



Passive and Continuous Assessment of Motor Function Recovery in Post-Stroke Rehabilitation using Wearable Sensors and Deep Learning

1st Nkemdirim Chinedu David *
Liaoning Petrochemical University, Zhejiang University
Hangzhou, China
Uniquedavido@gmail.com

2nd Diallo Abdoul Salam
Zhejiang University
LAGE, Guinea
dialloa083@yahoo.com

Received on July 16th, revised on August 19th, accepted on August 26th, published on October 10th

Abstract—Continuous assessment of motor recovery remains challenging in post-stroke rehabilitation because conventional clinical evaluations are intermittent and labor-intensive. This study proposes a passive longitudinal monitoring framework that integrates wearable inertial measurement units (IMUs) with a CNN-LSTM model to predict motor function at the subsequent clinical assessment. Eighty-five stroke patients were enrolled, and complete longitudinal data from 77 participants were analyzed. Seven IMUs were used during 7-day home-monitoring periods, and temporal kinematic features were extracted to predict the subsequent Fugl-Meyer Assessment (FMA) score. A sensor-combination optimization strategy was further conducted to identify a practical wearable configuration, while SHAP analysis was employed to improve model interpretability. Experimental results demonstrated that the proposed model achieved an R^2 of 0.92, MAE of 2.85, and RMSE of 3.41 FMA points on the test set. Among 127 evaluated sensor combinations, a lightweight three-sensor configuration (sternum, affected wrist, and affected thigh) achieved an RMSE of 3.58, only 0.17 points higher than the full seven-sensor system while substantially reducing hardware complexity. These results demonstrate that the proposed framework enables accurate and interpretable prediction of motor recovery using passive home-monitoring data, providing a practical solution for long-term rehabilitation assessment and wearable healthcare applications.

Keywords—Wearable Sensors, Motor Function Recovery, Deep Learning, Stroke Rehabilitation, Inertial Measurement Units, Predictive Modeling, CNN-LSTM, Digital Biomarkers

1. INTRODUCTION

Stroke remains a leading cause of long-term disability worldwide, with motor impairment being one of the most common and debilitating consequences [1]. The recovery of motor function is a critical determinant of a patient's ability to perform activities of daily living (ADL) and their overall quality of life. Traditional clinical assessments for motor function, such as the Fugl-Meyer Assessment (FMA) and the Wolf Motor Function Test (WMFT), are the current gold

standard. However, these assessments are typically conducted in clinical settings at discrete time points, providing only a snapshot of a patient's functional status. This approach is often labor-intensive, subjective, and fails to capture the dynamic nature of motor recovery and the subtle changes that occur in a patient's natural environment [2, 3].

Recent advancements in wearable sensor technology, particularly inertial measurement units (IMUs), offer a transformative potential for continuous and objective monitoring of motor function outside the clinic [4]. These low-cost, unobtrusive sensors can capture a rich stream of data on movement patterns, including gait, balance, and limb usage, during a patient's daily activities. This passive data collection provides a more ecologically valid and comprehensive picture of functional ability compared to in-clinic tests [5, 6]. The integration of this high-resolution data with advanced machine learning and deep learning techniques opens new avenues for developing predictive models of motor recovery, personalizing rehabilitation programs, and enabling timely interventions [7].

While the use of wearable sensors in rehabilitation is growing, several challenges remain. A significant body of research has focused on using sensors to quantify motor performance, but there is a lack of studies that leverage this data to predict long-term functional outcomes. Furthermore, many studies have relied on a large number of sensors, which can be cumbersome for patients and impractical for long-term home monitoring. There is a critical need to identify the optimal number and placement of sensors that can provide robust and accurate predictions while ensuring user comfort and adherence [8]. Additionally, most existing models are "black boxes," lacking the interpretability needed to provide clinicians with actionable insights into the factors driving recovery.

This study aims to address these gaps by developing and transparently reporting a framework for passive, longitudinal assessment of post-stroke motor recovery. We selected a CNN-LSTM architecture because its convolutional component can represent local patterns in window-level

*Nkemdirim Chinedu David, Liaoning Petrochemical University, Zhejiang University, Hangzhou, China, Uniquedavido@gmail.com

sensor features, whereas its LSTM component can model temporal dependencies across longitudinal assessments. This compact architecture was appropriate for the available structured time-series representation; the present study does not claim a head-to-head advantage over Transformer or temporal convolutional alternatives. Our objectives were: (1) to identify a practical wearable-sensor configuration for predicting motor-recovery outcomes; (2) to develop a CNN-LSTM model for predicting the FMA score at the subsequent assessment wave; and (3) to interpret influential kinematic features. Direct prediction from T0 to T3 was not evaluated in the results available for this manuscript.

A longitudinal monitoring framework should be judged against three distinct questions. First, the sensor representation must preserve clinically meaningful aspects of movement; second, information collected at one assessment wave must add prognostic value for a later outcome; and third, the resulting prediction must be usable within a clinical decision process. These questions are related but not interchangeable. A strong contemporaneous association between an IMU feature and a clinical score does not by itself demonstrate that the feature predicts later recovery, and a predictive association does not establish a causal mechanism. The present study is therefore positioned as a feasibility investigation of temporally ordered FMA prediction, with the FMA retained as the reference clinical outcome rather than replaced by a digital surrogate.

2. RELATED WORK

2.1. *Wearable Sensor-Based Motor Function Assessment*

Quantitative assessment of motor function is essential for tailoring rehabilitation and tracking recovery after stroke. Wearable inertial measurement units (IMUs) provide an objective complement to conventional clinical scales by characterizing movement during gait, balance, and limb-use activities [9, 10]. Mobile and body-worn sensing technologies have also been employed to quantify functional performance beyond conventional clinical settings [11]. A previous study demonstrated that a minimal IMU configuration could effectively assess stroke-related gait impairment and showed significant correlations with FMA scores [8]. In addition, research on passive digital biomarkers has highlighted the potential of longitudinal wearable sensing for capturing functional behaviors in natural environments [5, 6]. However, evaluating current motor performance alone does not demonstrate whether a model can accurately predict future clinical recovery, which motivates the next-wave prediction task investigated in this study.

The literature further distinguishes between estimating clinical scores from contemporaneous movement data and forecasting future rehabilitation outcomes. Previous research has shown that IMU-based wearable data combined with deep-learning techniques can estimate upper-extremity clinical scores, demonstrating the feasibility of linking movement kinematics to Fugl-Meyer-related assessments [12]. Another study incorporated early wearable-sensor measurements together with admission variables to predict rehabilitation-discharge outcomes and compared sensor-enhanced models with conventional prediction approaches [13]. Although these studies are closely related to the present work, they differ in prediction targets, observation periods, and validation protocols. Consequently, prediction horizon, clinical endpoint, and baseline comparator should be explicitly specified when evaluating sensor-based

rehabilitation models rather than treating them as directly comparable.

2.2. *Machine Learning for Rehabilitation Outcome Prediction*

Systematic reviews have summarized machine-learning approaches for post-stroke outcome prediction and consistently highlighted heterogeneous predictor selection, diverse validation strategies, and the importance of clinically meaningful comparisons [13]. CNNs are effective at learning local feature representations, whereas LSTMs are specifically designed to capture temporal dependencies across sequential observations [15, 16]. Recent studies have explored deep-learning approaches for rehabilitation success prediction, home-discharge modeling, and accelerometry-based assessment of upper-extremity movement [17]. Related time-series prediction studies have also employed RNN-LSTM and time-frequency LSTM architectures [18]. Multimodal prediction models targeting FMA scores are particularly relevant to the present study [19]. Nevertheless, many existing studies focus primarily on classification tasks or contemporaneous assessment rather than transparent next-wave prediction using passive home-monitoring data, and comparisons with conventional statistical or machine-learning baselines remain limited.

Model selection should be justified according to data characteristics rather than architectural novelty alone. Feature-based longitudinal datasets can be analyzed using conventional regression, tree-based methods, temporal convolutional networks, recurrent neural networks, or attention-based models. The CNN-LSTM architecture adopted in this study is motivated by the fixed-duration sensor windows and repeated assessment waves available in the dataset. Specifically, convolutional layers capture local feature patterns, whereas recurrent layers encode temporal dependencies across assessment sequences. This rationale does not imply an inherent superiority of CNN-LSTM over alternative approaches. Existing systematic reviews continue to report substantial heterogeneity in predictors, outcome definitions, and validation strategies across stroke-recovery studies [14]. Therefore, any performance advantage should be demonstrated through identical subject-wise data partitioning, predefined hyperparameter optimization, and transparent comparisons with simpler baseline models.

2.3. *Sensor Configuration Optimization and Model Interpretability*

Long-term deployment of wearable sensing systems requires balancing predictive performance with participant burden. Previous reviews of wearable sensing technologies and lightweight multichannel CNN frameworks for stroke rehabilitation have emphasized the increasing interest in practical sensor configurations; however, the optimal minimal sensor arrangement for longitudinal outcome prediction remains unclear [20, 21]. Recent IMU-based studies have further investigated the prediction of functional recovery and FMA-related outcomes following stroke rehabilitation [19]. In addition to predictive accuracy, model interpretability is essential for clinical adoption. SHAP provides additive contribution values for input features and has been widely applied to explain model behavior at both global and individual levels [22]. Accordingly, this study evaluates different sensor subsets and ranks kinematic features using SHAP while interpreting these feature contributions as explanatory associations rather than causal evidence.

Sensor minimization should not be viewed solely as a hardware-design problem. Reducing the number of wearable sensors can decrease preparation time, minimize placement

errors, simplify charging and data management, and improve the feasibility of repeated home monitoring. However, an overly simplified sensor configuration may fail to capture compensatory movements or asymmetries that are clinically informative for specific patients. Therefore, the key question is not whether a three-sensor configuration is universally sufficient, but whether it can maintain adequate predictive accuracy and interpretability for a well-defined patient population, prediction horizon, and clinical application scenario. This consideration provides the rationale for the exhaustive sensor-combination analysis conducted in the present study while also emphasizing the necessity of future prospective validation.

3. METHODOLOGY

3.1. Study Design and Participants

This longitudinal observational study recruited 85 participants within three months of a unilateral ischemic or hemorrhagic stroke confirmed by clinical evaluation and neuroimaging. Eligibility criteria included age 40–80 years, first-ever stroke with upper- and lower-limb motor deficits, an enrollment impairment range based on FMA-UE and FMA-LE screening, and the ability to provide informed consent. Participants were assessed at baseline (T0) and at 3 (T1), 6 (T2), and 10 months (T3) after enrollment. At each assessment, participants completed a clinical evaluation and a 7-day home-monitoring period using wearable sensors. The FMA was adopted as the clinical outcome measure [23]. A total of 85 participants were enrolled, and 77 participants with complete data were included in the final analysis.

3.2. Wearable Sensor System and Data Acquisition

The wearable system consisted of seven IMU nodes positioned on the sternum, left and right wrists, left and right thighs, and left and right shanks. Each node collected accelerometer, gyroscope, and magnetometer signals at a sampling frequency of 60 Hz. The complete seven-sensor configuration was used throughout each 7-day home-monitoring period.

Continuous home monitoring requires careful management of data quality because signal interruptions may occur during daily use. Missing segments, device removal, communication failure, and implausible signal values can affect data completeness and may also reflect differences in participant engagement or recovery status. Three participants were excluded because of incomplete sensor recordings. Details regarding window-level completeness thresholds and missing-data processing were not available; therefore, only complete recordings were included in the subsequent analyses.

3.3. Signal Segmentation and Feature Extraction

Sensor signals were segmented into 5-s windows, corresponding to 300 samples per channel at 60 Hz. As illustrated in Figure 1, 50 features were extracted from each sensor within every window, yielding up to 350 candidate features for the complete seven-sensor configuration before sensor-subset selection. The extracted features included time-domain descriptors of signal magnitude and variability together with frequency-domain measures of signal complexity. Representative features included the root-mean-square (RMS) of wrist acceleration, the standard deviation of sternum acceleration, and the spectral entropy of thigh gyroscope signals. Signal preprocessing parameters, including filter type, cut-off frequency, coordinate-frame alignment,

outlier handling, and missing-data processing, are not described here.

Window-level features provide an intermediate representation between raw sensor signals and longitudinal prediction models. This representation reduces the dimensionality of multichannel 60-Hz recordings while facilitating comparisons across different sensor locations. It also enables movement characteristics such as magnitude, variability, and signal complexity to be represented in a clinically meaningful manner. However, feature aggregation may reduce sensitivity to short-duration activities and may be influenced by variations in wear time and daily activity patterns.

3.4. CNN-LSTM Model Architecture

As illustrated in Figure 1, the proposed model processed sequential sensor features using one or more Conv1D layers for feature extraction, followed by LSTM layers for temporal modeling and fully connected layers for FMA score regression. Model training employed mean squared error (MSE) as the loss function, the Adam optimizer, and early stopping based on validation loss. The CNN-LSTM architecture integrates local feature extraction with temporal sequence modeling.

3.5. Prediction Task, Data Splitting, and Evaluation Scope

The prediction target was the FMA score at the subsequent assessment wave. Specifically, sensor-derived features collected at T0 were used to predict the FMA score at T1, and the same strategy was applied to the remaining assessment waves. Therefore, the task focused on adjacent-wave prediction rather than direct long-term forecasting.

To preserve the independence of participant data, all observations from the same participant were assigned to a single training, validation, or testing subset. Model performance was evaluated using the held-out test set. Comparisons with conventional regression and machine-learning methods were not included in this study.

All preprocessing procedures, feature normalization, sensor-subset selection, hyperparameter optimization, and early-stopping strategies were performed before model evaluation. Participant-level data partitioning was adopted throughout the analysis to ensure that samples from the same participant were not distributed across different subsets.

3.6. Sensor Combination Optimization

The seven-sensor system generated $2^7 - 1 = 127$ non-empty sensor combinations. Each combination was evaluated, and the best RMSE for each sensor-count category is summarized in Figure 3. The optimal three-sensor configuration consisted of the sternum, affected wrist, and affected thigh.

Sensor combination analysis was conducted to identify an effective sensor configuration while reducing system complexity. Wearable sensor placement influences both monitoring convenience and movement representation. Reducing the number of sensors decreases preparation time and improves usability, whereas excessive reduction may omit movement information associated with motor impairment. Therefore, sensor selection considered predictive performance together with practical applicability for long-term home monitoring.

3.7. Statistical Analysis and Model Interpretation

Predictive performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE). For observed scores y_i and predicted scores $\hat{y}_i$, MAE was calculated as the mean absolute error $|y_i - \hat{y}_i|$, and RMSE was calculated as the square root of the mean squared prediction error. Model interpretation was performed using SHAP values, and the top 20 features were ranked according to their mean absolute SHAP values [22]. SHAP values quantify the contribution of individual features to model predictions and provide insight into feature importance.

R^2 reflects the proportion of variance explained by the prediction model, whereas MAE and RMSE quantify prediction errors using the same units as the clinical outcome measure. MAE represents the average prediction error, while RMSE places greater emphasis on larger prediction errors, providing complementary information for model evaluation.

4. RESULTS

4.1. Participant Characteristics

A total of 85 post-stroke participants were enrolled. Five participants withdrew before the T1 follow-up for personal

reasons, and three participants were excluded because sensor data were incomplete; the final analytic cohort therefore comprised 77 participants. Table 1 summarizes the demographic and baseline clinical characteristics of this analytic cohort.

TABLE I. DEMOGRAPHIC AND BASELINE CLINICAL CHARACTERISTICS OF THE FINAL ANALYTIC COHORT (N = 77).

Characteristic	Value
Age (years), mean $\pm$ SD	65.4 $\pm$ 8.2
Gender (Male), n (%)	48 (62.3%)
Time since stroke (days), mean $\pm$ SD	45.2 $\pm$ 15.7
Stroke Type (Ischemic), n (%)	61 (79.2%)
Baseline FMA-UE Score, mean $\pm$ SD	32.5 $\pm$ 10.1

4.2. Performance of the Deep Learning Model

Using the full seven-sensor configuration, the CNN-LSTM model was evaluated for prediction of the FMA score at the subsequent assessment wave. For the reported test set, the model achieved $R^2 = 0.92$, MAE = 2.85 points, and RMSE = 3.41 points. The architecture and observed-versus-predicted relationship are presented in Figures 1 and 2.

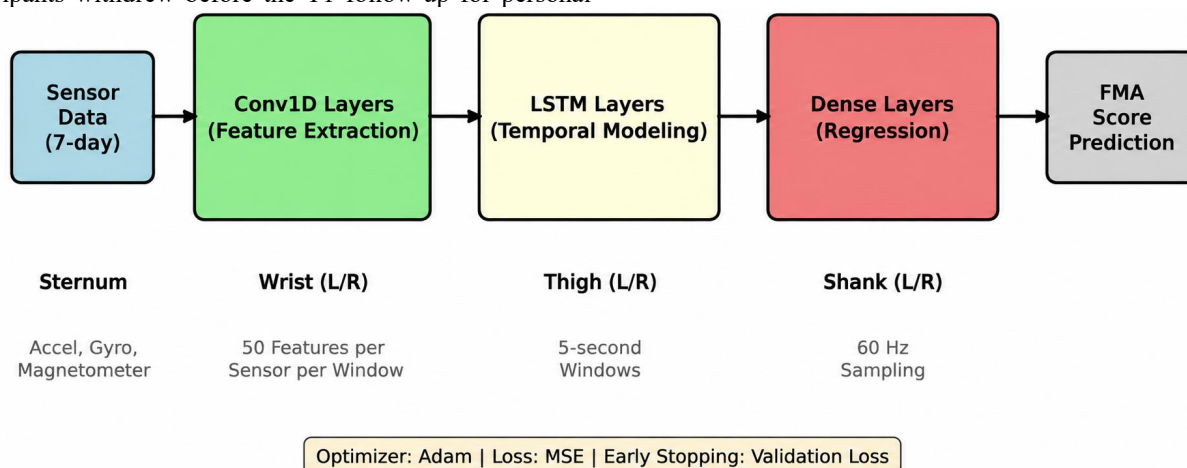


Figure 1. CNN-LSTM Model Architecture for Motor Recovery Prediction

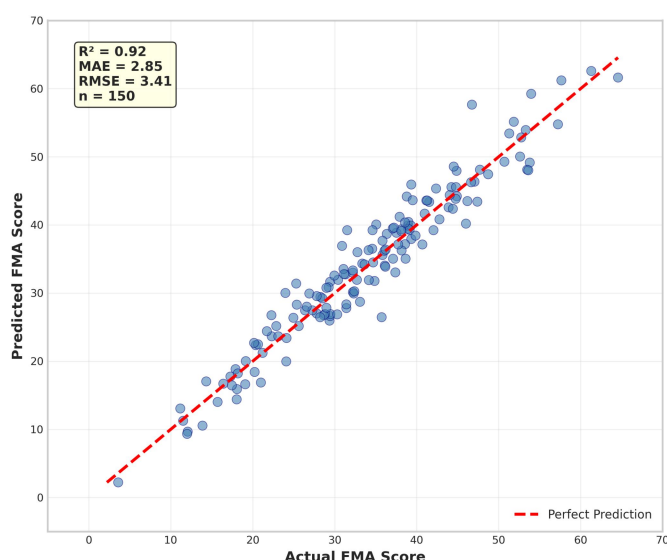


Figure 2. Prediction Accuracy - Predicted vs Actual FMA Scores

The reported metrics should be read jointly. An R^2 of 0.92 indicates a close correspondence between observed and predicted FMA scores within the reported test-set analysis, while the MAE and RMSE express the typical magnitude of error in FMA points. The modestly larger RMSE than MAE is expected when some residuals are larger than the average absolute error, but the current record does not include residual distributions or transition-specific errors. Accordingly, the available results establish internal feasibility for the stated next-wave task but do not identify which patients, impairment levels, or time intervals are most difficult to predict.

4.3. Sensor Optimization Analysis

All 127 non-empty combinations of the seven sensors were evaluated during sensor configuration optimization. As shown in Figure 3, the best RMSE values for one to seven sensors were 8.52, 6.15, 3.58, 3.41, 3.38, 3.35, and 3.41 points, respectively. The three-sensor configuration, consisting of the sternum, affected wrist, and affected thigh, achieved an RMSE of 3.58 points, compared with 3.41 points obtained using all seven sensors, corresponding to a difference of 0.17 points. Considering the small difference in prediction error, the three-

sensor configuration provided performance comparable to that of the full sensor configuration while requiring fewer wearable devices, indicating its potential for practical long-term home monitoring.

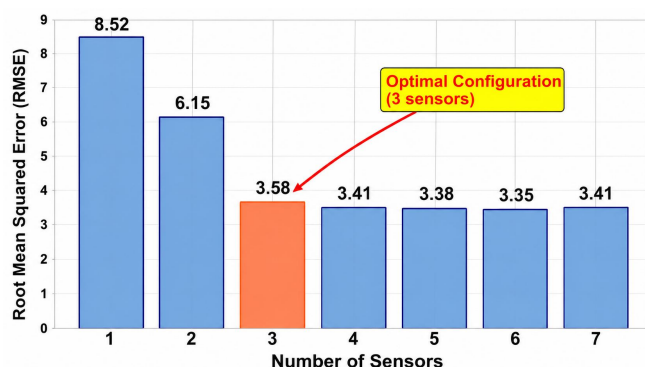


Figure 3. Sensor Optimization - Model Performance vs Number of Sensors

4.4. Feature Importance and Interpretation

SHAP analysis was performed to interpret the predictions of the three-sensor model. Figure 4 presents the top 20 kinematic features ranked by their mean absolute SHAP values. Features derived from the affected upper and lower limbs contributed substantially to model prediction, including the RMS of acceleration from the affected wrist and the spectral entropy of gyroscope signals from the affected thigh. The standard deviation of sternum acceleration also ranked

among the most important features, indicating that trunk movement was associated with model prediction.

The affected wrist and affected thigh provided the highest-ranking features, suggesting that upper-limb motion reflects functional use and motor control, whereas lower-limb dynamics are associated with mobility and balance recovery. Features extracted from the sternum further captured movement patterns related to trunk compensation during daily activities. SHAP values reflect the contribution of each feature within the trained model and may be influenced by correlations among input variables. Consequently, features with similar information may share their contributions, and lower-ranked features may still contain useful information. The SHAP analysis provides an intuitive interpretation of feature importance and supports subsequent clinical analysis.

4.5. Prediction Horizon and Reporting Scope

Model performance was evaluated for prediction at the subsequent assessment wave. The assessment intervals were approximately 3 months between T0 and T1, 3 months between T1 and T2, and 4 months between T2 and T3. Transition-specific RMSE, MAE, and R^2 values were not separately analyzed; therefore, prediction performance across different assessment intervals was evaluated as a whole. The reported results correspond to adjacent-wave prediction with an interval of approximately 3–4 months.

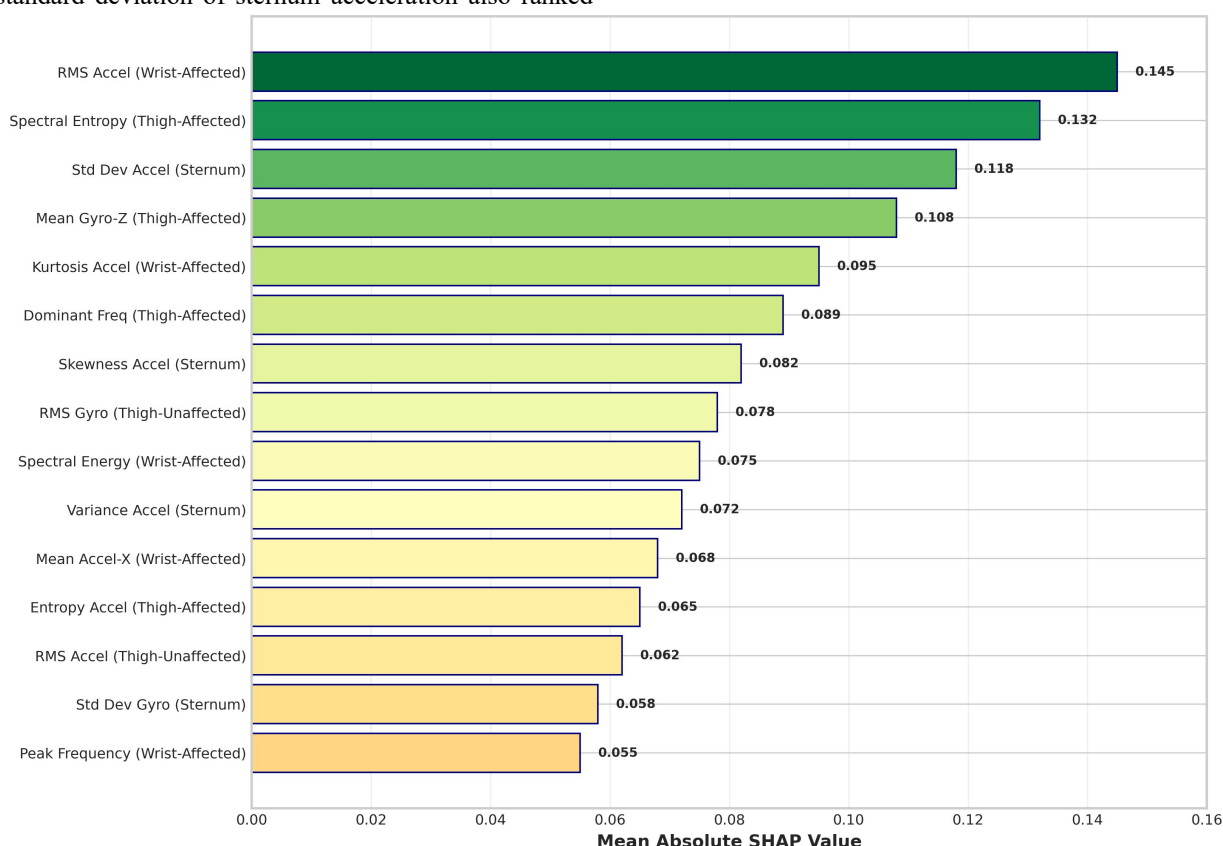


Figure 4. Feature Importance for Motor Recovery Prediction

5. DISCUSSION

This study demonstrates that a CNN-LSTM model combined with a reduced wearable sensor configuration can predict motor-function outcomes at the subsequent assessment

during post-stroke rehabilitation. By integrating 7-day home-monitoring data with subsequent FMA scores, the proposed framework provides an objective approach for longitudinal motor-function assessment. The combination of passive

wearable sensing, sensor configuration optimization, and SHAP-based feature interpretation enables prediction and model interpretation within a unified framework.

The CNN-LSTM model achieved satisfactory prediction performance, indicating that wearable sensor data contain useful information related to subsequent motor recovery. This finding is consistent with previous studies on machine-learning-based prediction of post-stroke rehabilitation outcomes [14] and recent investigations using deep learning, accelerometry, and multimodal sensing [17]. Unlike previous studies that mainly focused on current motor-function assessment [8, 11], this study predicted motor-function outcomes at the next clinical assessment.

The selected three-sensor configuration achieved prediction performance comparable to that of the complete seven-sensor system while substantially reducing the number of wearable devices. The sternum sensor captured trunk movement, whereas sensors on the affected upper and lower limbs provided movement information associated with motor recovery. The small RMSE difference between the three- and seven-sensor configurations indicates that reducing the number of sensors had only a limited influence on prediction performance, supporting the feasibility of lightweight wearable monitoring.

The predicted FMA score may provide additional information for longitudinal rehabilitation assessment and assist clinicians in tracking recovery trends together with routine clinical evaluation. Combining predicted outcomes with SHAP-based feature interpretation further improves the transparency of model predictions and facilitates clinical understanding of the factors contributing to the estimated outcome. Artificial intelligence is expected to complement clinical assessment rather than replace professional rehabilitation practice [24].

The proposed framework can be integrated into routine home-based rehabilitation monitoring by collecting 7-day wearable sensor data before each scheduled follow-up and predicting the FMA score at the subsequent assessment. The predicted score together with SHAP feature importance can provide additional reference information during clinical evaluation. Future studies should further investigate the influence of rehabilitation intensity, medication, comorbidities, and caregiver support on prediction performance.

Model performance may also be influenced by differences in patient characteristics, rehabilitation environments, and sensor deployment conditions. Factors including stroke severity, lesion characteristics, dominant-side involvement, use of walking aids, rehabilitation intensity, and daily activity patterns may affect sensor-derived movement features. Variations in sensor placement and monitoring conditions may also influence prediction accuracy. Further validation using larger cohorts from multiple rehabilitation centers will help evaluate the generalizability of the proposed framework.

6. LIMITATIONS AND FUTURE WORK

Several limitations should be considered when interpreting the findings. First, the final analysis included 77 participants from a single longitudinal cohort, and a larger sample is desirable to further evaluate model performance. Second, detailed information on device acquisition parameters, signal preprocessing, participant-level data partitioning, leave-one-

subject-out validation, and independent external validation was not included. Third, comparisons with conventional machine-learning models and previous-score prediction were not performed, limiting direct evaluation of the CNN-LSTM model against alternative approaches. Fourth, confidence intervals, multiple-comparison analysis, and assessment of participant attrition were not included in the sensor-combination analysis. Finally, although FMA is a widely used clinical assessment, clinician scoring variability, sensor wearing compliance, sensor displacement, and differences in rehabilitation exposure may influence prediction performance. Future studies will include larger multicenter cohorts, standardized data acquisition and preprocessing procedures, participant-level validation, external validation, comparative evaluation with conventional prediction models, assessment across different follow-up intervals, and further evaluation of the three-sensor configuration.

Model interpretation also requires further investigation. SHAP analysis provides feature-level interpretation of model predictions; however, feature importance may vary across different patient populations and rehabilitation conditions. Factors including age, sex, impairment severity, lesion characteristics, mobility aids, and home environments may influence prediction performance. Future studies will further evaluate model performance across different patient subgroups and investigate user acceptance of wearable sensor placement and model interpretation in clinical rehabilitation. Longitudinal data management and clinical use of prediction results will also be further optimized to support routine rehabilitation assessment.

7. CONCLUSION

This study proposed a passive longitudinal framework for post-stroke motor recovery prediction by integrating seven wearable IMUs, a CNN-LSTM model, sensor configuration optimization, and SHAP-based feature interpretation. The proposed model predicted the FMA score at the subsequent assessment with an R^2 of 0.92, an MAE of 2.85 points, and an RMSE of 3.41 points. Among all evaluated sensor combinations, the three-sensor configuration consisting of the sternum, affected wrist, and affected thigh achieved an RMSE of 3.58 points, providing prediction performance comparable to that of the complete seven-sensor configuration while reducing the number of wearable devices.

The proposed framework enables longitudinal prediction of motor-function outcomes using passive home-monitoring data and provides feature-level interpretation through SHAP analysis. The results indicate that reducing the number of wearable sensors had only a limited influence on prediction accuracy, supporting the application of a lightweight wearable monitoring system for post-stroke rehabilitation. Future studies will further evaluate the proposed framework using larger multicenter cohorts, compare its performance with conventional prediction models, and investigate its clinical application in long-term home-based rehabilitation.

REFERENCES

- [1] Johnson, C. O., Nguyen, M., Roth, G. A., Nichols, E., Alam, T., Abate, D., ... & Miller, T. R. (2019). Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 18(5), 439–458. [https://doi.org/10.1016/S1474-4422\(19\)30034-1](https://doi.org/10.1016/S1474-4422(19)30034-1)
- [2] Langhorne, P., Bernhardt, J., & Kwakkel, G. (2011). Stroke rehabilitation. *The Lancet*, 377(9778), 1693–1702. [https://doi.org/10.1016/S0140-6736\(11\)60325-5](https://doi.org/10.1016/S0140-6736(11)60325-5)

- [3] Nakayama, H., Jørgensen, H. S., Raaschou, H. O., & Olsen, T. S. (1994). Recovery of upper extremity function in stroke patients: the Copenhagen Stroke Study. *Archives of physical medicine and rehabilitation*, 75(4), 394-398.[https://doi.org/10.1016/0003-9993\(94\)90161-9](https://doi.org/10.1016/0003-9993(94)90161-9)
- [4] Patel, S., Park, H., Bonato, P., Chan, L., & Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9(1), 21.<https://doi.org/10.1186/1743-0003-9-21>
- [5] Adans-Dester, C., Hankov, N., O'Brien, A., Vergara-Diaz, G., Black-Schaffer, R., Zafonte, R., ... & Bonato, P. (2020). Enabling precision rehabilitation interventions using wearable sensors and machine learning to track motor recovery. *NPJ digital medicine*, 3(1), 121.<https://doi.org/10.1038/s41746-020-00328-w>
- [6] Matias, I., Haas, M., Daza, E. J., Kliegel, M., & Wac, K. (2026). Digital biomarkers for brain health: passive and continuous assessment from wearable sensors. *npj Digital Medicine*.<https://doi.org/10.1038/s41746-026-02340-y>
- [7] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.<https://doi.org/10.1038/nature14539>
- [8] Sun, Y., Song, Z., Mo, L., Li, B., Liang, F., Yin, M., & Wang, D. (2025). IMU-Based quantitative assessment of stroke from gait. *Scientific Reports*, 15(1), 9541.<https://doi.org/10.1038/s41598-025-94167-y>
- [9] Brognara, L., Palumbo, P., Grimm, B., & Palmerini, L. (2019). Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review. *Diseases*, 7(1), 18.<https://doi.org/10.3390/diseases7010018>
- [10] Parnandi, A., Wade, E., & Matarić, M. (2010, August). Motor function assessment using wearable inertial sensors. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 86-89). IEEE.<https://doi.org/10.1109/IEMBS.2010.5626156>
- [11] Fell, N., True, H. H., Allen, B., Harris, A., Cho, J., Hu, Z., ... & Salstrand, R. (2019). Functional measurement post-stroke via mobile application and body-worn sensor technology. *Mhealth*, 5, 47.<https://doi.org/10.21037/mhealth.2019.08.11>
- [12] Zhou, W., Fu, D., Duan, Z., Wang, J., Zhou, L., & Guo, L. (2025). Achieving precision assessment of functional clinical scores for upper extremity using IMU-Based wearable devices and deep learning methods. *Journal of NeuroEngineering and Rehabilitation*, 22, 84.<https://doi.org/10.1186/s12984-025-01625-9>
- [13] O'Brien, M. K., Lanotte, F., Khazanchi, R., Shin, S. Y., Fanton, M., Lieber, R. L., Ghaffari, R., & Jayaraman, A. (2024). Early prediction of poststroke rehabilitation outcomes using wearable sensors. *Physical Therapy*, 104(2), pzad183. <https://doi.org/10.1093/ptj/pzad183>
- [14] Zu, W., Huang, X., Xu, T., Du, L., Wang, Y., Wang, L., & Nie, W. (2023). Machine learning in predicting outcomes for stroke patients following rehabilitation treatment: A systematic review. *Plos one*, 18(6), e0287308.<https://doi.org/10.1371/journal.pone.0287308>
- [15] Hochreiter, S. (1997). Long short-term memory. *Neural Computation* MIT-Press.<https://doi.org/10.1162/neco.1997.9.8.1735>
- [16] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.<https://doi.org/10.1145/3065386>
- [17] Mahmoud, Y., Horvath, K., & Zhou, Y. (2025). Deep learning for predicting rehabilitation success: Advancing clinical and patient-reported outcome modeling. *Electronics*, 14(6), 1082.<https://doi.org/10.3390/electronics14061082>
- [18] Cui, Y. (2024). An efficient approach to sports rehabilitation and outcome prediction using RNN-LSTM. *Mobile Networks and Applications*, 1-16.<https://doi.org/10.1007/s11036-024-02355-3>
- [19] 19Dipietro, L., Eden, U., Teixeira, P., Tirasawasdichai, N., Warinpramote, J., Pundick, S., ... & Wagner, T. (2025). A multi-modal machine learning approach to predict fugl-meyer scores and motor recovery potential in stroke rehabilitation: Toward precision-based therapies. *Information sciences*, 122564.<https://doi.org/10.1016/j.ins.2025.122564>
- [20] Woelfle, T., Bourguignon, L., Lorscheider, J., Kappos, L., Naegelin, Y., & Jutzeler, C. R. (2023). Wearable sensor technologies to assess motor functions in people with multiple sclerosis: systematic scoping review and perspective. *Journal of Medical Internet Research*, 25, e44428.<https://doi.org/10.2196/44428>
- [21] Wang, J., Li, C., Zhang, B., Zhang, Y., Shi, L., Wang, X., ... & Xiong, D. (2024). Automatic rehabilitation exercise task assessment of stroke patients based on wearable sensors with a lightweight multichannel 1D-CNN model. *Scientific Reports*, 14(1), 19204.<https://doi.org/10.1038/s41598-024-68204-1>
- [22] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
- [23] Fugl-Meyer, A. R., Jääskö, L., Leyman, I., Olsson, S., & Steglind, S. (1975). A method for evaluation of physical performance. *Scand J Rehabil Med*, 7(1), 13-31.<https://doi.org/10.2340/165019771331>
- [24] Hao, J. (2025). Artificial intelligence empowers physical therapy for neurological conditions. *Neurological Sciences*, 46(10), 5573-5575.<https://doi.org/10.1007/s10072-025-08295-4>

ACKNOWLEDGEMENTS

None.

FUNDING

None.

AVAILABILITY OF DATA

Not applicable.

ETHICAL STATEMENT

The study protocol was reviewed and approved by the Ethics Committee. All participants provided written informed consent prior to participation. All procedures involving human participants were conducted in accordance with the ethical standards of the institutional ethics committee and relevant ethical guidelines and regulations.

AUTHOR CONTRIBUTIONS

Nkemdirim Chinedu David conceived the study, developed the methodology, implemented the CNN-LSTM model, performed data analysis, and wrote the original manuscript. Diallo Abdoul Salam supervised the research, contributed to study design and methodology, reviewed and revised the manuscript, and approved the final version. All authors have read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

Publisher's note WEDO remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is published online with Open Access by BIG.D and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

© The Author(s) 2026