Multimodal Sequence Classification of force-based instrumented hand manipulation motions using LSTM-RNN deep learning models

Abhinaba Bhattacharjee, Sohel Anwar, Lexi Whitinger, M. Terry Loghmani

***Abstract*—The advent of mobile ubiquitous computing enabled sensor informatics of human movements to be used in modeling and building deep learning classifiers for cognitive AI. Expanding deep learning approaches for classifying instrumented hand manipulation tasks, especially the art of manual therapy and soft tissue manipulation, can potentially augment practitioner's performance and enhance fidelity with computer assisted guidelines. This paper introduces a dataset of 3D force profiles and manipulation motion sequences of controlled soft tissue manipulation stroke pattern applications in thoracolumbar, upper thigh and calf regions of a single human subject performed by five experienced manual therapists. The multimodal 3D force, 3D accelerometer and resultant gyro raw data were preprocessed and experimentally fed into a multilayered Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) deep learning model to observe sequence classifications of two manipulation motion techniques (Linear "Strumming" motion and curvilinear "J-Stroke" arched motion) of manual therapy performed using a handheld, localizing Quantifiable Soft Tissue Manipulation (QSTM) medical tool. Each of these motion sequences were further labeled with corresponding best practice technique from validated video tapes and reclassified into "Correct" and "Incorrect" practice based on defined criteria. The deep learning model resulted in 90-95% classification accuracy for individual intra-therapist reduced dataset. The classification accuracy varied between 78%-93% range, when trained with multivariate characteristic feature set combinations for the complete spectrum of inter-therapist dataset.**

***Clinical Relevance* — AI informed online therapeutic guidelines can be leveraged to minimize practice inconsistencies, optimize educational training of therapy using data informed protocols, and study progression of pain and healing towards advancing manual therapy.**

Keywords—Motion Sequence Classification, Long Short-Term Memory, Recurrent Neural Networks, Deep Learning.

I. INTRODUCTION

Human behavior studies have been a field of research in the domain of kinesiology for decades. Human reflexes in response to stimuli for environmental perception, self-adaptation and maneuvering define the state of actions in human activities. Mathematical computation of human motion states has been possible by analyzing multivariate sequential outputs of modern-day inertial sensors integrated in wearables or handheld tools recording associated activities. Triaxial force sensors enable tactile perception [1] of textures and allow pressure adjustments in manipulation tasks using haptic feedback [2]. Deep learning techniques have crept in popular health monitoring applications ranging from physical exercise tracking with gait recognition [3] to computer assisted rehabilitation using exoskeletons [4], and developing data informed assistive intelligence in body sensor networks [5]. In context, Convolution Neural Networks has proved to achieve promising results in video-based activity recognition [6] with 3D depth information. Furthermore, Recurrent Neural Networks has performed optimally well to classify multimodal time series sequences acquired by wearables for human activity recognition [7] and speech analysis [8].

Similar deep learning techniques can be extended to analyze force regimens and temporal motion sequences used for instrument-assisted digital manual therapy. The Quantifiable Soft Tissue Manipulation (QSTM) medical device system is a precursor of this work which includes two handheld medical tools (localized [9] and dispersive [10] QSTM applicators) for treating musculoskeletal pain conditions. Pre-clinical studies using these tools resulted in recording the massage routines comprising dynamic force profiles and hand motion sequences of soft tissue palpation and treatment by a custom-made data requisition software (Q-Ware©) [11]. An observational preclinical study of 15 experienced therapists, each with more than 8 years clinical experience in instrument-assisted soft tissue manipulation (IASTM) was conducted with institutional review board approvals. Each therapist performed two massage motion types on the same human subject's thoracolumbar, upper thigh and calf areas of left and right side of body on different days for subject comfort. After careful graphical and visual assessments individual datasets of 5 therapists were selected for multimodal sequence classifications out of 15 therapists. These temporal sequences were labelled and fed into a deep learning model for identical sequence classifications on untrained data. The deep learning model is comprised of 3 layered Recurrent Neural Networks engaging a Long-short Term Memory (LSTM) layer [12], a fully connected layer and a soft max layer to perform the training.

This paper introduces a multivariate time series dataset of two fundamental hand manipulation motion sequences of manual therapy (namely: Linear "Strumming" straight motion and Curvilinear "J-Stroke" arched motion) which are further classified based on the technique of application as indicated by practice guidelines [13]. A total of four fundamental classes with associated motion types and techniques indicating "Strumming-Correct", "Strumming-Incorrect", "J-Stroke-Correct" and "J-Stroke-Incorrect" were selected and labeled on the inter-therapist dataset based

Abhinaba Bhattacharjee, and Sohel Anwar are with the Department of Mechanical and Energy Engineering, Purdue School of Engineering and Technology, Purdue University Indianapolis, IN 46202 USA {email: abhbhatt@iupui.edu, soanwar@iupui.edu}.

Lexi Whitinger is with the Department of Biomedical Engineering, Purdue School of Engineering and Technology, Purdue University Indianapolis, IN 46202 USA {email: ajwhitin@iupui.edu}.

M. T. Loghmani, is with the Department of Physical Therapy, Indiana University School of Health and Human Sciences, Indiana University Indianapolis, IN 46202 USA, {e-mail: mloghman@iu.edu}.

This Research was initially supported by Indiana Clinical Translational Sciences Institute, Indiana Center for Biomedical Innovation & Indiana University Research and Technology Corp, and Indiana University Purdue University Indianapolis (IUPUI). This project was supported in part by the NIH National Center for Complementary & Integrative Health (Award Number R41AT011494). The content is the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

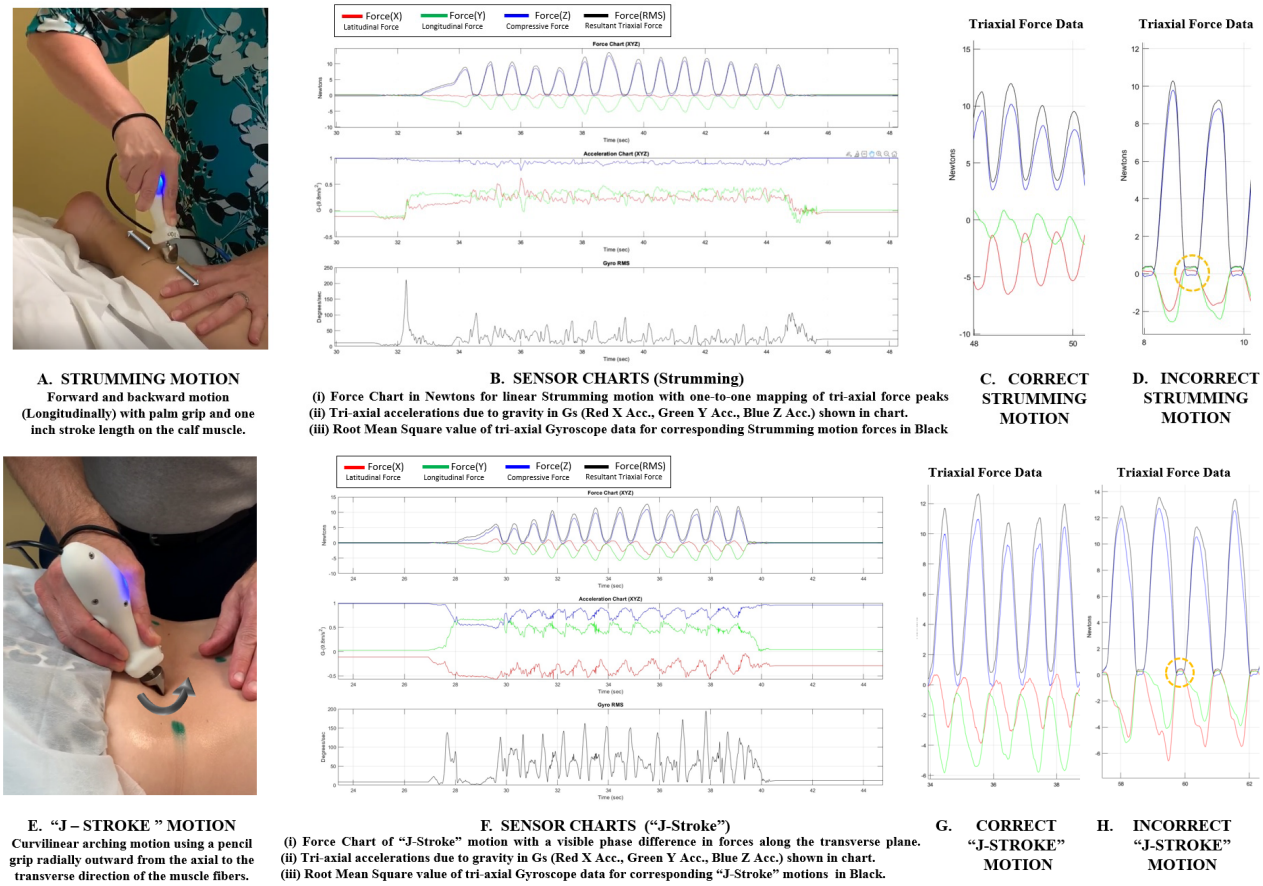


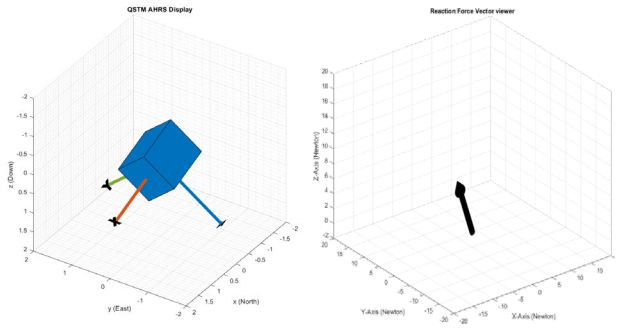
Figure 1. Manipulation motions of Manual Therapy using the localized QSTM applicator following the Strumming and "J" type turn sequences and their corresponding multichannel data depicting correct and incorrect techniques of practice.

on animated data observations validated by their corresponding recorded videotapes. An additional fifth class "Random-N/A" is labeled for motions associated to preparatory and tool positioning random sequences on the subject's body with no applicable (N/A) techniques. Feature extraction of multimodal temporal sequences is usually performed by handpicking characteristic attributes based on visual observations and requires domain knowledge. Supervised deep learning models with higher processing capacities serves the purpose by feeding raw sensor data sequences. In our case, a combination of features including tri-axial force; angular pose defined by Yaw, Pitch and Roll sequences; triaxial acceleration and gyro readings of angular velocity were experimentally chosen and fed into the deep learning model to perform the classification of motion sequences and examine its relative accuracy on a prepared testing dataset.

II. Dataset Preparation and Labelling

A. Manual Therapy Motions and Techniques

The mentioned manipulation motion types linear "Strumming" motion and curvilinear "J-Stroke" arched motion with their characteristic multimodal data sequences of triaxial force, acceleration and gyro data are shown in Fig. 1(a)-(d) and Fig. 1(e)-(h), respectively. Strumming refers to the back-and-forth longitudinal traversal (Fig. 1(a)) of the tooltip along or across the muscle fibers palpating ridges or rocky subsurface anatomies of the fascia. Whereas the "J-Stroke" indicates a force sustained 90º turn on the skin surface from the axial to transverse directions of muscle fibers engaging stimulations and blood flow in different underlying muscles and connective tissues. Hence, these



(a) (b)

Figure 2. Animation viewers (a) Helper Orientation Viewer (MATLAB) to view the Euler Angles or quaternions of angular pose of the motion sequence. (b) The Arrowhead denotes the magnitude of compressive force and the arrow tilts and grows based on the resultant force magnitude.

two soft tissues manipulation motion sequences are fundamental for the localized QSTM applicator tool in treating musculoskeletal pain. However, manipulation motion sequences using the dispersive QSTM applicator [14] is beyond this paper's scope.

B. Visual Data Observations and Labelling

The Dataset prepared from the preclinical study includes performed fundamental "Strumming" linear and "J-Stroke" turning motions using the localized QSTM applicator by several practitioners (of which best five were chosen) on three body regions (lower back, upper thigh and calf) of both left and right side of the subject's body. In total, every practitioner performed 12 trials, each of 15 seconds for each individual manipulation motion. Graphical observation of the triaxial force data (Fig. 1(b)) of the "Strumming" motion sequence represents one-to-one mapping of the local

compressive force maxima with corresponding local lateral or longitudinal force max or mins respectively. However, a visible phase difference (Fig. 1(f)) along the planar forces is noticed along the graphical representations of the J-Type turning motion sequences. The correct techniques of practicing these motions engage a ramped up sustained pressure on and off, back and forth the linear or curved trajectories, respectively, without losing contact of the tooltip with the skin for the complete trial time. The practice sequences of correct manipulation techniques are well represented by tri-axial force charts in Fig. 1(c) and 1(g) for Strumming and J-Stroke motions. If the contact of tooltip and skin breaks during the motion cycle, then it is considered as incorrect technique of performing the motion sequence. The overlap of triaxial-force data at the zero-crossings of every sequence cycle indicates tooltip-skin contact releases and are denoted by the dashed yellow circle in respective force charts of Fig. 1(d) and 1(h) representing incorrect techniques of practice.

Hence for labelling the dataset with respective motion type and technique we manually approached the leave one out cross validation (LOOCV) [15] process by animating the raw data. Two animations – one for the angular pose using *Helper Orientation Viewer* of the *Sensor Fusion Toolbox* in MATLAB for assessing the motion type shown in Fig. 2(a); and another custom-made 3D reaction force vector animation using *Plot3()* function to visualize the tooltip-skin contact release shown in Fig. 2(b), determining the technique's correctness, were performed. These animations were further validated by their corresponding videotapes recorded during data collection. The animations assisted in identifying the state change timestamps both visually and graphically for individual motion sequence types and techniques in the raw datafiles. Subsequently, each datapoint on every timestamp was labelled with their respective categorical classes, forming a total of five classes. Thus, two motion types and two techniques make a combination of four classes "Correct Strumming" and "Incorrect Strumming"; and "Correct J-Stroke" or "Incorrect J-stroke", while "Random N/A" class indicates the unknown sequences which couldn't be categorized.

C. *Dataset Preparation for Deep Learning*

Deep learning classifiers require the complete dataset to split into training and testing data where each training vector needs to include feature specimens from all labelled classes for optimum convergence of the solver. Raw datafiles of our internal dataset included maximum of two classes which were either unknown data of "Random N/A" class or any one of the other four. Hence, these raw files needed to be sorted to form training vectors with all labelled classes. The raw input contained high frequency noise, except for the triaxial force data, which were significantly eliminated by a moving average or rolling mean filter of window size 25 datapoints. The filtered data were segregated into body region specific manipulation motion sequences of both left and right sides for five experienced therapists, out of which two fundamental motion sequences linear "Strumming" and curvilinear "J-Stroke" motions were cross validated from graphs, videotapes and vector animations to be labeled as correct or incorrect techniques of the fundamental motion sequences respectively. Hence, preparation of training dataset for the deep learning model was challenging and tedious. To compensate for feature specimens from all available classes into each training vector, the raw sensor data extracted for individual body regions were shuffled based on the labelled data and concatenated such that each training vector contains adequate proportions of all 5 classes of labelled data. Fig. 3 is a visual representation of the training and testing vectors of the inter-therapist dataset along with their corresponding label distribution and proportion statistics for each of the five class labels. A total of five training vectors were generated, each of which were formed by appending data from at least two identical motion sequences of all five individual class labels performed by different therapists in different body regions. Whereas the

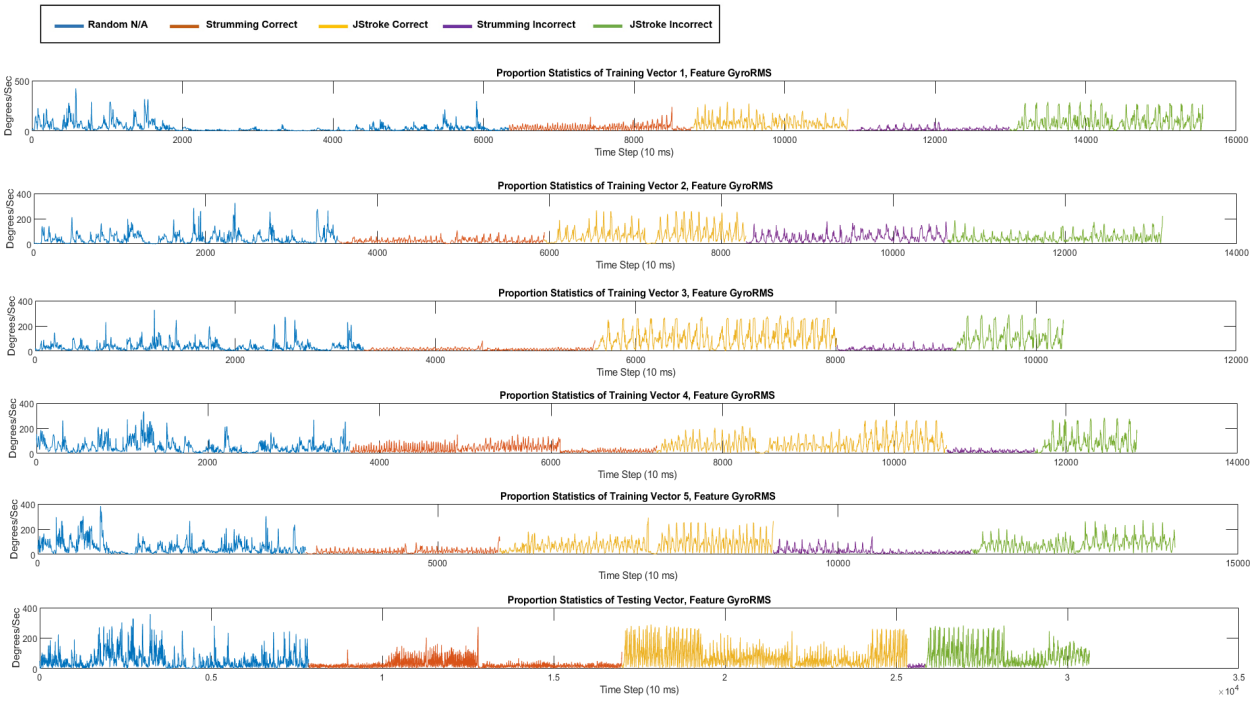


Figure 3. Segregation of the Dataset into training and testing vectors with the corresponding proportion statistics and distributions of the five labelled classes in each training and testing vector. The inter-therapist dataset is compiled into five training vectors and one testing vector represented with their corresponding labels. Gyroscope readings are considered as feature vector to visually distinguish between "Strumming" and "J-Stroke" motion sequence labels as observed from the difference in peak amplitude of the waveforms.

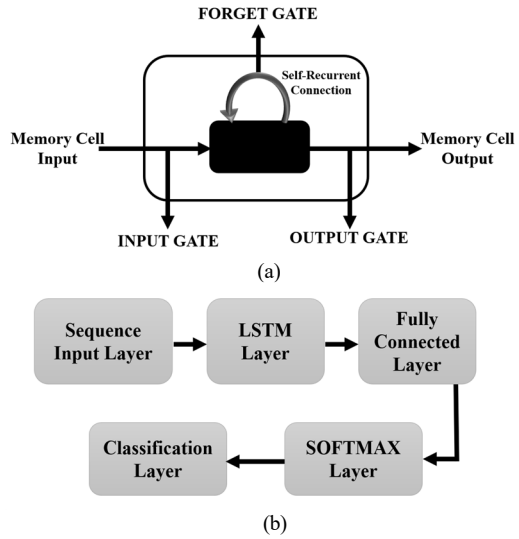


Figure 4. (a) Memory Cell Unit of a LSTM Layer. (b) Implemented Layers of the deep learning LSTM-RNN classification model.

Figure 5. Confusion Matrix of the Predicted and Trained Classes of data for individual intra-therapist reduced dataset, training a combination of 7 features producing the best class prediction accuracy.

remaining unused labelled data was compiled into a testing vector to complete the dataset preparation of the learning model. The feature vector shown in Fig. 3 represents the root mean square of triaxial gyroscope readings (GyroRMS), from which the characteristic traits from fundamental motion sequences "Strumming Correct" and "J-Stroke Correct" can be visually demarcated from the waveform amplitudes of colored labelled classes.

III. Deep Learning Model Development

Conventional Recurrent Neural Networks suffer from cumulatively incremental (exploding) or decremental (vanishing) gradient [16] problems associated with the error magnitudes of loss function during model training. This issue severely impacts the training process as the biases of the training vectors are not updated. Tracking for Long- and short-term updates of these training vector biases into memory cells significantly improves the vanishing and exploding gradient problems. Hence, an LSTM layer comprising a memory cell unit [17] shown in Fig.4(a) overcomes this issue by storing the state of the changing gradients. It allows the input gate to retain or change its state impacting other connected neurons through the output gate. The forget gate enables modulation of the self-recurrent connection by remembering or forgetting its last state as necessary. Finally, the classification layer categorizes an unknown dataset with predicted classes trained on known feature sets.

The Deep Learning Toolbox of MATLAB is used to build the LSTM-RNN classification model where the specific layers mentioned in Fig. 4(b) are structured and initialized. The sequence input layer is fed with raw data signals as input features, transmitted through 200 hidden units of the LSTM layer followed by feature classification at the fully connected SoftMax layers. Finally, the classification layer categorizes an unknown dataset with predicted classes trained on known feature sets.

The hyperparameters of the deep learning models i.e. learning rates [18] and drop-out rates [19] severely depend on the chosen training options and controls the error magnitudes of the loss function during training. The selected Adam optimization algorithm [20] which is appropriate for large datasets minimizes errors by optimizing the biases for training vectors and maximizes accuracy. The gradient threshold was set to 1, assisted with training vector reshuffling at each epoch for a total of 200 epochs. The training was performed on 6GB of dedicated GPU (NVIDIA

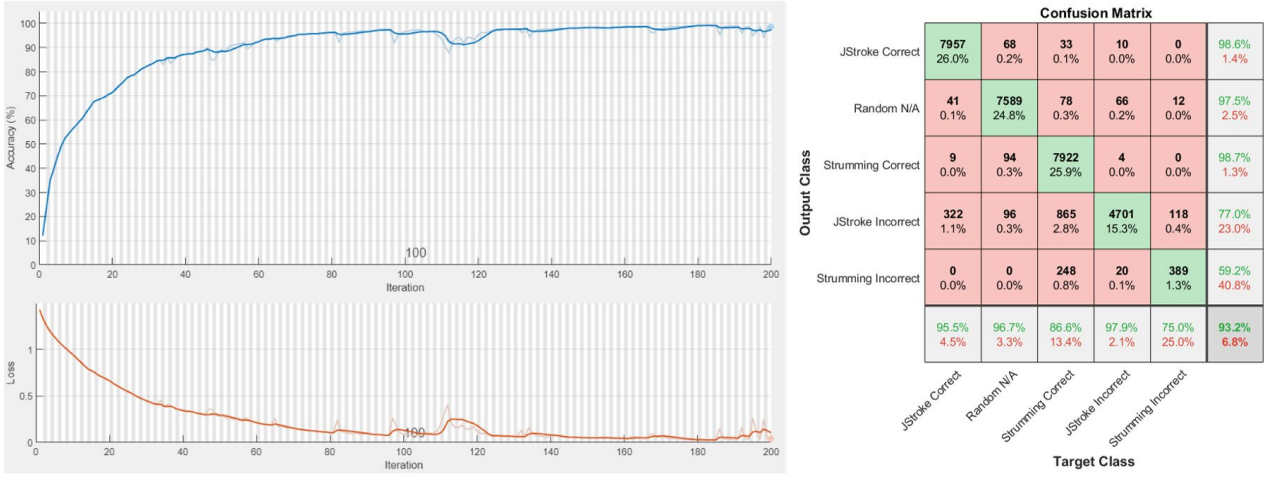


Figure 6. (a) Training Progress of the Deep Learning model for 200 epochs with 7 Feature sets (Triaxial Force, Triaxial Acceleration and GyroRMS). (b) Confusion Matrix of the Predicted vs Trained Classes of data using a combination of 7 features producing a prediction accuracy of 93.2%.

GeForce GTX 1060) followed by predicting classes of the test dataset to achieve classification accuracy of the model.

IV. Experiments and results

The training and testing process was experimentally conducted by feeding an escalating combination of individual features into the training network to investigate the best group of features for suitable classifications. Initially features from intra therapist raw data were fed into the model to examine the network performance in smaller dataset. Intra-therapist data for individual therapist comprise of only two known classes either "Random N/A" and "Correct Strumming", or "Random N/A" and "Correct J-Stroke" from four raw data files, one for Strumming and one for J-Stroke on both left and right sides of the body. Hence, data acquired on the left side of the subject's body served as training dataset, while that of the right side of the body was assigned for testing the model's accuracy. Results from these experiments proved the model to achieve high accuracy in the range of 90-95% prediction rate for a total of 7 characteristic feature set training; the best-case confusion matrix is shown in Fig. 5.

These positive results fueled our study to train the network on the prepared training dataset of the full inter-therapist spectrum with five fundamental classes as shown in the confusion matrix of Fig. 6(b) with the loss and training accuracy of the best trained outcomes in Fig. 6(a). The multimodal feature combinations that were experimentally selected for training the deep learning model on the complete spectrum of inter-therapist dataset are enlisted in Table. 1.

TABLE I. Performance of learning model based on Input features

Classification Accuracy of the LSTM-RNN Deep Learning Model			
No. of Features	**Features Sets**	**Accuracy**	**Convergence and Sensitivity Parameters**
2	Gyro RMS and Force RMS	78.2%	Training Accuracy: 95% Loss: 0.1 Micro AUC: 0.9198 Macro AUC: 0.9252 Mean TPR: 0.8282
3	Tri axial Forces (X, Y, Z)	84.4%	Training Accuracy: 93% Loss: 0.2 Micro AUC: 0.9547 Macro AUC: 0.9548 Mean TPR: 0.8327
5	Tri axial Forces (X, Y, Z); Resultant Force and GyroRMS	90.0%	Training Accuracy: 97% Loss: 0.1 Micro AUC: 0.9687 Macro AUC:0.9544 Mean TPR: 0.8549
7	Tri-axial Force (X,Y,Z); Tri-axial Acceleration (X,Y,Z); GyroRMS	***93.2%***	Training Accuracy: 98% Loss: 0.15 Micro AUC: 0.9876 Macro AUC: 0.9833 Mean TPR: 0.8678
11	Tri-axial Force (X,Y,Z); ForceRMS; Angular Pose (Yaw, Pitch Roll); Tri-axial Acceleration (X,Y,Z); GryoRMS.	85.3%	Training Accuracy: 100% Loss: 0.1E-2 Micro AUC: 0.9432 Macro AUC: 0.9571 Mean TPR: 0.8448

The training experiments proved that the loss function errors in most of the experiments converged almost to zero as the number of features being fed into the model increased. However, the training accuracies for all experiments reached a level of 90% – 100% with an initial learning rate of 0.01. The sensitivity metrics of the prediction results of the deep learning model were measured in terms of the Receiver Operator Characteristics (ROC), in which particularly the micro and macro Area Under the Curve (AUC) of the ROC plots discriminating true positive rate (TPR) versus the false positive rate (FPR) are reported in Table 1. The prediction accuracy of the model trained with combinations of multivariate sensor metrics as feature sets improved experimentally by escalating the number of characteristic features from two to seven as evident from the table. The classification results are reliant on critical spatial characteristics of multivariate waveform patterns in the sequences. The cross validation of the raw data reproducing animations of the motion sequences showed great variances in feature attributes. Experimental training of the model with derived features, in particular angular pose sequences of Strumming and J-Stroke patterns, produce no significant characteristic differences resulting in least prediction accuracy. Whereas that of the Resultant Gyro (GyroRMS) feature shows a significant change in amplitude for J-Stroke waveforms as compared to Strumming waveforms visible from the sensor charts in Fig. 1 and training vector proportion statistics in Fig. 3. Model training with too many features, especially derived features like angular orientation/pose, diminish the effect of fundamental attributes contributed by the characteristic feature sets (raw sensor data) on the prediction accuracy for identical sequence or pattern classification. This reveals the reason for the decreased accuracy (85.3%) of the model when trained with all available features. Therefore, the resulting deep learning model training experiments assist in identifying the best combination of feature sets to achieve higher prediction accuracy. However, there is a necessity for dimensionality reduction using feature transformations for characteristic attribute extraction in the ground truth data, as well as validating the training network with a validation dataset to optimize model performance for real-time classifications.

V. Future works

Time series classifications for motion sequence or human activity detection involve challenges based on characteristic spatiotemporal traits observed in shapes of waveforms to classify an identical pattern. In particular, the dataset described in this study incorporates observational shape variations of the characteristic waveforms based on the multiple physical attributes related to both reaction force and manipulation motion measurements. Hence, fundamental investigation is necessary to improve the accuracy of the model performance. An intermediate multiscale shapelet transform layer [21] or elastic distance tracking to map the peak-valley pairs using Dynamic Time Warping [22] may be beneficial to capture the spatial characteristics of the waveform patterns which can then be fed into our current LSTM RNN architecture with data validation vectors to ensure model enhancements. More improvements can be examined by experimenting with the shapelets based on reduced multivariate feature sets with collectives of transformation-based ensembles (COTE) [23] techniques and upgrading the LSTM-RNN architecture to Transformer [24] based deep learning

models. Investigation of our current and extended clinical dataset with the aforementioned approaches are in place for future works to examine the real-time classification of the hand motion sequences of therapists during both educational training and therapeutic treatment.

VI. Conclusion

A multi-sensory time series dataset of force-based multimodal hand manipulation motions from preclinical manual therapy is presented for sequence classifications. The dataset was preprocessed, analyzed and segregated into training and testing feature vectors to develop a classification model using deep learning LSTM RNN architecture. A combination of raw sensor data features has been introduced for training the deep learning network and estimating the model prediction accuracy on the testing datasets. The model trained with seven fundamental features showed the best performance with classification accuracy of 93.2% for the complete spectrum of inter therapist dataset. The classification performance can be further improved by feature transformations in the preprocessing step followed by a validation dataset entry during training or approaches mentioned in the future works section. Building such deep learning models for digital manual therapy opens new horizons in the field for data driven AI guided clinical training. AI assisted digital therapy can prove to be a practice enhancement tool for both novice and experienced therapists curbing the variance in therapeutic practice with precision and fidelity. Pain affected musculoskeletal areas of the body are susceptible to tolerate less dynamic force as compared to non-involved regions. Progression of force tolerance and pressure endurance can prove to be a tool to track the healing effect of pain conditions. Thus, AI assisted digital manual therapy is necessary to advance healthcare.

Supplementary Materials

The dataset and the code repository for this study is available in the GitHub Repository Link: https://github.com/AbhiBjee/Multimodal-Sequence-Classifications-with-Deep-Learning

References

[1] Okamoto, Shogo, Hikaru Nagano, and Yoji Yamada. "Psychophysical dimensions of tactile perception of textures." *IEEE Transactions on Haptics* 6, no. 1 (2012): 81-93.

[2] Danion, Frédéric, Jonathan S. Diamond, and J. Randall Flanagan. "The role of haptic feedback when manipulating nonrigid objects." *Journal of neurophysiology* 107, no. 1 (2012): 433-441.

[3] Wan, Changsheng, Li Wang, and Vir V. Phoha, eds. "A survey on gait recognition." *ACM Computing Surveys (CSUR)* 51, no. 5 (2018): 1-35.

[4] Bouteraa, Yassine, Ismail Ben Abdallah, and Ahmed M. Elmogy. "Training of hand rehabilitation using low cost exoskeleton and vision-based game interface." *Journal of Intelligent & Robotic Systems* 96 (2019): 31-47.

[5] Chetty, Girija, and Matthew White. "Body sensor networks for human activity recognition." In *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 660-665. IEEE, 2016.

[6] Yao, Guangle, Tao Lei, and Jiandan Zhong. "A review of convolutional-neural-network-based action recognition." *Pattern Recognition Letters* 118 (2019): 14-22.

[7] Pienaar, Schalk Wilhelm, and Reza Malekian. "Human activity recognition using LSTM-RNN deep neural network architecture." In *2019 IEEE 2nd wireless africa conference (WAC)*, pp. 1-5. IEEE, 2019.

[8] Liu, Chaojun, Yongqiang Wang, Kshitiz Kumar, and Yifan Gong. "Investigations on speaker adaptation of LSTM RNN models for speech recognition." In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5020-5024. IEEE, 2016.

[9] Bhattacharjee, Abhinaba, Stanley YP Chien, Sohel Anwar, and Mary T. Loghmani. "Quantifiable Soft Tissue Manipulation (QSTM™)–A novel modality to improve clinical manual therapy with objective metrics." In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 4961-4964. IEEE, 2021.

[10] Bhattacharjee, Abhinaba, Sohel Anwar, Stanley Chien, and M. Terry Loghmani. "A handheld Quantifiable Soft Tissue Manipulation device for tracking real-time dispersive force-motion patterns to characterize Manual Therapy treatment." *IEEE Transactions on Biomedical Engineering* 70, no. 5 (2022): 1553-1564.

[11] Bhattacharjee, Abhinaba. "A Data Requisition Treatment Instrument For Clinical Quantifiable Soft Tissue Manipulation." PhD diss., Purdue University Graduate School, 2021.

[12] Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." *Physica D: Nonlinear Phenomena* 404 (2020): 132306.

[13] Graston Technique Training Manual, 4th ed., 2014.

[14] Bhattacharjee, Abhinaba, Terry Loghmani, and Sohel Anwar. "Finite Element Simulation and Analysis of Drop Tests to Improve the Mechanical Design of a Handheld QSTM Medical Device." In *ASME International Mechanical Engineering Congress and Exposition*, vol. 86663, p. V004T05A008. American Society of Mechanical Engineers, 2022.

[15] Cawley, Gavin C., and Nicola LC Talbot. "Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers." *Pattern recognition* 36, no. 11 (2003): 2585-2592.

[16] Grosse, Roger. "Lecture 15: Exploding and vanishing gradients." *University of Toronto Computer Science* (2017).

[17] Pulver, Andrew, and Siwei Lyu. "LSTM with working memory." In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 845-851. IEEE, 2017.

[18] Lin, Huangxing, Weihong Zeng, Yihong Zhuang, Xinghao Ding, Yue Huang, and John Paisley. "Learning rate dropout." *IEEE Transactions on Neural Networks and Learning Systems* (2022).

[19] Rennie, Steven J., Vaibhava Goel, and Samuel Thomas. "Annealed dropout training of deep networks." In *2014 IEEE Spoken Language Technology Workshop (SLT)*, pp. 159-164. IEEE, 2014.

[20] Jais, Imran Khan Mohd, Amelia Ritahani Ismail, and Syed Qamrun Nisa. "Adam optimization algorithm for wide and deep neural network." *Knowledge Engineering and Data Science* 2, no. 1 (2019): 41-46.

[21] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining", *Proc. 15th ACM Int. Conf. Knowl. Discovery Data Mining*, pp. 947-956, 2009.

[22] Wang, Lei, and Piotr Koniusz. "Uncertainty-DTW for time series and sequences." In *European Conference on Computer Vision*, pp. 176-195. Cham: Springer Nature Switzerland, 2022.

[23] Bagnall, Anthony, Jason Lines, Jon Hills, and Aaron Bostrom. "Time-series classification with COTE: the collective of transformation-based ensembles." *IEEE Transactions on Knowledge and Data Engineering* 27, no. 9 (2015): 2522-2535.

[24] Wen, Qingsong, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. "Transformers in time series: A survey." *arXiv preprint arXiv:2202.07125* (2022).