Comparative Analysis of Hybrid Deep Learning Model and N-BEATS model for Human Activity Recognition

Recent Advances in Machine Learning

Universität Siegen

Srinadh Alugu 1717015 **Sanketh Ganesh** 1712320 **Tarun Phanindra** 1593084

Abstract

We proposed two deep learning models, such as a hybrid model and the Neural Basis Expansion Analysis for Interpretable Time Series(N-BEATS) model, for Human Activity Recognition (HAR) using the WEAR dataset. The Hybrid Deep Learning Model leverages various neural network architectures, including CNN, Bi-LSTM, GRU, and transformer-encoded layers. In contrast, the N-BEATS Model, which was initially created for time-series forecasting, has been adapted for human activity recognition. Its architecture is general and flexible (1). The Hybrid Deep Learning model has shown an accuracy of 85.64%, average precision score of 85.87%, average F1-score of 85.58%, and average Recall score of 85.65%. The N-BEATS Model has shown an accuracy of 79.57%, average precision score of 79.76%, average F1-score of 79.58%, and average Recall score of 79.57%. Our results demonstrate that the Hybrid Deep Learning Model surpasses the N-BEATS Model in terms of accuracy, precision, recall, and F1 score. Furthermore, it demonstrates the N-BEATS Model's applicability to new areas such as HAR, establishing a benchmark for future time-series classification research and applications.

1 Introduction

Human Activity Recognition is a research field focusing on creating and testing novel methods for reliably detecting human activities (2). HAR uses inertial body sensor data which is useful for various applications, including health monitoring and activity tracking(3). Wearables in fitness and lifestyle can greatly develop one of the most prevalent aspects of HAR applications(3). These sensors expand the use of activity recognition outside instrumented rooms enabling smart support(4). We use an outdoor sports dataset called WEAR, the dataset includes acceleration and egocentric video data from 18 participants engaged in 18 different workout activities by placing inertial sensors on wrists and ankles along with a head-mounted camera(5). Combining wearable sensors and deep learning for feature learning improves variation, and generality, and addresses challenges in human activity recognition(6). In this study, we aimed to work on how two models with different neural network architectures perform on the given time series classification task for HAR. One of the models is the Hybrid Deep Learning Model which consists of CNN, Bi-LSTM, GRU, and Transformer encoded layer, the other is the N-Beats Model. Upon doing the training and testing on the WEAR dataset by using the K-fold cross-validation technique, we obtained their respective evaluation metrics. We also compared both models to understand their strengths. As expected, the Hybrid Model has shown better performance compared to the N-Beats model. However, the N-Beats model has given satisfactory results being a forecasting model.

2 Methodology

The process involved several key steps such as preprocessing, model building, hyperparameter tuning, cross-validation, and evaluation. Below is an explanation of each architecture and the methodology used.

2.1 Hybrid Deep Learning Models

Hybrid deep learning models combine it with several neural network designs to maximize their respective strengths, which leads to more accurate and robust predictions. We explored the implementation of CNN, Bi-LSTM, GRU, and Transformer Encoder as hybrid models for activity recognition tasks.

'Convolutional Neural Networks (CNN)', are made to handle data structures that resemble grids, such as time series or image data. It creates feature maps that draw attention to significant patterns like edges and textures by using convolutional layers to apply filters to the input data. These patterns are later transmitted through pooling layers to lower the dimensionality of the data while keeping the most important features, and activation functions (like ReLU) to add non-linearity.

'Bidirectional Long Short-Term Memory (Bi-LSTM)' networks are a form of Recurrent Neural Network (RNN) that detects connections in input sequences from both previous and future contexts. LSTM cells are meant to retain information for long periods, which overcomes the vanishing gradient problem found in standard RNNs. resulting in a more complete knowledge of the sequential input. This bidirectional processing is especially valuable for time-series data, as context from both ways can increase prediction accuracy.

Another kind of RNN that resembles an LSTM but has a simpler architecture is the 'Gated Recurrent Unit (GRU)'. By controlling the input flow through gating mechanisms, especially update and reset gates, GRUs can solve the vanishing gradient problem more effectively during training than LSTMs. GRUs may be trained and inferred more quickly, which makes them useful in applications where processing speed is essential.

The 'Transformer Encoder' is a component of the Transformer architecture, which has transformed sequence modeling jobs, particularly natural language processing. Transformer Encoders, unlike RNNs, analyze sequences in parallel employing self-attention processes that allow the model to balance the importance of each piece in the sequence. Positional encoding is employed to incorporate the sequence's order into the model, as the self-attention mechanism is order-agnostic. Multi-head attention improves the model's capacity to focus on multiple sections of the sequence at once.

The outline of the model can be seen below in figure 1

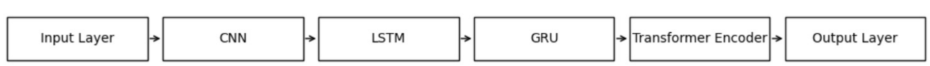

Figure 1: Outline of hybrid deep learning model

2.2 N-BEATS Model

A deep learning architecture created especially for time-series forecasting is the N-BEATS model. N-BEATS breaks down the time series into trend and seasonal components using a stack of completely connected layers arranged into basis expansion blocks. The outline of the architecture can be seen in figure2. By enhancing gradient flow, these blocks make the model more stable and easier to train. They do this by using both forward and backward residual linkages(1). Smooth information flow across the network is guaranteed by the residual links, which is especially helpful for very deep architectures.

A primary benefit of N-BEATS is its generic architecture, which enables extensive customization-free application to a broad range of time-series forecasting applications. Because of its completely connected layers and base expansion technique, the model can identify complex patterns in the data and produce forecasts that are accurate and clear. Because of this, N-BEATS is an effective tool for applications like financial forecasting, demand planning, and activity detection that call for accurate time-series forecasts.

In this study, we rigorously trained the N-BEATS model using the WEAR datasets, assessing its predictive power in the context of activity identification by contrasting its performance with hybrid deep learning models. Each model's efficacy was thoroughly assessed, which included preprocessing, model building, hyperparameter tuning, cross-validation, and evaluation.

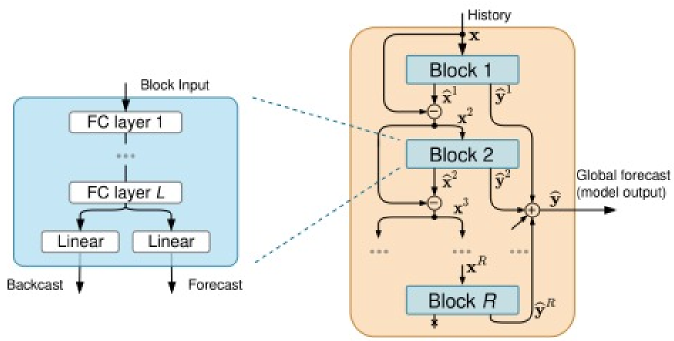

Figure 2: Outline of N-BEATS model

3 Experiment

- Preprocessing: To make the raw WEAR dataset consistent and model-compatible, it was cleaned, normalized, and transformed. In this step, the data was segmented into fixed-length windows, scaling features, and addressing missing values. To improve the quality of the dataset, other methods including noise filtering and data augmentation were also used.

- Model Building: Keras or sklearn, two suitable machine learning libraries, were used to implement each architecture. Based on preliminary experiments and research of the literature, the network topologies were meticulously created, taking into account factors like the number of layers, units per layer, and activation functions.

- Hyperparameter Tuning: Using grid search or random search techniques, hyperparameters including learning rates, batch sizes, number of epochs, and dropout rates were tuned. To determine the ideal parameters, this stage entailed training many versions of each model with various hyperparameter configurations. Used keras_tuner in a random search. Conducted a maximum of 10 trials, each with a single execution.

- Cross-Validation: Stratified K-Fold, used 5-fold stratified cross-validation to ensure balanced class representation. Early stopping, implemented early stopping based on validation loss to avoid overfitting.

- Evaluation: The models were assessed using performance criteria such as confusion matrices, accuracy, precision, recall, and F1-score. The ability of each model to accurately identify activities from the WEAR dataset was demonstrated by these indicators. Furthermore, the model's performance was examined at various folds to evaluate its reliability and consistency.

By using this approach, we aimed to present a thorough analysis of various neural network topologies and their performance in activity recognition tasks. We were able to assess and determine the most suitable model for the given dataset and application.

Table 1: Hybrid model fold accuracies

Overall Test Accuracy	0.8565
Fold-wise Test Accuracies:	
Fold 1	0.8556
Fold 2	0.8570
Fold 3	0.8569
Fold 4	0.8610
Fold 5	0.8518

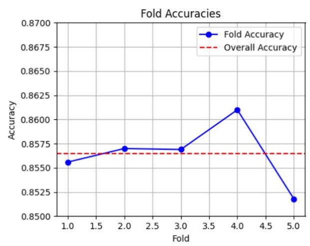

Figure 3: Hybrid line metrics

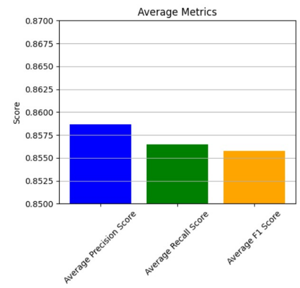

Figure 4: Hybrid bar metrics

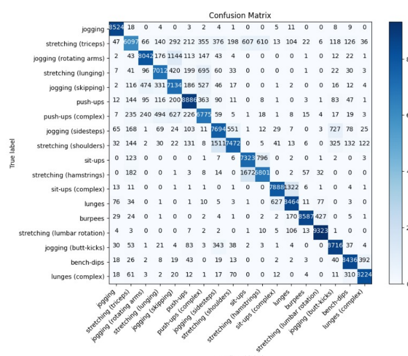

Figure 5: Confusion matrix for hybrid deep learning model

Our observations for the Hybrid model as per the confusion matrix are as follows:

- The hybrid model demonstrates higher performance, with most activities exhibiting high true positive rates.
- There is some ambiguity between similar activities, showing that there is still a place for progress in the recognition of subtle distinctions in complicated activity patterns.

Table 2: N-BEATS model fold accuracies

Description	**Accuracy**
Overall Test Accuracy	0.7957
Fold-wise Test Accuracies:	
Fold 1	0.7955
Fold 2	0.7956
Fold 3	0.7961
Fold 4	0.7957
Fold 5	0.7958

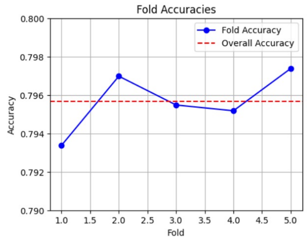

Figure 6: N-BEATS line metrics

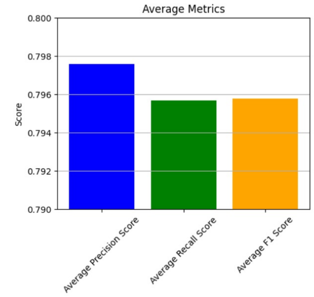

Figure 7: N-BEATS bar metrics

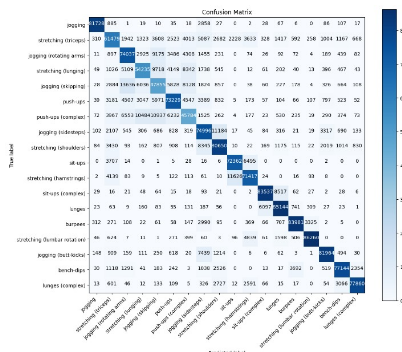

Figure 8: Confusion matrix for N-BEATS model

Our observations for the N-BEATS model as per the confusion matrix are as follows:

- The N-BEATS model performance shows more confusion between the activities indicating that it is struggling with variations in complex activities.
- This model sets a strong benchmark for time-series classification but requires additional modification for the task of Human Activity Recognition.

4 Conclusion

Our observations show that the Hybrid Model, comprising of CNN, Bi-LSTM, GRU, and Transformer encoded layers, outperformed the other model. This is mainly because of its ability to capture local features (by CNN), long-term dependencies (via Bi-LSTM and GRU), and the Transformer's attention mechanism, which focuses on significant areas of the input data. Such a combination enables a more robust and accurate time series data classification. However, the N-Beats model provides a feasible alternative due to its agility and effectiveness, making it an asset for specialized applications that function in real time.

Future work can delve into domain adaptation techniques to facilitate the transfer of knowledge from one domain to another by pre-training the N-BEATS model across new, multiple related datasets, and evaluation through various metrics will help to determine its limitations and strengths and enhance its generalization across various applications. Our findings contribute to the advancement of machine learning and lay the groundwork for future studies in this area.

References

[1] B. N. Oreshkin, D. Carpov, N. Chapados, and Y. Bengio, "N-beats: Neural basis expansion analysis for interpretable time series forecasting," *arXiv preprint arXiv:1905.10437*, 2019.

[2] A. O. Ige and M. H. Mohd Noor, "A survey on unsupervised learning for wearable sensor-based activity recognition," *Applied Soft Computing*, vol. 127, p. 109363, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494622005191

[3] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa, "Deep learning in human activity recognition with wearable sensors: A review on advances," *Sensors*, vol. 22, no. 4, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/4/1476

[4] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, jan 2014. [Online]. Available: https://doi.org/10.1145/2499621

[5] M. Bock, H. Kuehne, K. Van Laerhoven, and M. Moeller, "Wear: An outdoor sports dataset for wearable and egocentric activity recognition," *arXiv preprint arXiv:2304.05088*, 2023. [Online]. Available: https://arxiv.org/abs/2304.05088

[6] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Systems with Applications*, vol. 105, pp. 233–261, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417418302136