Proceedings of the 11th International Conference of Control Systems, and Robotics (CDSR 2024)
Toronto, Canada – 10 -12 June 2024
Paper No. 115
DOI: 10.11159/cdsr24.115

Advanced Exercise Classification with a Hybrid CNN-GRU Model: Utilising IMU Data from Cell Phones

Jing Zhang, Meng Cheng Lau, Ziping Zhu
Bharti School of Engineering and Computer Science
Laurentian University
935 Ramsey Lake Rd, Sudbury, ON, Canada
jzhang17@laurentian.ca; mclau@laurentian.ca; zzhu4@laurentian.ca

***Abstract* -** We introduce a hybrid CNN-GRU model in this study to classify exercises using IMU time-series data, with a focus on jumping jacks, lunges, and squats. By combining Convolutional Neural Networks with Gated Recurrent Units, our model effectively manages the high dimensionality and variable sampling rates of IMU data. We employed data normalisation and augmentation techniques to refine the dataset. Our model showed high accuracy in classifying types of exercises, highlighting its potential in motion classification and fitness-tracking applications. These results emphasise the value of hybrid deep learning methods in analysing complex time-series data and make a significant contribution to the understanding of human exercise movement patterns.

***Keywords*: Motion Classification, IMU, Deep Learning, CNN-GRU.**

1. Introduction

Advancements in exercise physiology and sports psychology have driven the evolution of exercise classification, a multidisciplinary field crucial to health and fitness, sports science, and public health policy. This classification plays a vital role in crafting targeted exercise plans and gauging the impact of physical activities. With the integration of machine learning and data analysis, classification has grown more sophisticated, offering more personalised fitness advice and improving injury prevention in sports medicine. Recent technological strides have pivoted the focus to time-series data from Inertial Measurement Unit (IMU) sensors, which are widespread in smartphones and wearable devices. These sensors, which we prefer to video data, provide precise measurements of physical movements and are less intrusive, making them perfect for capturing the subtleties of exercises while preserving privacy. Traditional classification models like Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbours (KNN), and Hidden Markov Models (HMM) struggle with the high-dimensional nature of time-series data. Each model comes with strengths—for instance, SVM's robustness in high dimensions and RF's proficiency in unravelling complex data relationships. Yet, they also grapple with issues like computational inefficiency and challenges in capturing time-series data's temporal dynamics. As a result, exercise classification is increasingly harnessing advanced IMU sensor data, driving the need for sophisticated models capable of efficiently and accurately navigating its complexity and dimensionality for motion recognition and classification.

Our study sets out to improve motion classification using IMU time-series data by developing a hybrid CNN-GRU model. This model marries Convolutional Neural Networks (CNNs) for spatial feature extraction with Gated Recurrent Units (GRUs) to capture temporal dynamics, tackling the high-dimensional challenge of IMU data. We focus on the model's ability to process IMU data with varying sampling rates and to boost classification efficiency and accuracy. Through a comparative analysis with current models, we will confirm the superiority of this hybrid approach. We anticipate that our model, which integrates CNNs with GRUs, will make a significant mark in managing complex IMU time-series data for exercise motion classification.

2. Background

Significant evolution marks the field of exercise classification, which now blends traditional methods with cutting-edge technologies. Lu et al. [1] have pioneered the use of IMU data for fine-grained activity recognition, thus paving new paths in this domain. Wang et al. [2] advanced the field by using HMMs for arm gesture classification with IMUs in medical rehabilitation, highlighting the method's diverse applications. Recently, researchers have been investigating the capabilities of CNNs and GRUs under various circumstances. Ahmed et al. [3] employed a combination of CNN, LSTM, and GRU models, including hybrid CNN-GRU architectures, for speech emotion recognition. Chiu et al. [4] put into action a CNN-

GRU hybrid network to forecast building energy consumption, underlining the combined strengths of CNNs and GRUs in capturing spatiotemporal features. In a similar vein, Khan et al. [5] utilised a CNN-LSTM model for motion classification with depth camera sensors, proving the adaptability of these models with different types of sensor data.

IMU data has played a pivotal role in classification tasks. Eyobu et al. [6] have brought forward innovative data augmentation methods such as window slicing and jittering for IMU data in Human Activity Recognition (HAR). Wang et al. [7] evaluated individual models like CNN, LSTM, and GRU for animal behaviour classification, demonstrating their efficacy. Concurrently, Ferrari et al. [8] employed CNN-based ResNet models for human activity recognition using accelerometer data, thus expanding the horizons of IMU data usage. Although current research is expansive, it reveals discernible gaps. Theissler et al. [9] underscore the demand for explainable AI in time series classification, a niche that current models have scarcely filled. Small et al. [10] have probed the impact of reduced accelerometer sampling rates on activity monitoring, hinting at more efficient data collection methods. The resampling methods that Wang et al. [11] suggested for sensor data augmentation could address challenges associated with scarce labelled data. Broad overviews provided by Lima et al. [12] and Kim et al. [13] suggest a strong need for further empirical research that utilises detailed datasets and testing to confirm these methods' practicality. These identified gaps offer chances for future research endeavours to develop more efficient, explainable, and empirically proven methods in exercise classification.

3. Methodology

3.1. CNN and GRU

We deploy this research on a CNN-GRU hybrid deep learning model that combines CNN and GRU architectures to focus on motion analysis from IMU time series data. This hybrid model uses CNN for spatial feature extraction and GRU for temporal sequence processing. The CNN component structures itself to effectively process the time-series input data, with convolutional layers that employ a set of learnable filters. These filters capture spatial dependencies through convolution operations between the filters and the input, creating feature maps. Batch normalisation follows the convolutional layers to stabilise the learning process and enhance the model's efficiency. We can optionally insert a pooling layer to reduce the output's spatial dimensions, which helps lower computational demands and prevent overfitting.

The GRU component, succeeding the CNN, processes the temporal features we extracted earlier. GRUs are adept at handling data sequences, with each unit featuring an update and a reset gate that govern the flow of information, crucial for capturing the temporal dynamics and dependencies in IMU data. We integrate by extracting spatial features with the CNN, then normalising these features and transposing the matrix for GRU compatibility. The GRU processes these normalised features, culminating in a final classification output from a fully connected layer that merges spatial and temporal insights, as illustrated in Fig 1.

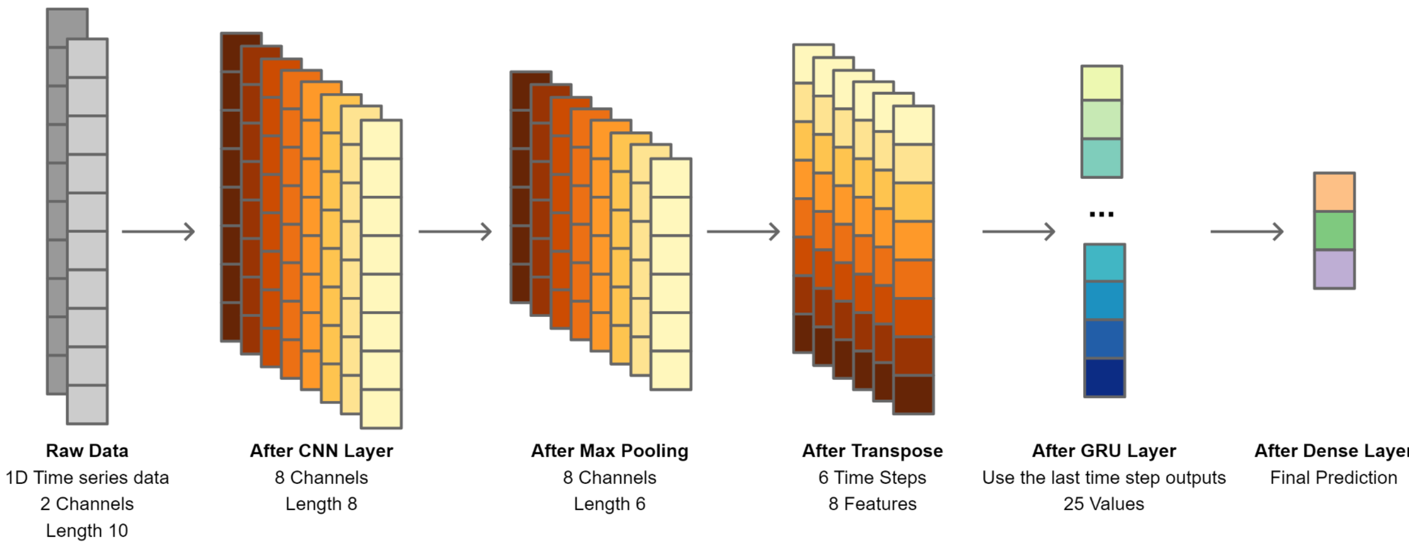


Fig 1. CNN-GRU Hybrid Model

3.1. Data Collection

We gathered data for this study using an Android application named "Sensor Logger"[14] that is selected for its precise capture of a wide spectrum of motion data. The participant executed a series of exercises – Jumping Jack, Lunge, and Squat – repeating each exercise in ten distinct sets to ensure a diverse and comprehensive dataset that captures a range of bodily motions and exercise dynamics. The app collected IMU data, which included accelerometer readings along the x, y, and z axes to track linear acceleration movements, and orientation data through quaternions (qx, qy, qz, qw) and Euler angles (roll, pitch, yaw). We recorded total acceleration data on the x, y, and z axes, providing a full view of the dynamic and static forces on the body during exercises. However, the sampling rates of these sensor readings varied due to the performance constraints of mobile devices and limitations in the Android operating system. For example, the accelerometer's sampling rate during the Jumping Jack exercise fluctuated between 52.7128 Hz and 52.7148 Hz, and the orientation sensor's rate varied between 59.5660 Hz and 60.4022 Hz.

To standardise the variable sampling rates, we applied a uniform down-sampling procedure to all the data, setting it to a consistent rate of 50 Hz. We used interpolation methods to align all data types to this fixed frequency, ensuring dataset consistency and comparability. After standardisation, we segmented the continuous raw sensor data into discrete sets corresponding to the specific exercises performed, resulting in 30 labelled datasets, with ten for each exercise type. We based the segmentation on pattern detection through human observation and expertise, and we implemented this in Python to automatically distinguish and separate different exercise movements. To augment the original dataset's limited size of 30 exercise groups and enhance its size and diversity, we applied data augmentation techniques. We introduced realistic variability by adding random multiplicative noise to 30% of the sampling points, each multiplied by a random coefficient ranging from 0.8 to 1.1. Moreover, we generated new subsequences from each exercise sequence by randomly selecting start and end points within the original sequences, thereby expanding the dataset and introducing randomness crucial for a diverse and robust training set.

Another key aspect of data preprocessing was addressing the variability in the time durations taken by the participant to complete each exercise instance, resulting in inconsistent numbers of samples per set. We used interpolation to standardise each dataset to a fixed number of samples, as demonstrated in Fig 2. This method did not compromise the model's accuracy but significantly reduced the computational load during training. Additionally, we employed data visualisation techniques to confirm that the critical motion characteristics and patterns remained intact, ensuring the dataset's integrity and applicability for training the CNN-GRU model.

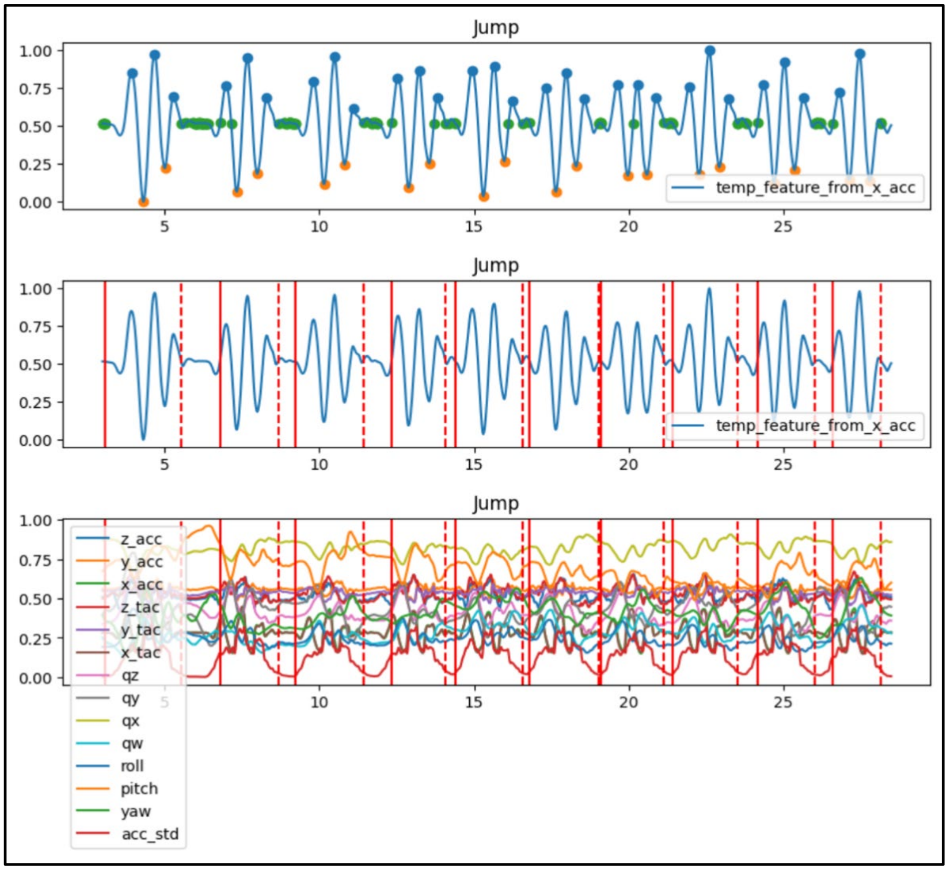


Fig 2. Sampling for 30 jumping jack

3.3. Model Training

We structured the training of the hybrid CNN-GRU model and divided the IMU dataset into an 80% training set and a 20% test set to ensure comprehensive training and substantial unbiased evaluation. We fine-tuned significant architectural parameters, such as convolutional kernel sizes and the choice between max and average pooling layers, to optimise the model, deliberately excluding padding from these layers to concentrate on core motion patterns. We selected training parameters like learning rate, batch size, and the number of epochs with an eye on balancing efficiency and the use of computational resources. We chose all hyperparameters through a grid search cross-validation strategy to boost machine learning performance metrics like accuracy and efficiency. Since we set hyperparameters before learning, the system could not correlate them with the data, necessitating an exhaustive exploration of their combinations to find the optimal set. Our grid search cross-validation strategy varied parameters within predefined ranges to rigorously test the model's sensitivity to different hyperparameter combinations. This systematic exploration of the parameter space helped us find a balance between maximising prediction accuracy and the efficient use of computing resources. We used accuracy as the primary performance metric during training, providing direct and clear feedback on the model's effectiveness, especially during regular assessments with the 20% test set. This focused evaluation strategy allowed for timely and precise model adjustments, securing its accurate exercise classification capability and demonstrating the utility and efficacy of our methodology in deep learning applications for motion classification.

4. Implementation

We implemented the CNN-GRU model using Python and the PyTorch framework, renowned for its neural network modelling efficiency. Google Colab served as our primary platform, providing access to high-performance GPUs. This setup offered us a powerful combination of flexibility and computational strength, perfectly suited for deep learning tasks without the need for sophisticated local hardware. At the project's outset, we established the Python environment in Google Colab, installing PyTorch and preparing dependencies. We formatted and normalised the pre-processed IMU data, preparing it for the training phase. Then, we constructed the CNN-GRU model, coding the convolutional and GRU layers and integrating essential components for effective learning. Key training parameters like learning rate, batch size, and epoch count were set, along with tools for monitoring the model's training progress. The model was trained with the training set, constantly monitored, and fine-tuned based on performance metrics. After training, we tested the model on the test set to assess its accuracy and generalisation capability. We addressed challenges like variable IMU data sampling rates and durations by applying interpolation for standardisation. To enrich the dataset, we used data augmentation methods, including noise addition and subsequence extraction. We managed the computational demands by utilising Google Colab’s GPU resources, which ensured an efficient training process. These steps were critical in developing a high-performing and robust CNN-GRU model for motion classification, overcoming obstacles to enhance performance, and proving the model's practicality in deep learning for motion analysis.

5. Results and Discussion

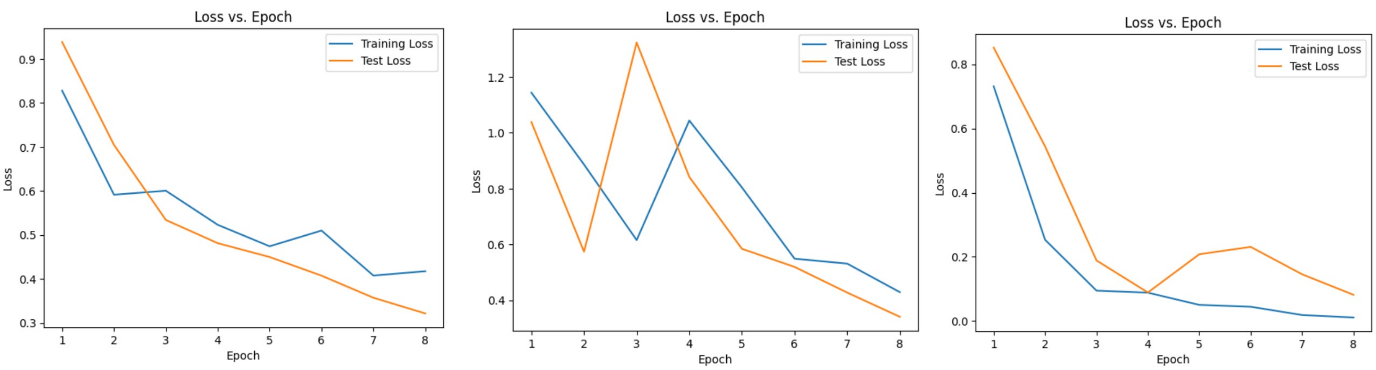
5.1. Results

Table 1 that shows the evaluation of various models on both the training and test sets revealed significant findings. The Random Forest model showed an accuracy of 0.81 on the training set and 0.85 on the test set, while the Support Vector Machine (SVM) model exhibited higher accuracy, with 0.91 on the training set and 0.9 on the test set. The CNN model improved accuracy, achieving 0.94 on the training and 0.91 on the test sets. The GRU model matched CNN's test set accuracy at 0.91 but was slightly lower on the training set with 0.9. The CNN-GRU hybrid model outperformed all individual models, achieving perfect accuracy on the training set and 0.97 on the test set. Compared to traditional methods like Random Forest and SVM, the deep learning models, specifically the hybrid CNN-GRU, demonstrated superior performance. While traditional models provided a solid baseline, the deep learning approaches, particularly the hybrid model, showcased significant improvements in precision, recall, and F1 scores across training and test datasets.

Fig. 3 shows the training and test loss as a function of epochs for CNN, GRU, and CNN-GRU models. The CNN graph shows both training and test losses decrease steadily, indicating good generalisation and stable learning. The GRU graph shows high volatility with significant spikes in test loss, suggesting possible overfitting or other issues like improper learning rate settings. The CNN-GRU hybrid graph exhibits a smooth decline in training loss and a more variable test loss that flattens and then briefly increases, hinting at an initial overfitting that the model later overcomes. This comparison suggests that the CNN is learning most consistently, the GRU may have difficulty capturing the temporal patterns or could be sensitive to hyperparameters, and the CNN-GRU combines elements of both, with a potential for overfitting that needs to be monitored. The results underscore the effectiveness of deep learning models in handling complex classification tasks. The CNN-GRU hybrid model's perfect training set scores and near-perfect test set scores highlight its capability to learn and generalise from the IMU data efficiently. This signifies a substantial advancement in exercise classification accuracy, demonstrating the potential of combining convolutional and recurrent neural networks.

Table 1: Execution results of all learning models(1.00=100%)

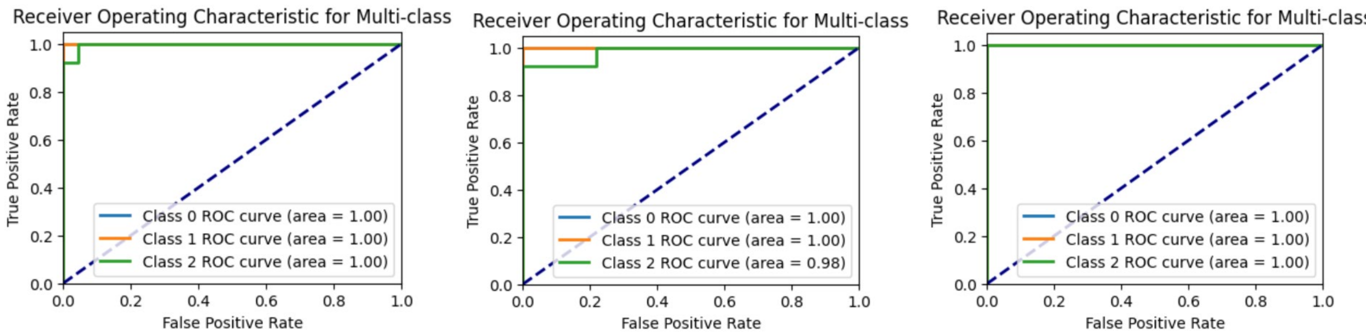
	Data Type	Precision	Recall	F1-Score
Random Forest	Train Set	0.81	0.72	0.65
	Test Set	0.85	0.81	0.78
Support Vector Machine	Train Set	0.91	0.87	0.87
	Test Set	0.90	0.86	0.85
CNN	Train Set	0.94	0.94	0.94
	Test Set	0.91	0.89	0.88
GRU	Train Set	0.90	0.89	0.88
	Test Set	0.91	0.89	0.88
CNN-GRU	Train Set	1.00	1.00	1.00
	Test Set	0.97	0.97	0.97



CNN **GRU** **CNN-GRU**

Fig 3. Comparison of the training and test loss for CNN, GRU, and CNN-GRU models.

Fig.4 provides an overview of ROC curves for the three models that exhibit exceptional classification performance on the test data for this multi-class problem where class 0, class 1, and class 2 represent jumping jack, lunge, and squat movements, respectively. All three models achieve perfect or near-perfect Area Under the Curve (AUC) scores of 1.00 across most classes, indicating an excellent true positive rate without increasing the false positive rate. The slight deviation seen in the GRU model for class 2, with an AUC of 0.98, suggests a marginally lower but still outstanding ability to classify squats compared to the other movements. Overall, the near-identical AUC scores across all classes for each model suggest that all models are highly effective for this specific test dataset, although such perfect classification is uncommon in practice and could warrant further investigation to ensure the models' robustness.



CNN **GRU** **CNN-GRU**

Fig 4. ROC curves for CNN, GRU, and CNN-GRU models' classification performance on test data: class 0, class 1, and class 2 represent jumping jack, lunge, and squat movements respectively.

5.2. Discussion

The high accuracy and F1 scores achieved by the CNN and GRU models indicate their strength in extracting spatial and temporal features, respectively. However, the hybrid CNN-GRU model's superior performance suggests that integrating these models' strengths can capture the nuances of IMU data more effectively, leading to better classification outcomes. The findings align with existing literature that posits the superiority of hybrid deep learning models in various classification tasks. Studies like Lu et al. [1] and Ahmed et al. [15] have previously demonstrated the effectiveness of combining different neural network architectures, and our results further corroborate these observations in the context of exercise classification. The CNN-GRU model's high accuracy and generalisation capability indicate its potential for real-world applications, such as fitness tracking and rehabilitation monitoring. However, the model's complexity and the need for substantial computational resources might limit its deployment in resource-constrained environments. Future research could focus on optimising the model's computational efficiency and exploring its applicability in broader real-world scenarios.

6. Conclusion

This study evaluated the performance of various models, including traditional machine learning and advanced deep learning techniques, for classifying exercises using IMU data. The findings unambiguously demonstrate the superior performance of deep learning models, with the CNN-GRU hybrid model standing out by achieving nearly perfect accuracy. Specifically, the CNN-GRU model outperformed traditional models such as Random Forest and Support Vector Machine and individual deep learning models like CNN and GRU in terms of precision, recall, and F1-score across both training and test sets. The results of this study have significant implications for the development of advanced exercise classification systems. By demonstrating the effectiveness of the CNN-GRU hybrid model, this research highlights the potential of combining convolutional and recurrent neural networks to handle the spatial-temporal complexity inherent in IMU data. This insight is precious for health and fitness technology applications, where accurate exercise classification can contribute to personalised fitness tracking and rehabilitation programs. While the CNN-GRU model shows promising results, this study opens several avenues for future research. One potential area is exploring the model's applicability in real-time exercise classification systems, considering the computational demands of deep learning models. Additionally, future work could investigate the model's performance across a broader range of physical activities and in more diverse datasets, including those with varying levels of complexity and granularity. Another promising direction is enhancing the model's interpretability and explainability, which is crucial for applications in clinical settings where understanding model decisions can be as important as the decisions themselves. Lastly, research could also focus on optimising the model for deployment on edge devices, enabling more widespread and accessible fitness and health monitoring solutions. In conclusion, this study contributes to the growing body of knowledge on exercise classification, affirming the value of hybrid deep learning models in achieving high accuracy in complex classification tasks. With its robust performance, the CNN-GRU model sets a new benchmark in the field and serves as a foundation for future innovations in exercise recognition and related areas.

References

[1] Y. Lu and S. Velipasalar, “Autonomous Human Activity Classification From Wearable Multi-Modal Sensors,” *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11403–11412, Dec. 2019, doi: https://doi.org/10.1109/jsen.2019.2934678.

[2] D. Wang, X. Meng, J. Wang, and Y. Liu, “HMM-based IMU data processing for arm gesture classification and motion tracking,” *International Journal of Modelling, Identification and Control*, vol. 42, no. 1, p. 54, 2023, doi: https://doi.org/10.1504/ijmic.2023.10053831.

[3] Md. Rayhan Ahmed, S. Islam, A. K. M. Muzahidul Islam, and S. Shatabda, “An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition,” *Expert Systems with Applications*, vol. 218, p. 119633, May 2023, doi: https://doi.org/10.1016/j.eswa.2023.119633.

[4] M. Chiu, H.-W. Hsu, K.-S. Chen, and C.-Y. Wen, “A hybrid CNN-GRU based probabilistic model for load forecasting from individual household to commercial building,” *Energy Reports*, vol. 9, pp. 94–105, Oct. 2023, doi: https://doi.org/10.1016/j.egyr.2023.05.090.

[5] I. U. Khan, S. Afzal, and J. W. Lee, “Human Activity Recognition via Hybrid Deep Learning Based Model,” *Sensors*, vol. 22, no. 1, p. 323, Jan. 2022, doi: https://doi.org/10.3390/s22010323.

[6] O. Steven Eyobu and D. Han, “Feature Representation and Data Augmentation for Human Activity Classification Based on Wearable IMU Sensor Data Using a Deep LSTM Neural Network,” *Sensors*, vol. 18, no. 9, p. 2892, Aug. 2018, doi: https://doi.org/10.3390/s18092892.

[7] L. Wang, R. Arablouei, F. A. P. Alvarenga, and G. J. Bishop-Hurley, “Classifying animal behavior from accelerometry data via recurrent neural networks,” *Computers and Electronics in Agriculture*, vol. 206, p. 107647, Mar. 2023, doi: https://doi.org/10.1016/j.compag.2023.107647.

[8] A. Ferrari, D. Micucci, M. Mobilio, and P. Napoletano, “On the Personalization of Classification Models for Human Activity Recognition,” *IEEE Access*, vol. 8, pp. 32066–32079, 2020, doi: https://doi.org/10.1109/access.2020.2973425.

[9] A. Theissler, F. Spinnato, U. Schlegel, and R. Guidotti, “Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions,” *IEEE Access*, vol. 10, pp. 100700–100724, Jan. 2022, doi: https://doi.org/10.1109/access.2022.3207765.

[10] S. Small, S. Khalid, P. Dhiman, S. Chan, D. Jackson, A. Doherty, and A. Price, “Impact of Reduced Sampling Rate on Accelerometer-Based Physical Activity Monitoring and Machine Learning Activity Classification,” *Journal for the Measurement of Physical Behaviour*, vol. 4, no. 4, pp. 298–310, Dec. 2021, doi: https://doi.org/10.1123/jmpb.2020-0061.

[11] J. Wang, T. Zhu, J. Gan, L. L. Chen, H. Ning, and Y. Wan, “Sensor Data Augmentation by Resampling in Contrastive Learning for Human Activity Recognition,” *IEEE Sensors Journal*, vol. 22, no. 23, pp. 22994–23008, Dec. 2022, doi: https://doi.org/10.1109/jsen.2022.3214198.

[12] W. Sousa Lima, E. Souto, K. El-Khatib, R. Jalali, and J. Gama, “Human Activity Recognition Using Inertial Sensors in a Smartphone: An Overview,” *Sensors*, vol. 19, no. 14, p. 3213, Jul. 2019, doi: https://doi.org/10.3390/s19143213.

[13] H. Kim and I. Kim, “Human Activity Recognition as Time-Series Analysis,” *Mathematical Problems in Engineering*, vol. 2015, pp. 1–9, 2015, doi: https://doi.org/10.1155/2015/676090.

[14] K. Choi, “One-tap Sensor Logger” https://www.tszheichoi.com/sensorlogger. (accessed Jan. 3, 2024).

[15] N. Ahmad, R. A. R. Ghazilla, N. M. Khairi, and V. Kasi, “Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications,” *International Journal of Signal Processing Systems*, vol. 1, no. 2, pp. 256–262, 2013, doi: https://doi.org/10.12720/ijsps.1.2.256-262.