Using LSTM for Detection of Wrist Related Disorders

Miroslav Jaščur
Department of Cybernetics and Artificial Intelligence
Technical university of Kosice,
Kosice, Slovakia
miroslav.jascur@tuke.sk

Norbert Ferenčík
Department of Cybernetics and Artificial Intelligence
Technical university of Kosice,
Kosice, Slovakia
norbert.ferencik@tuke.sk

Marek Bundzel
Department of Cybernetics and Artificial Intelligence
Technical university of Kosice,
Kosice, Slovakia
marek.bundzel@tuke.sk

Iveta Zolotová
Department of Cybernetics and Artificial Intelligence
Technical university of Kosice,
Kosice, Slovakia
iveta.zolotova@tuke.sk

***Abstract*— We have build a device named RepaiR for strength measurements and isokinetic rehabilitation of the wrist joint. We have performed series of measurements on 25 healthy individuals and 10 patients with neuromuscular and traumatic impairments. Our initial goal was to verify that the measured data contain sufficient information to distinguish between healthy and not healthy subjects as a proof of concept. We have implemented Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) that processes the time structured measurements. LSTM effectively models the varying length of the input vector and the long term dependencies. We compare performances of our models on the data sets with varying minimal input vector lengths. We have proven that the measurements can be used to detect neuromuscular impairments and our best performing model worked with 77,6 % accuracy.**

I. Introduction

People suffering from movement disorders usually undergo some form of physical rehabilitation to improve motor functions of the impaired limbs. The individuals whose upper extremities are affected are particularly limited in the activities of daily life. The outcome of the physical rehabilitation process depends on the patientś health condition, his/her motivation, duration, intensity and selection of the rehabilitation activities, [1]. Repeating exercises many times is necessary to achieve improvement but this also leads to boredom, disengagement, exhaustion, frustration both on the side of the patient and the therapist especially if the rate of recovery is very slow, [2].

Physical rehabilitation can be aided and supervised by robotic devices. Robots reduce the strain on the therapist who can then tend to more patients. Patients' condition and progress can be measured more precisely and, in more detail, using the sensors of a robot than with the conventional methods, [3]. Physiological exercise data is collected by sensors attached both to the human and to the robot. The collected data are processed to extract information related to the diagnosis of the exercising subject and can be used as an input to a decision support system that advises on the ways the physical rehabilitation process is conducted. Precise diagnosis and quantification of the extent of the patient's impairment enable to suggest the optimal structure of the exercises to accelerate the recovery. The patient's condition changes over time and the challenge is to evaluate the patient's state often and unobtrusively.

We have developed a tensometric device RepaiR shown in Fig. 1 that we use for measurements and isokinetic rehabilitation of the wrist. RepaiR measures forces in flexion, extension, ulnar and radial deviation of the wrist. Resistance exercises proved to be effective for rehabilitation of both, the lower and the upper extremities, [4], [5].

Fig. 1. Repair device for wrist diagnostics

Today, the standard procedures for measuring the wrist strength are performed with objects of known weight and special exercises that enable the patient's performance to be ranked, [6]. The disadvantages of this approach are low repeatably, loss of much of the available data and subjectivity on the side of the supervising person. RepaiR's sampling frequency is in the range of 100ms and the data is evaluated by a machine learned model.

We have performed series of strength measurements on 25 healthy and 10 wrist impaired individuals. Each measurement was divided into small portions representing isokinetic strength of a basic movement and labeled with the verified diagnosis. We have used the labeled data to train and to test the machine learned models. The absolute values of the patient's measured forces provide only a limited insight into his/her condition. The time structure of the data is more important. We have used the LSTM models to classify the motion patterns and proved that RepaiR device records useful data to be used for detection and quantification of wrist related disorders.

II. Related Work

A. Rehabilitation robotics

The attempts to use robots in physical rehabilitation have a long history. Robots conserve human resources on the side of the care personnel. Robots do not get bored, frustrated and may work with higher accuracy than humans. MIT-MANUS was a pioneering rehabilitation robot for upper limbs introduced in 1992, [7]. Robot assisted therapy has become widely accepted by clinicians and researchers in the following decades. Various studies have proved that robot assisted physical rehabilitation therapy is effective although usually not more than a conventional therapy, [3] [8] [9].

NU-Wrist was a novel exoskeleton robot for wrist and forearm rehabilitation, [10]. Its authors had the same reasons for its development as we did with the RepaiR system. Currently, most of the rehabilitation robotic systems are very expensive and available only in specialized rehabilitation centers in limited quantities. NU-Wrist was designed as an table mounted exoskeleton containing three joints corresponding to the three DOF of the wrist:

1) Flexion and extension.
2) Ulnar and radial deviation.
3) Pronation and supination.

The prototype design was done in 3D CAD software allowing direct prototyping with 3-D printing technology. The robot was driven by electric motors that are widely available in various sizes and torque capabilities thus suiting most of the robotic applications including the robot-assisted wrist rehabilitation. The robot's range of motion was set on healthy subjects in isolated movements when angular deviations of the actuators were measured while the subject exercised within the maximum ranges of his/her motions, [10].

Besides the wrist motion ranges also wrist strength may be impaired e.g. by epicondylitis, soft tissue damage or fractures of the distal forearm or carpus. Yoshii et al. developed a similar device to our RepaiR for the measurement of flexural and extension torques in different forearm positions, [11]. They examine the differences between the wrist bend and elongation torque in various forearm postures in healthy subjects. The wrist torque measuring device consisted a force sensor handle stick similar to the one used in RepaiR and a table top rig for positioning of the forearm. The handle stick inclination could also be adjusted. The arm of the subject was attached to the table, unlike in RepaiR where the arm is placed in a splint. The subjects were asked to develop maximum isometric contraction for flexion or wrist extension and hold it for 5 seconds.

The patients with the neural system affected form a separate category. Miller et al., [12], studied chronic hemiparetic stroke patients using a wrist and finger force sensor module. Their device was designed to measure isometric flexion/extension forces using strain gauges on fingers, wrist, and thumb during robot-mediated 3-D dynamic movements of the upper limb. They have collected data from eight subjects with chronic hemiparetic stroke. The range of wrist and fingers force measurement was from 0 to app. 450 N what is enough to measure the maximum effort in paresis, non-paresis and healthy control subjects without the loss of resolution. This range was adapted for hand flexion and extension.

B. Machine learning for medical data

Hammerla et al. have used deep learning to assess the state of Parkinson disease, [13]. They collected input data by triaxial accelerometer that measures acceleration along three perpendicular axes with high temporal resolution (100 Hz). They used generative model based on Restricted Bolztman machines with fine-tuned softmax layer for detection of four classes: asleep, off, on and dyskinetic. Their model outperformed other approaches, despite unreliable labeling of the input data.

Stanlescu et al., [14], developed Hierarchical Linear Dynamical System for detection of sepsis in neonates. The authors monitored the cardiovascular system (heart-rate), thermoregulatory system (body temperature) and respiration system (saturation of oxygen) at 1 Hz sampling frequency and classified the data into three regimes: stability, known factors and unknown factors. The study did not find the Hierarchical Linear Dynamical System to perform better than Auto-regressive Hidden Markov Model.

Recurrent neural networks (RNN) have been successfully used on medical data. Lipton et al. [15] developed RNN with LSTM to classify 128 diagnoses using 13 clinical measurements. Multivariate input series contained: diastolic and systolic blood pressure, peripheral capillary refill rate, end-tidal CO_2, fraction of inspired O_2, Glascow coma scale, blood glucose, heart rate, pH, respiratory rate, blood oxygen saturation, body temperature, and urine output. Their best performing model was using two layers of LSTM, dropout layer with probability 0.5, fully connected layer and cross-entropy layer for multiclass classification. The model outperformed MLP with hand engineered features. This study gives strong evidence, that LSTM works well with different types of medical data.

III. The Methods Used

A. Long Short-Term Memory

A recurrent neural network is a class of artificial neural network used for classification and regression of sequential information. Information persists in the network through recurrent connections that pass information from previous time step to the current one. The most commonly used type of RNN is Long Short-Term Memory (LSTM), that was introduced by Hochreiter and Schimdhuber in 1997, [16]. We decided to use RNN with vanilla LSTM. A large-scale study concludes that it performs reasonably well on various datasets, [17].

LSTM is designed to store long-term dependencies in time series data. Long term dependencies are stored in cell states. The cell state C_t (fig. 2) passes through LSTM with simple linear modifications. LSTM is capable of modifying cell state.These modifications are performed by gates, non-linear functions, that control the amount of information that is passed further or forgotten. While h_t represents current output of LSTM, C_0 and h_0 represent the initial state of LSTM. Information in C_t and h_t is computed in every time-step t of the input time series.

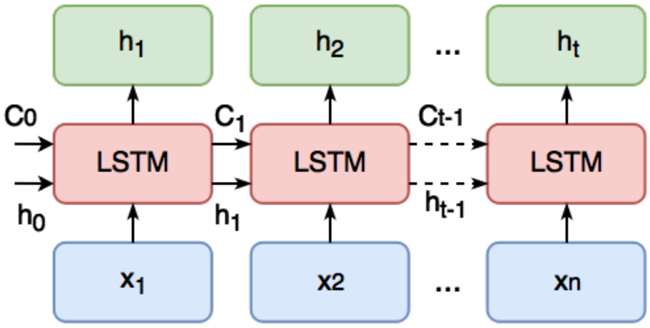

Fig. 2. Flow of memory and output in Long Short-Term Memory

LSTM unit takes three inputs x_t, h_{t-1}, C_{t-1} and computes two outputs C_t, h_t. The inputs are processed through gates. Cell state C_t is modified by two operations; element-wise multiplication and summation. Element-wise multiplication is performed by forget gate f_t (Eq. 1). Here f_t is a sigmoid function, that controls which information should be forgotten.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{1}$$

Inputs x_t and h_{t-1} are also used by the input gate i_t and vector of proposed cell state values C_t^*. The input gate i_t (Eq. 2) is again represented by a sigmoid function that controls the amount of information that will influence the new cell state C_t. The input gate uses multiplication to filter the proposed cell state C_t^* (Eq. 3), which is computed as hyperbolic tangent function of the new information in time step t.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{2}$$

$$C_t^* = \tanh\left(W_c.[h_{t-1}, x_t] + b_c\right) \tag{3}$$

New cell state C_t (Eq. 4) is then calculated as a sum of old cell state $C_t - 1$ filtered by forgot gate f_t and proposed cell state C_t^* filtered by input gate i_t.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t^* \tag{4}$$

New output h_t (Eq. 6) is computed as cell state C_t filtered by the output gate o_t (Eq. 5). The output gate o_t (Eq. 5) represents how much should the input x_t the previous output h_{t-1} and the current cell state influence the current output of the LSTM unit.

$$o_t = \sigma(W_o.[h_{t-1}], x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

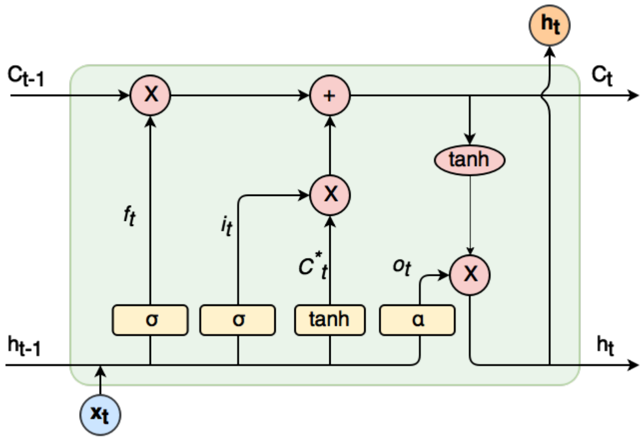

Fig. 3. Long Short-Term Memory structure

We have used a RNN architecture that contained an input layer, an LSTM layer a fully connected layer and a softmax layer. Softmax was used for multiclass classification and the loss was computed by a cross-entropy cost function.

B. Experimental Setup

The goal of our experiments was to verify that the RepaiR device can collect data that enable detection of wrist related problems. We have examined 25 healthy individuals and 10 individuals with neuromuscular impairments. Participants varied in age (20 - 68), sex (13 females, 22 males). We have labeled the data based on the patient's diagnose that was determined by medical experts prior to the experiments. Experiments were performed under supervision of physicians at The Department of physiotherapy, balneotherapy and clinical rehabilitation in Louis Pasteur University Hospital. All of the experiments were performed with formal approval of the human subjects.

We have asked the subjects to consecutively perform flexion, extension, radial deviation and ulnar deviation and to hold exert the force for 2 seconds at their maximum. We did not provide the subject any feedback on their performance. This allowed us not only to measure the basic movements' maximal forces but also to observe the changes in the forces through time, Fig. 4.

Fig. 4. Exercising on Repair device.

The raw input data are measured by eight strain gauges in the RepaiR device. The sensors are located inside the handle stick, 4 in the top and 4 in the bottom placed on the sides of 2 squares, forming two parallel cartesian planes. The sampling frequency was 10 Hz. The raw input data is thus a time series of 8 dimensional vectors.

We have divided the data into subsets based on the type of the attempted movement (flexion, extension etc.). We have implemented a rule-based system to classify the input data this way based on the responses of the individual strain gauges. The subsets served as inputs to the machine learned models (Eq. 7).

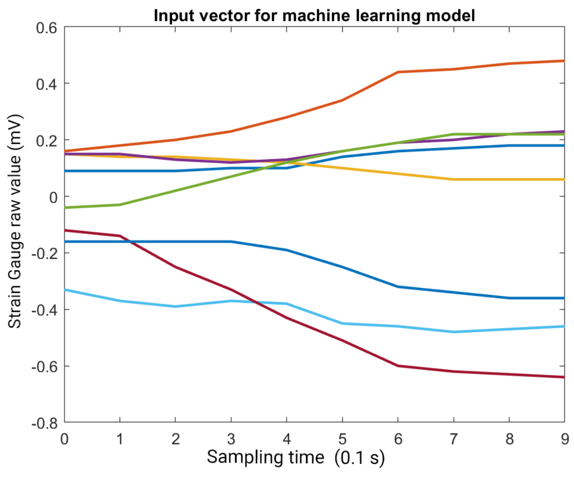

Fig. 5. Input vector of raw data used for classification

Each of the sequential input maps represented a single attempt of a subject to exert maximal force in a specific movement, Fig. 5. We have evaluated the performance of two LSTM models. Both models were designed as dichotomous classifiers (healthy or impaired) detecting neuromuscular impairments. We have varied the minimal size T_{min} of the input sequence to find the optimal setup and to answer the question if increasing the length of the input sequence improves the accuracy of the trained model.

$$\begin{gathered} x_i = \{f_1, f_2 \ldots f_8\} \\ X_{input} = \{x_1 \ldots x_t\} \\ where\ t\ is\ number\ of\ measurements \\ t > T_{min} \end{gathered} \tag{7}$$

IV. Experimental Results

We have produced 5 different datasets from the raw data increasing the minimal value of T_{min} from 1 to 5. The size of the datasets varied from 1024 observations at T_{min}=1 to 280 observations at T_{min}=5. We have used 60 % of the data for training and 40 % for validation. Each class was represented by equal proportion of samples - 50%.

The RNN models were trained using stochastic gradient descent with fixed learning rate α=0.1 for 250 epochs. Each training procedure was repeated 10 times and the results obtained are reported in Tab. I. The values displayed represent averaged performances over the 10 repetitions. The top of Tab. I shows the performance of the RNN with a single LSTM layer with 50 units and the bottom of the RNN with 100 LSTM units and a dropout layer with 0.5 dropout probability.

The individual trained models performed with 0.7 ± 0.1 accuracy. The accuracy does not vary significantly neither between the two models nor across the different training sets. slightly improving with the increasing length of the input sequence. The most accurate model trained on the dataset with $T_{min} = 4$ has the average accuracy of 0.7671.

It is also important to evaluate the sensitivity and the specificity. The first model has sensitivity of 0.55 or lower. The second model performs similarly on the first three datasets. However, there is a substantial increase in sensitivity to 0.7181 ($T_{min} = 4$) and 0.6275 ($T_{min} = 5$) on the two latter datasets. The specificity of the models remains high on the datasets; i.e. 0.8. The results show that the first model without the dropout layer has the tendency to classify majority of the input patterns in the negative class what leads to high specificity, but low sensitivity.

V. Discussion

We have performed a study that aim was to verify the diagnostic capability of the RepaiR device. We have developed a computer game to test the rehabilitation potential of the RepaiR device but that is a work in progress. We have performed the experiments in the Louis Pasteur University Hospital in Kosice, Slovakia. We acknowledge here their kind support. We have found several design issues during the tests. The subject needs to grasp the handle. This is problematic when the subject suffers on partial paralysis of the hand. Sometimes the subject's physical proportions do not match the splint or the handle although the RepaiR device is partially adjustable. The handle should be free rotating to avoid twisting the handle that affects the input data negatively. We have observed that if the subject is provided with the visual feedback he/she exerts higher forces. Overmotivated subjects tend also to twist the handle at higher

Average classification performance					
LSTM - 1 layer - 50 units					
	T_{min}=1 (n = 1024)	T_{min}=2 (n = 726)	T_{min}=3 (n = 524)	T_{min}=4 (n = 380)	T_{min}=5 (n = 280)
Accuracy	0.7090	0.6897	0.6876	0.7020	0.7107
Sensitivity	0.5043	0.4248	0.4170	0.5170	0.5495
Specificity	0.9188	0.9470	0.9535	0.8899	0.8690
PosPredVal	0.8642	0.8861	0.8981	0.8267	0.8047
NegPredVal	0.6440	0.6289	0.6246	0.6446	0.6626
PosLikHood	6.2121	8.0108	8.9695	4.6963	4.1959
NegLikHood	0.5394	0.6074	0.6114	0.5428	0.5183
LSTM - 1 layer - 100 units with dropout (P=0.5)					
Accuracy	0.7044	0.7138	0.6957	**0.7671**	0.7357
Sensitivity	0.5611	0.4598	0.4908	**0.7181**	**0.6275**
Specificity	0.8488	0.8843	0.8959	0.8151	0.8413
PosPredVal	0.7890	0.7948	0.8217	0.7918	0.7941
NegPredVal	0.6574	0.6269	0.6429	0.7470	0.6984
PosLikHood	3.7111	3.9734	**4.7162**	3.8837	3.9532
NegLikHood	0.5170	0.4934	0.5683	0.3459	0.4428

TABLE I

AVERAGE CLASSIFICATION PERFORMANCES OF TWO RNN WITH LSTM ON THE DATASETS WITH VARYING INPUT LENGTHS.

forces what is the most prominent in flexion. The subject tends to move the forearm towards the handle to increase the force. One of the solutions is to use an auto adjustable splint. The ergonomy of the device should also be improved. It is placed on the top of a table and the handle was difficult to reach for wheelchair patients. The device should be mounted on a dedicated height adjustable mount.

The strain gauges measure opposite forces and work in couples placed opposite to each other. The strain gauges are flexible and produce counter forces when deformed that affect the measurement. The measured time series are generally useful only when the measured force increases. The time series with the measured forces decreasing usually correspond to the return to zero values due to the physical characteristics of the device rather than due to the subject's input.

There are several practical issues that might affect model's performance, accuracy and sensitivity. The data used for experiments contained the measurements of healthy individuals and patients with impairments caused by various injuries and diseases such as forearm trauma, stroke or multiple sclerosis. In most cases we only had one or two patients within each group and we have measured maximal strengths of both wrists in four motions. However, not all of these conditions influence the strength. These measurements might pollute the data by providing contradicting samples.

In terms of variability of the conditions a significant proportion of the data was measured on stroke patients. This lack of diversity in training samples might lead to bias towards features specific for stroke patients and influence the model's accuracy for patients with different conditions. Another factor that is likely to affect the model robustness is the relatively small number of the study participants (35).

The optimization method we have used was the stochastic gradient descent that is trained in mini-batches. This training combined with the fixed learning rate causes fluctuations in the training progress (Fig. 6) that affect the performance of the classifier. Using dynamic learning rate caused over fitting on the subjects with the low wrist strength. This can be solved by using a more heterogeneous dataset.

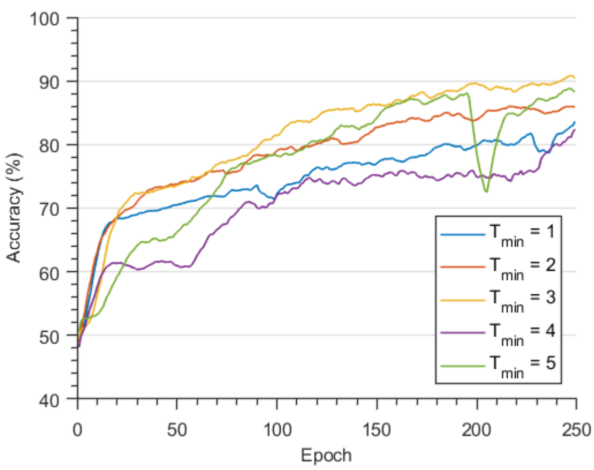

Fig. 6. LSTM training progress - 250 epochs

VI. CONCLUSION

We have built the RepaiR device to measure forces exerted by the wrist attempted movements and to aid in the physical rehabilitation process. Our goal is to make the rehabilitation process adaptive. For that the subject's condition has to be evaluated often. We have presented here the initial research of the capability of the RepaiR device to detect wrist related problems. We have collected experimental data from healthy individuals and patients with various neuromuscular impairments. We have implemented a recurrent neural network with LSTM layer that was able to distinguish between the healthy and the affected subjects. The best performing model had 77.6 % accuracy with 71.8 % sensitivity.

The results obtained show, that recurrent neural network with LSTM is able to classify continuous data from RepaiR.

The RepaiR collected data contain more information that can be used to focus and specify the rehabilitation process.

Several suggestions arise from the results of this study some of them related to improving the ergonomy of the device. More experimental data need to be collected to distinguish between different diagnoses and the extent of the problem. So far, we have found that certain fluctuations in the measured force indicate that the neural system is affected.

VII. Acknowledgment

This work was supported by the Scientific Grant Agency VEGA under Grant VEGA No. 1/0663/17; Tatra bank foundation under Grant E-Talent No. 2016et019; Slovak Research and Development Agency under Grant No. APVV-015-0731.

References

[1] Hilde Feys, Willy De Weerdt, Geert Verbeke, Gail Cox Steck, Chris Capiau, Carlotte Kiekens, Eddy Dejaeger, Gustaaf Van Hoydonck, Guido Vermeersch, and Patrick Cras. Early and repetitive stimulation of the arm can substantially improve the long-term outcome after stroke: a 5-year follow-up study of a randomized trial. *Stroke*, 35(4):924–929, 2004.

[2] Julie Luker, Elizabeth Lynch, Susanne Bernhardsson, Leanne Bennett, and Julie Bernhardt. Stroke survivors' experiences of physical rehabilitation: a systematic review of qualitative studies. *Archives of physical medicine and rehabilitation*, 96(9):1698–1708, 2015.

[3] GB Prange, MJA Jannink, CGM Groothuis-Oudshoorn, HJ Hermens, and MJ Ijzerman. Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke. *Journal of rehabilitation research and development*, 43(2):171–184, 2009.

[4] S Robineau, B Nicolas, P Gallien, S Petrilli, A Durufle, G Edan, and P Rochcongar. Eccentric isokinetic strengthening in hamstrings of patients with multiple sclerosis. In *Annales de readaptation et de medecine physique: revue scientifique de la Societe francaise de reeducation fonctionnelle de readaptation et de medecine physique*, volume 48, pages 29–33, 2005.

[5] A Escalante, L Miller, and TD Beardmore. Resistive exercise in the rehabilitation of polymyositis/dermatomyositis. *The Journal of rheumatology*, 20(8):1340–1344, 1993.

[6] Lee D Walsh, Uwe Proske, Trevor J Allen, and Simon C Gandevia. The contribution of motor commands to position sense differs between elbow and wrist. *The Journal of physiology*, 591(23):6103–6114, 2013.

[7] Neville Hogan, Hermano Igo Krebs, J Charnnarong, P Srikrishna, and Andre Sharon. Mit-manus: a workstation for manual therapy and training. i. In *Robot and Human Communication, 1992. Proceedings., IEEE International Workshop on*, pages 161–165. IEEE, 1992.

[8] Gert Kwakkel, Boudewijn J Kollen, and Hermano I Krebs. Effects of robot-assisted therapy on upper limb recovery after stroke: a systematic review. *Neurorehabilitation and neural repair*, 22(2):111–121, 2008.

[9] Peter S Lum, Charles G Burgar, Peggy C Shor, Matra Majmundar, and Machiel Van der Loos. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. *Archives of physical medicine and rehabilitation*, 83(7):952–959, 2002.

[10] Nurdos Omarkulov, Kuat Telegenov, Maralbek Zeinullin, and Iliyas Tursynbek. Preliminary mechanical design of nu-wrist: a 3-dof self-aligning wrist rehabilitation robot. *International Conference on Biomedical Robotics and Biomechatronics*, 6(1):962–967, 2016.

[11] Yuichi Yoshii, Hiroshi Yuine, Ohashi Kazuki, Wen-lin Tung, and Tomoo Ishii. Measurement of wrist flexion and extension torques in different forearm positions. *BioMedical Engineering OnLine*, 14(1):115–118, 2015.

[12] Laura C Miller, Ricardo Ruiz, Torres Arno, H. A. Stienen, and P. A. Dewald. A wrist and finger force sensor module for use during movements of the upper limb in chronic hemiparetic stroke. *IEEE transactions on bio-medical engineering*, 56(9):2312–2317, 2009.

[13] Nils Yannick Hammerla, James Fisher, Peter Andras, Lynn Rochester, Richard Walker, and Thomas Plötz. Pd disease state assessment in naturalistic environments using deep learning. In *AAAI*, pages 1742–1748, 2015.

[14] Ioan Stanculescu, Christopher K. I. Williams, and Yvonne Freer. *A Hierarchical Switching Linear Dynamical System Applied to the Detection of Sepsis in Neonatal Condition Monitoring*. 2014.

[15] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.

[16] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[17] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2017.