meisam Yadollahzadeh-Tabari, Aidin Zehtab-Salmasi, and Rasul Enayatifar. Wearable sensor-based human activity recognition system employing bi-lstm algorithm. *The Computer Journal*, 67(3):961–975, 2024.

[6] Chao Yang, Wenxiang Jiang, and Zhongwen Guo. Time series data classification based on dual path cnn-rnn cascade network. *IEEE Access*, 7:155304–155312, 2019.

[7] Hang Yuan, Shing Chan, Andrew P Creagh, Catherine Tong, Aidan Acquah, David A Clifton, and Aiden Doherty. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *NPJ Digital Medicine*, 7(1):91, 2024.