UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Machine-learning-based biomarkers towards customization of robotic rehabilitation treatments for stroke patients

Caio Benatti Moretti

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Caio Benatti Moretti

Machine-learning-based biomarkers towards customization of robotic rehabilitation treatments for stroke patients

Doctoral dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. *EXAMINATION BOARD PRESENTATION COPY*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. Alexandre Cláudio Botazzo Delbem

USP – São Carlos
February 2021

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M844m   Moretti, Caio Benatti
Machine-learning-based biomarkers towards
customization of robotic rehabilitation treatments
for stroke patients / Caio Benatti Moretti;
orientador Alexandre Cláudio Botazzo Delbem. --
São Carlos, 2021.
176 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2021.

1. Machine learning. 2. Biomarkers. 3.
Rehabilitation robotics. 4. Stroke. I. Delbem,
Alexandre Cláudio Botazzo , orient. II. Título.

Caio Benatti Moretti

**Biomarcadores baseados em aprendizado de máquina para
customização de tratamentos de reabilitação robótica para
pacientes com AVC**

Tese apresentada ao Instituto de Ciências
Matemáticas e de Computação – ICMC-USP,
como parte dos requisitos para obtenção do título
de Doutor em Ciências – Ciências de Computação
e Matemática Computacional. *EXEMPLAR DE
DEFESA*

Área de Concentração: Ciências de Computação e
Matemática Computacional

Orientador: Prof. Dr. Alexandre Cláudio
Botazzo Delbem

**USP – São Carlos
Fevereiro de 2021**

*To Denise, one of the pillars of
my inspiration and character.
Friend. Aunt. Woman.*

ACKNOWLEDGEMENTS

As I write this, I feel that it is not only being on the verge of achieving a big milestone that triggers a plenitude of joy inside me. It is especially about the wonderful people that were part of the outstanding experiences I had along this journey and their huge contributions. People I do owe a debt of gratitude big time.

The taste of every struggle towards this current moment, since when I started applying to my Master's, reminds me of the intense moments when my family were there for me. My big brother Ralph, my mom Ana and aunt and uncle Denise and Wagner were my heroes that always helped me through every burden of a fight. More importantly, they were also the first to know when it was time to celebrate every achievement that was worth fighting for. Thank you!

My initiation into the academia world could not have been possible if it wasn't for Professors Dr. Kleber Andrade, Dr. Pedro Grosso and Dr. José Martins, who kindly prepared me for Master's. This takes me to Professor Dr. Glauco Caurin, who introduced me to the field of research this thesis arises from and took me even deeper into the scientific labor, getting me ready for pursuing a PhD degree. Thank you!

This hectic routine from the last five years as a PhD student has been an interesting practice for my future endeavors. I am very grateful for the kind willingness of Professor Dr. Alexandre Delbem, not only for supervising and contributing to my research project, but also for the huge help on every problem or request that was outside my purview. I also extend this gratitude to the professors from the courses I took, which were paramount for the amount of knowledge I have acquired and will be of big use in the future. Thank you!

This journey was also made of some (not to say lots) of fun moments. I made a lot of good friends for life, especially in GAPeria CAASO samba band from the campus at São Carlos, and my roommates Lemuel, Erick and Oliver, Tiago and Giba. You made me a very happy person with lots of awesome memories. Thank you!

I had the opportunity to step my feet on foreign soil for the first time during my period at MIT. My dear friend Adriana picked me up at the airport and gave me the first coordinates on how to live abroad. This golden opportunity was possible thanks to Professor Dr. Hermano Krebs, who kindly invited me to conduct research at The 77 Laboratory, which culminated in huge gains for this thesis. Despite the ton of work to do, the routine was very fun with my awesome labmates Satoshi, Moises, Andres, Ronnappee,

Takahiro, Taya, Ben, Xu, Lvyan and Kazuo. Not to mention the golden moments I could visit laboratories, watch seminars and meet those people that have their names written in the books that are the pillars of my education.

The Visiting Students Association (MIT VISTA) also brought me the opportunity to make more awesome friends for life. Lots of good memories from the talent show, halloween, thanksgiving, christmas and all of the amazing events we had. Even the hunts for free food around the campus (especially the golden mine known as E40) warm up my heart everytime I remember. So thanks Taka, Ari, Jorge, Lisa, Claudio, Lorenzo, Yori, Lorenz, Sofia, Katie, Richard, Simone, Niki, Kata, Cyril, Michael, Eleonora, Marcella, Jurgis and everyone from VISTA (lots of names, sorry) for such amazing moments. I also had the best roommates Anitah, Duncan and Bissrat. With them, my dear friend Alicia and some VISTA friends I had the opportunity to experience different cultures food wise, as well as share a bit of mine with some brazilian barbecue and caipirinha. Thank you!

This research was only possible because of the brave stroke patients from Lucy Montoro institute, Burke Neurological Institute and Feinstein Institute for Medical Research, who never gave up and had the discipline to pursue their recovery, as well as the kindness to consent in contributing to research. Thank you!

I also thank São Paulo Research Foundation (FAPESP, grant numbers 18/26493-7 and 19/06551-5), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, finance code 001), MISTI MIT-Brazil (grant number 2160632) and NICHD of the NIH (award number R01HD069776) for the financial support.

Caio Benatti Moretti

Piracicaba, January 16, 2021.

“Everything should be made as simple as possible, but not simpler.”
(Albert Einstein)

RESUMO

MORETTI, C. B. **Biomarcadores baseados em aprendizado de máquina para customização de tratamentos de reabilitação robótica para pacientes com AVC.** 2021. 174 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Acoplados a sensores, robôs para reabilitação de AVC descrevem o comportamento motor de pacientes como grandezas cinemáticas e dinâmicas, pouco exploradas no contexto de ciência de dados, devido à custosa tarefa de obter um volume significativo de dados. Além disso, a definição de biomarcadores para uma avaliação mais confiável da evolução do paciente permanece um problema aberto na literatura. Quatro estudos diferentes foram conduzidos com o objetivo de abordar tal questão. É proposta também uma ferramenta modular para organizar programas para análise de dados, com cálculos de mais de vinte métricas implementadas. O primeiro estudo consiste em um método puramente baseado em dados para definir uma região no espaço de dados, aludindo a um estado de reabilitação, baseado na incerteza da classificação de lados hemiparéticos de pacientes com AVC crônico. Nosso segundo estudo levantou evidências de que a tDCS anódica pode ter uma interação desvantajosa com o hemisfério afetado em pacientes com déficits muito severos dos membros superiores. O terceiro estudo correlaciona as métricas implementadas com escalas clínicas tradicionais, de forma que os modelos de aprendizado de máquina treinados possam desempenhar o mesmo papel de forma quantitativa e determinística, eliminando a natureza subjetiva dos métodos tradicionais de avaliação. Descobrimos que utilizar o modelo treinado com dados de apenas de um tipo de robô (ombro/cotovelo ou pulso) para estimar escalas clínicas é tão eficiente quanto combinar dados de ambos. Encontramos no quarto estudo evidências, sob uma perspectiva clínica, de um potencial de previsão de resultados clínicos de pacientes em estágios iniciais do tratamento. Nossos resultados indicam a possibilidade de melhorar o processo de tomada de decisão alertando, ao final da segunda sessão, se o paciente potencialmente não apresentará uma resposta significativa à terapia. Os projetos aqui descritos avançam o estado da arte no desenvolvimento de biomarcadores para avaliar e acompanhar o progresso do paciente em terapia robótica; propõem padrões a fim de simplificar o compartilhamento de dados; simplificam os estudos clínicos com amparo estatístico, bem como a ferramenta proposta, e otimizam a tomada de decisão clínica, impactando o orçamento em tratamentos de reabilitação e otimizando recursos do paciente para uma melhor qualidade de vida.

Palavras-chave: Aprendizado de Máquina, Biomarcadores, Robótica de Reabilitação, Acidente Vascular Cerebral.

ABSTRACT

MORETTI, C. B. **Machine-learning-based biomarkers towards customization of robotic rehabilitation treatments for stroke patients**. 2021. 174 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2021.

Coupled with sensors, robotic devices for Stroke rehabilitation describe the motor behavior of patients as kinematic and kinetic data, which are underexplored in the data science and machine learning context, due to the time-consuming task of pursuing enough data volume. Moreover, the definition of biomarkers for a quantitative and deterministic assessment of patient progress remains an open problem in the literature. Four different studies were carried out aiming to address such an issue. We also propose a modular framework for organizing pieces of software for data analysis with the calculations of more than twenty robot-based metrics implemented. Our first study concerns a clinically-naïve method for defining a region in the data space, alluding to a state of rehabilitation, based on the uncertainty in the classification of hemiparetic sides of chronic stroke patients. Our second study raised evidences that anodal tDCS may have a detrimental or maladaptive interaction with the affected hemisphere in patients with very severe upper-extremity impairments. Our third study correlates the implemented robot-based metrics with traditional clinical scales, so trained machine learning models can play the same role on a quantitative and deterministic way, eliminating the subjective nature from traditional evaluation methods. We found that having a model to estimate clinical scales only from one kind of robot (shoulder/elbow or wrist) is as good as combining data from both. We found in our fourth study evidences from a clinical perspective on the prediction of clinical outcomes of patients at early stages of the treatment. Our results indicate the possibility of improving the decision-making process by alerting, at the end of the second therapy session, when patients will potentially not present a significant response. The four projects here described enabled to push the state of the art towards the development of biomarkers to evaluate and track patient's progress on rehabilitation robotics treatments; proposing standards for simplifying the data sharing; simplifying clinical studies with traditional statistical tools and our framework API and optimize the clinician's decision-making process towards impacting patient's budget on rehabilitation treatments for optimizing quality of life.

Keywords: Machine learning, Biomarkers, Rehabilitation Robotics, Stroke.

LIST OF FIGURES

Figure 1 – Data flow of the Fayyad et al. KDD model.	36
Figure 2 – Effort spent in KDD steps.	39
Figure 3 – Examples of data arrangements susceptible to projections in lower dimensions	40
Figure 4 – Usage of k principal components, according to their cummulative explained variance of the original data. (a) Depicts the cummulative explained variance as a function of the first k principal components and (b) defines the ratio of the function in the graph	41
Figure 5 – Comparison of projections for supervised learning.	42
Figure 6 – Hypothetical space $\mathcal{F}_{\text{all}}$, encompassing $\mathcal{F}$. (a) Bias-variance tradeoff represented by $\mathcal{F}$; (b) Consistencies of a classification algorithm; (c) Estimation and approximation errors.	46
Figure 7 – Hypothetical space depicting interpretations of estimation and approximation errors.	47
Figure 8 – Possible classifications in $\mathbb{R}^2$ from an algorithm of linear bias and a sample of size 3.	51
Figure 9 – Cases in which a linear bias fails to shatter a sample of size 4 in $\mathbb{R}^2$ . . .	51
Figure 10 – Illustrations of some robotic devices from Table 1.	57
Figure 11 – JSON structure of preprocessed data. Left and right sides of nodes are respectively the field name and data type.	67
Figure 12 – Directory tree of MIT MANUS device	68
Figure 13 – UML Sequence diagram of the execution of tasks, modes and metrics dynamically added to the framework for a single subject. Dashed rectangles are placeholders for programs to be dynamically located according to the naming format with words in gray.	68
Figure 14 – Example of development and implementation processes of metrics in the framework. After developing a metric that does calculations from raw data (a), it must be allocated within a task and such task must be appended in the task list of the respective device (b).	71
Figure 15 – Progress of Patients A (recovering) and B (not recovering) described by Deviation and Aim robot metrics along evaluation sessions.	72

Figure 16 – The virtual environment (Clock game) used by patients for rehabilitation exercises. On the left, the yellow circle indicates the end-effector position to be moved towards the current target (red circle). On the right, the flow of a cycle of 16 movements. All the targets (black circles) are 14cm away from the center.	76
Figure 17 – Aim metric rationale. On the left, aim is the angle difference θ between the true path (black solid line) towards NE target and the point of when the patient starts to take action along the movement trajectory (blue dotted line). The X mark indicates such moment both in the speed profile (right) and in the position space (left).	79
Figure 18 – Duration metric rationale. On the left, the blue dotted line traces the patient's true movement trajectory (with the respective blue part on the right side) and the black solid line indicates the straight path towards NE target. On the right, the gray part of the speed profile indicates idle state, whereas the blue part (exceeding 10% of peak speed) indicates true action on performing the movement and its time length as the metric itself.	80
Figure 19 – Deviation metric rationale. On the left, the blue dotted line traces the patient's movement trajectory and the black solid line indicates the straight path towards NE target in the form $-x + y = 0$. The squared area, magnified on the right side, indicates the distances between each point and the straight path, aggregated in Equation 5.1.	81
Figure 20 – Dwell time metric rationale. On the left, the blue dotted line traces the patient's behavior during idle state. Respectively, on the right, the blue part of the speed profile indicates the idle state. Indicating 10% of peak speed, the X mark indicates when idle state ends, so that dwell time is defined by its time length.	81
Figure 21 – Mean speed and peak speed rationale. Peak speed is denoted by the star mark and the mean speed is denoted by the dashed line.	82
Figure 22 – Rationale for major axis and minor axis metrics. The blue dotted line traces patient's movement trajectory (robot data) for ccw3 task. The black ellipse is the ellipse fitted to the robot data and dashed and dotted lines are, respectively, major and minor axes of the fitted ellipse.	85

Figure 23 – Joint independence rationale. (a) Arm model (top view) of the subject (gray) performing a cw9 task, where l_1 and l_2 are upper and lower arm, respectively, θ_1 and θ_2 are shoulder and elbow angles, respectively and $(x, y) = (X - d_1, Y + d_2)$. (b) Female (F) and male (M) parameters and body measurements extracted from Dipietro *et al.* (2007): d'_2 is the distance between device's table border and subject's body midline (0.114m for male, 0.107m for female), $dist$ is the distance between device's table border and the center of robot's workspace (0.23m), so that $d_2 = d'_2 + dist$; d_1 is the distance between shoulder and subject's body midline (0.175m male, 0.157m female); l'_2 is the forearm length (0.254m male, 0.234m female), wgd is the distance from the wrist to the grip line (0.076m male, 0.071m female) and rh is the robot handle radius (0.03m), so that $l_2 = l'_2 + wgd + rh$; l_1 is variable, so that by plugging the upper arm length (0.282m male, 0.264m female) into Equation 5.10, it yields a projection ranging from l'_1 (0.141m male, 0.132m female) to l''_1 (0.265m male, 0.248m female). 86

Figure 24 – Definition of l_1 . (a) upper arm angle (vertically) as a function of the distance d with respect to the closest target to the arm, defined in Equation 5.9. (b) Projection of the upper arm length as a function of d (Equation 5.10). 87

Figure 25 – Iterative process described in Algorithm 1. The green curve indicates the current fitted gaussian function $\mathbf{s}$, given sm_{peak} (star mark) and the points within r_0 and r range (left and right X marks, respectively). The solid blue curve is the speed profile $\mathbf{v}$, whereas dotted blue is $\mathbf{v} - \mathbf{s}$, i.e., dotted blue curve from the i -th iteration is the solid blue curve of the $(i + 1)$ -th iteration. The red curve indicates an invalid submovement. . . 90

Figure 26 – Submovements from the iterative process illustrated in Figure 25 in (a) the order they were extracted and (b) sorted by $sm_{peaktime}$ (line 28). . . 91

Figure 27 – Submovement overlap rationale. The time length that the 4th submovement overlaps the third is defined as $sm^{4_{overlap}}$ 92

Figure 28 – Submovement interpeak rationale. The time-offset difference between 4th and 3rd submovement peaks, defined as $sm_{4_{interpeak}}$ 93

Figure 29 – The robotic device used in physical therapy sessions. (a) InMotionArm™ robotic device for the robot-aided rehabilitation (adapted from < http://int.lynxdesign.com/wp-content/uploads/2012/11/InMotionARM240.jpg >). (b) At the left, the virtual environment (Clock game) used for rehabilitation exercises from which data were collected in this work; at the right, the flow of cycles of movements performed by patients. (c) Visualization of position data where a cycle of movements is performed by a chronic patient; black circles are the actual target positions.	97
Figure 30 – Visual inspection of subspaces where the proneness to separability is more evident. (a) $\bar{f}_x \times \sigma_{f_y}$ subspace. (b) $\gamma_{f_x} \times \gamma_{f_y}$ subspace.	100
Figure 31 – Accuracies and dimensionalities of reduced sets of features along the feature selection strategy of Algorithm 5.	105
Figure 32 – Generalization bound of the SVM classifier f , with $\delta = 0.05$ and $R_{\text{emp}}(f) = 0.19$, indicating the bound with respect to the sample from Table 4.	106
Figure 33 – Estimation of $\mathcal{N}(\mathcal{F}, 2n)$ via coefficients of a quadratic polynomial. . . .	106
Figure 34 – Upper bound of Chernoff/Hoeffding inequality as a function of the sample size. (a) shows the curve and points out the size of the available sample (Table 4). (b) shows the same curve within the range of interest. . . .	107
Figure 35 – Order of emergence of instances from a single patient (#15, from Table 4) over time in $\bar{f}_x \times \sigma_{f_y}$ subspace. Black and gray points arose respectively from devices A and B. (a) Region in $\bar{f}_x \times \sigma_{f_y}$ to be enlarged in b and c. (b) The first 15 instances. (c) Instances from the 15 th to the 32 nd	111
Figure 36 – Emergence of instances (vertical axis) over time (horizontal axis). Black and gray bars indicate instances which arose respectively on devices A and B. (a) Attendance of patient #15 in rehabilitation sessions. (b) Ideal attendance with random numbers of cycles per session.	112
Figure 37 – Outputs of the SVM model, for each instance from patient #15, over time. Blue curve is a linear model fitting the series.	113
Figure 38 – Distribution of the angular coefficients from each time series related to patient performance. The mean is indicated in blue.	113
Figure 39 – CONSORT flow diagram	118
Figure 40 – Brute-force search process by reducing and offsetting a window along the FMA scale. The patients that fall in such window, i.e., with an FMA score within the window, were divided into $Robot_{tDCS}$ and $Robot_{Sham}$ groups and submitted to the unpaired test for each metric. Then, the first quartile (Q1) is obtained, as the criterion for finding the window in which metrics are impacted the most.	119

Figure 41 – Combined $Robot_{tDCS}$ and $Robot_{Sham}$ mean and standard error of kinematic and kinetic outcome metrics for S/E device. Significant changes ($p < 0.05$) between admission (ad), discharge (dc), and follow-up (fu) are marked with an * mark.	120
Figure 42 – Combined $Robot_{tDCS}$ and $Robot_{Sham}$ mean and standard error of kinematic and kinetic outcome metrics for wrist device. Significant changes ($p < 0.05$) between admission (ad), discharge (dc), and follow-up (fu) are marked with an * mark.	121
Figure 43 – Example movement plots of p2p (S/E and wrist) and circle (S/E) tasks for Patient #42627 (FMA of 9 at admission and 16 at follow-up). . . .	124
Figure 44 – Example movement plots of p2p (S/E and wrist) and circle (S/E) tasks for Patient #42658 (FMA of 15 at admission and 25 at follow-up). . . .	124
Figure 45 – Example movement plots of p2p (S/E and wrist) and circle (S/E) tasks for Patient #42633 (FMA of 24 at admission and 34 at follow-up). . . .	125
Figure 46 – Improvements in joint independence over time. (a) and (b): individual values of joint independence metric, for all patients, sorted by their performance at admission (squares) in comparison to discharge (diamonds), and follow-up (circles). A lower number indicates greater/improving joint independence. (c) and (d): changes in the joint independence metric, where filled circles indicate significance ($p < 0.05$) at discharge (c) and follow-up (d).	127
Figure 47 – Heatmap of S/E kinematic and kinetic data correlation with clinical scales. Significant correlations ($p < 0.05$) are marked with a triangle at the upper-right corner of the respective element.	134
Figure 48 – Heatmap of wrist kinematic data correlation with clinical scales. Significant correlations ($p < 0.05$) are marked with a triangle at the upper-right corner of the respective element. The * mark refers to metrics from PS tasks, whereas the others are from 2D tasks.	135
Figure 49 – Performance of patient #42658 along therapy sessions described by deviation metric. (a) Ebbinghaus’ forgetting curve. Each therapy session is represented by a solid curve. Dashed curves are exponential functions fitted to sessions individually (shorter curves) or to all sessions (longer curve). (b) gains kept (retained) from the previous session, defined as $i - 1$ -th session’s last point minus i -th session’s first point. . .	145
Figure 50 – Dataset format with patients included in the post analysis (size $n = 76$, see Figure 39) for LSTM models. (a) Metric values at movement level are fed at each time step, for all examples. (b) The expected output: FMA score for the first strategy or Responder/non-responder label for the second strategy.	146

Figure 51 – Approval from CAPPesq ethics committee. 174

LIST OF ALGORITHMS

Algorithm 1 – Submovement decomposition algorithm. Adapted from Krebs (1997).	89
Algorithm 2 – Submovement overlap calculation	93
Algorithm 3 – Submovement interpeak calculation	94
Algorithm 4 – Fitness function	102
Algorithm 5 – Recursive population initialization for feature selection with genetic algorithms	104

LIST OF TABLES

Table 1 – Summary of the main features of the robotic devices.	58
Table 2 – Demographics from data samples of related work	60
Table 3 – Format of an overall report of statistical tests (per metric) over a patient sample. P-values of the provided statistical tests (e.g., t-test and Wilcoxon signed rank test) are reported towards comparing changes in metric values between phases like admission (adm), discharge (dc) and follow-up (fu).	72
Table 4 – Demographics from the collected data.	98
Table 5 – Grid search parameter configuration.	102
Table 6 – Significant changes from admission to discharge for the combined analysis of kinematic and kinetic data.	122
Table 7 – Post-hoc analysis results: significant changes seen in the <i>Robot_{sham}</i> group in the severe patient sub-group.	123
Table 8 – Participant demographics and admission results of clinical measures . .	136
Table 9 – Correlation between estimator models and and clinical scales. $**p < 0.001$, $*p < 0.05$	139
Table 10 – Best parameter configuration for the LSTM models (grid search results), given different (time-dependent) input data combinations.	147
Table 11 – Best parameter configuration for the MLP models (grid search results), given different (static) input data combinations.	147
Table 12 – Weights of the ensemble of MLP neural networks that constitute the nonlinear model for the FMA-UE correlation, given S/E data.	171
Table 13 – Weights of the ensemble of MLP neural networks that constitute the nonlinear model for the FMA-UE correlation, given S/E and wrist data.	172

LIST OF ABBREVIATIONS AND ACRONYMS

AA	Adduction/Abduction
ADL	Activities of Daily Living
AUC	Area Under ROC Curve
BI	Barthel Index
CE	Cross-Entropy
CONSORT	Consolidated Standards of Reporting Trials
DBMS	Database Management Systems
DOF	Degrees of Freedom
ERM	Empirical Risk Minimization principle
FE	Flexion/Extension
FMA	Fugl-Meyer Assessment
GA	Genetic Algorithms
iid	independently and identically distributed
KDD	Knowledge Discovery in Databases
LDA	Linear Discriminant Analysis
LSTM	Long Short Term Memory
MCC	Matthews Correlation Coefficient
MCID	Minimum Clinically Important Difference
MEP	Motor Evoked Potential
MLP	Multi-Layer Perceptron
MP	Medical Research Council Motor Power score
MRI	Magnetic Resonance Imaging
owr	One Way Record
p2p	Point to Point
PCA	Principal Component Analysis
PS	Pronation/Supination
pstatic	Playback static
RBF	Radial-Basis Functions
rdyn	Round dynamic
S/E	Shoulder/Elbow
SG	Savitzky-Golay

SLT	Statistical Learning Theory
SOR	state of rehabilitation
SSE	Sum of Squared Errors
SVM	Support-Vector Machines
tDCS	Transcranial Direct Current Stimulation
UE	Upper Extremity
WMFT	Wolf Motor Function Test

CONTENTS

1	INTRODUCTION	31
1.1	Motivation	32
1.2	Objective	33
1.3	Chapter structuring	33
2	BACKGROUND	35
2.1	Knowledge Discovery and Healthcare data	35
2.1.1	<i>KDD process</i>	36
2.1.2	<i>Feature extraction and selection</i>	39
2.1.3	<i>Robotic Rehabilitation</i>	43
2.2	Statistical Learning Theory	44
2.2.1	<i>Empirical Risk Minimization</i>	47
2.2.2	<i>Generalization bound</i>	52
2.2.3	<i>VC Dimension</i>	53
3	RELATED WORK	55
3.1	Robotic devices and data acquisition	55
3.2	Feature extraction	57
3.3	Analysis strategies and goals	59
3.3.1	<i>Human labor</i>	61
3.3.2	<i>Clinical scales and supervised learning</i>	61
3.3.3	<i>Analysis based on evidences in data</i>	63
4	KINEMATIC AND KINETIC DATA ANALYSIS FRAMEWORK	65
4.1	Related work	66
4.2	Development	66
4.2.1	<i>Data structure</i>	66
4.2.2	<i>Software organization</i>	69
4.3	Application	69
4.4	Discussion	71
4.5	Conclusions	73
5	ROBOT-BASED METRICS	75
5.1	Tasks	75

5.2	Preprocessing	77
5.2.1	<i>Savitzky-Golay filter</i>	77
5.2.2	<i>Radial-Basis Functions</i>	78
5.3	Macrometrics	78
5.4	Micrometrics	88
6	DEFINITION OF THE HEMIPARESIS-INHERENT DATA SPACE TOWARDS PATIENT'S PROGRESS ASSESSMENT	95
6.1	Domain understanding and definition of KDD goals	97
6.2	Selection of the target set	98
6.3	Preprocessing	99
6.4	Feature extraction	99
6.5	Data mining	101
6.5.1	<i>Definition of techniques</i>	101
6.5.2	<i>Machine learning and metaheuristics</i>	101
6.6	KDD results and interpretation	103
6.6.1	<i>Model configuration for fitness function</i>	103
6.6.2	<i>Final descriptor for the data space</i>	104
6.6.3	<i>Learning guarantees from SLT</i>	105
6.6.4	<i>Avoidance of bias in sampling</i>	108
6.6.5	<i>Testing on data with different protocol bias</i>	109
6.7	Progress assessment based on uncertainty in classification	110
6.7.1	<i>State of rehabilitation</i>	110
6.7.2	<i>Time dependency</i>	110
6.7.3	<i>Results and discussion</i>	113
7	THE IMPACT OF TDCS PRECEDING INTENSIVE ARM TRAIN- ING IN CHRONIC STROKE	115
7.1	Study Overview	116
7.2	Data Analysis	117
7.2.1	<i>Analysis of kinematic and kinetic data</i>	117
7.2.2	<i>Post-hoc exploratory analysis</i>	119
7.3	Results	119
7.4	Discussion	125
7.5	Conclusion	130
8	ESTIMATION OF CLINICAL SCALES BASED ON KINEMATIC MEASURES FOR CHRONIC STROKE	131
8.1	Clinical Outcome Measures for the UE	133
8.2	Dataset format and selection of UE Kinematic and Kinetic features	133

8.3	Clinical estimator models	135
8.4	Results	136
8.4.1	<i>Correlation between clinical scales</i>	137
8.4.2	<i>Correlation between robotic data and clinical scales</i>	137
8.5	Discussion	139
8.6	Conclusions	141
9	EARLY PREDICTION OF CLINICAL OUTCOMES FOR CHRONIC STROKE PATIENTS	143
9.1	Time dependency	144
9.2	Machine learning	144
9.3	Results	145
9.4	Discussion	148
9.5	Conclusion	149
10	CONCLUSION	151
10.1	Publications	153
	BIBLIOGRAPHY	155
APPENDIX A	MATHEMATICAL PERSPECTIVE ON THE SPREAD OF CLINICAL SCORES	169
APPENDIX B	NONLINEAR MODELS FOR ESTIMATING THE FMA- UE	171
APPENDIX A	APPROVAL FROM CAPPESQ ETHICS COMMITTEE	173

INTRODUCTION

An increasing life expectancy worldwide entails larger contingents susceptible to rehabilitation treatments and hence an increasing demand of such healthcare services. One of these expanding contingents is composed by people involved in incidents that caused severe wounds, such as car accidents. According to the IRTAD Road Safety Annual Report, 1.3 million people die on the roads, whereas tens of millions become injured every year ([ITF, 2016](#)); in Brazil, occurrences of accidents involving injured people overcome around nine times death occurrences ([DNIT, 2010](#)). Survivors who remain in motor recovery conditions are likely to be enrolled in rehabilitation programs. Furthermore, as a result of advances in healthcare technologies, elderly is another expanding contingent with more stability in their health conditions. However, such aged contingent has higher proneness to demand rehabilitation services either as a preventive measure, keeping good motor condition in upper or lower limbs, or to gradually redevelop motor abilities which occasionally present functional deficits due to a diagnosed pathology of physical or neurological order.

Motor behavior disorders do not arise only from local and physical injuries. Extrinsic reasons are also to be pondered, such as the integrity of electric pulses sent from the nervous system to the limb to properly perform its motor functions ([HORAK, 1990](#)). Not necessarily succeeding traumatic brain injury, a stroke is also likely to occur from natural causes, unrelated to external events. Every year, 15 million people suffer a stroke worldwide, having two thirds of survivors. From these survivors, only a third presents recovery conditions ([WHO, 2012](#)).

Such an increasing demand for rehabilitation has been motivating research projects ([KREBS *et al.*, 2008](#); [KREBS *et al.*, 2003](#)), aiming to propose novel protocols and rehabilitation approaches, as a paradigm shift involving robotic devices and computer games. Upper-limb robotic rehabilitation for stroke patients was recommended by the American Heart Association, as well as the Department of Veterans Affairs and the Department of Defense from the United States, as a complementary treatment to conventional therapy

(MILLER *et al.*, 2010; Dept. of Veterans Affairs; Dept. of Defense, 2010). In comparison to traditional rehabilitation, robotic rehabilitation affords a significantly higher number of repetitions per treatment session, keeping patients motivated. This approach consists of a robotic device, manipulated by patients, as an interface to a computer game. Gaming environments enhance patient's attention, turning physical effort to an engaging activity.

Apart from the mechatronic context, the computational activity of robotic devices extends to other pieces of software, such as games and the management of the acquired data arising from sensors. It is often possible to benefit from the collected data, as well as the available CPU for the purpose of extracting performance information of patients as an additional knowledge to the therapist's decisions along treatments.

1.1 Motivation

Since robotic devices are coupled to computers and sensors, there comes up the ability to store and analyze data, describing the motor behavior of patients during sessions. This data is stored in a database, enabling further and thorough analysis, as well as possible detections of patterns potentially tied to a particular pathology, among other non trivial knowledge representations that, arising from large data repositories, would not be likely to arise only from human findings. Despite this rich potential of data analysis, a lack of such practices is still quite noticeable (MORETTI; ANDRADE; CAURIN, 2013), and is starting to be filled, as more related work, addressing data analysis as the main purpose, arise.

Analysis approaches take place either within sessions, or from an overall perspective of the treatment. Regarding employment and proposal of methods under these approaches in related work (BOIAN *et al.*, 2003; KREBS *et al.*, 2003; BURKE *et al.*, 2009), one can observe such efforts being employed secondarily, as a complementary outcome to the robotic device being proposed. Although this is not related to the relevance of data analysis outcomes, a bigger effort, addressing the extraction of novel information as the main goal, would improve what is expected as a result when it comes to hypothesis formulation, manipulation of data and feature spaces, as well as proper interpretation.

Robotic devices typically analyze the collected data at the end of therapy sessions. The extracted information from these restricted time domains is left to human labor, if any, through data visualization. Even seeming feasible to lean on visual inspection, this task becomes less prone to entail consistent findings, as the number of sessions takes larger proportions, culminating in more representative and statistically significant datasets. Describing the entire treatment of a patient, the overall data is suitable to the usage of computational mechanisms for the detection of the underlying pattern inherent to hemiparesis or another pathology arising from stroke. Hereupon, high-level information,

arising from patterns, are more likely to entail consistent human findings either from data science perspective, for further analysis iterations, or from clinical perspective, offering human intuitive information regarding progress levels.

1.2 Objective

This PhD project proposes standards on the organization and analysis of this kind of data, as well as computational methods for tracking the patient's progress along robotic rehabilitation treatments based on different machine learning and data science perspectives.

The data for such kind of analysis comprise components arising from time-dependent kinematic and kinetic measures which describe the motor behavior of patients. The projects presented in this PhD thesis have different goals, according to the different ways the data can be processed and analyzed, from time-dependent to summarized time-static approaches. With premises from the Statistical Learning Theory, the impact of the methods here presented are pondered, when applicable, aiming the due learning guarantees, which may provide additional clarity on the consolidation in the clinical field.

Our efforts to investigate the aforesaid objectives attain different subjects multidisciplinary, which can be seen as the core contributions of this PhD project, as enumerated below:

1. Computational methods from a data science perspective for the analysis of rehabilitation robotics data towards evaluating and tracking patient's performance.
2. Standards for organization and sharing of patient data (demographics, clinical and robot-based) in terms of a data structure.
3. Modular tool for preprocessing of rehabilitation robotics data, allowing easy implementation of new features as needed.
4. Indicators of chronic stroke severity as biomarkers, in the form of machine learning models.
5. Clinical findings: impact of robot-aided rehabilitation and administration of neurostimulation on chronic patients; correlations between clinical scales and robot-based measures; responsiveness of patients to robotic rehabilitation treatments.

1.3 Chapter structuring

The remaining of this thesis is organized as follows: **Chapter 2** provides a background on the core concepts used along this work, encompassing subjects from Knowledge

Discovery over healthcare data and Statistical Learning Theory, as the means to ponder our results, when applicable. In **Chapter 3** we introduce a discussion concerning the main methods and devices found in the literature that aim to extract knowledge over robotic rehabilitation data, narrowing the focus to the assessment of the progress of stroke patients.

In **Chapter 4** we propose a modular framework (MORETTI; DELBEM; KREBS, 2020) for organizing programs for data preprocessing, visualization, and analysis, as well as the preprocessed data itself within a structure that can be exported in a portable format. The standards from our framework's data structure are intended to be used for simplifying the data sharing from The 77 Laboratory at MIT, which has one of the largest data sources in the field.

The implementation of robot-based metrics are explained in detail in **Chapter 5**. Owing to the fact that they remain subjectively explained in the literature, we aim to elucidate the way we interpreted, implemented and adapted such metrics, which also contributes for clarification in the field. Such metrics were implemented in our framework.

A clinically-naive method is presented in **Chapter 6**, concerning the definition of a data space, based on a Knowledge Discovery roadmap, for the classification of hemiparetic sides based on performance data from 19 chronic stroke patients. The defined space is the subject of study, analyzing the performance of patients towards a spatial representation of the state of rehabilitation, based on classification uncertainty.

In **Chapter 7** we investigate the impact of administering anodal transcranial direct current stimulation prior to therapy sessions of 82 chronic stroke patients. We employed our framework (**Chapter 4**) to conduct statistical analyses towards tracking overall trends on data at the end of the treatment.

Chapter 8 concerns the estimation of clinical scores from nonlinear models, so such kind of evaluation can be carried out deterministically (i.e., free of subjectivity). We investigate the value of kinematic data from the wrist device (combined with data from S/E device) to strengthen the correlation between regression models and traditional clinical evaluations of upper-extremity.

Based on the findings from **Chapter 8**, in **Chapter 9** we focused on the early prediction of clinical outcomes, i.e., whether patients are going to respond to the treatment, from a clinical perspective, after the first two therapy sessions. We implemented strategies both from time-dependent and time-static perspectives.

Finally, in **Chapter 10** we draw some conclusions over our results, as well as future directions of the research. Details on the publication of the aforementioned projects are also provided.

BACKGROUND

This chapter encompasses as a review some important concepts used along this thesis for a better understanding of the further sections. Subjects such as Knowledge Discovery over healthcare data and Statistical Learning Theory are discussed. It is important to point out that only the core ideas are raised along this chapter, for the purpose of providing basic knowledge regarding involved techniques.

2.1 Knowledge Discovery and Healthcare data

Knowledge Discovery in Databases (KDD) is a roadmap to build a process in which data is submitted for analysis, resulting in the extraction of knowledge, i.e., new high-level information, over the data domain of low-level informativeness provided as input. KDD processes handle data since its raw aspect, until yielding high-level knowledge.

Once it is possible to collect and store volumes of data, as well as to define knowledge formats as output, KDD becomes feasible in every domain which such computational abilities are possible to be performed, such as several domains of healthcare: biomechanical and biomedical variable, as well as every way of describing organic and physical phenomena arising from human behavior.

Pursuing knowledge demands the cooperation of experts in the field to be exploited. Such cooperation brings complementary expertise from technical perspective for better understanding the data domain, as providing solid foundations to properly formulate a concrete analysis process. Prior data understanding is the primary means to establish criteria for the choice of techniques and algorithms involved in the knowledge discovery process. Otherwise, poor data understanding leads to a poor analysis strategy which yields illegitimate information, usually called data dredging ([ANDERSON; BURNHAM, 1999](#); [FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996](#)). Besides field expertise, the study of the involved variables and the arrangement of its instances in space also adds to prior data

understanding. The mapping of data from an n -dimensional space $\mathbb{R}^n$ to an m -dimensional space $\mathbb{R}^m$, such that $m < n$ enables the possibility of better inspecting data (HOLZINGER; DEHMER; JURISICA, 2014) through visual inspections or mathematical tools from linear algebra further elucidated in Section 2.1.2.

Inspection of data during a KDD process (see Section 2.1.1) is a way to verify whether the expected output format at the end of a particular step is correct and hence ponder the relevance of the technique applied at that step. For instance, after data mining step, one may inspect the arrangement of decision boundaries for classification, distribution of clusters, slope of a linear function in regression, or any other spatial aspect that concerns interpretation. Inspection also helps in the detection of outliers (to be ignored), specially at preprocessing steps, avoiding to propagate noise to the overall output of the process. Techniques of visual inspections can be seen in (BARTKE, 2005; WARE, 2004; FAYYAD; GRINSTEIN; WIERSE, 2002).

2.1.1 KDD process

Among the main KDD models proposed in the literature (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; CABENA *et al.*, 1998; ANAND; BUCHNER, 1998; SHEARER, 2000; CIOU *et al.*, 2000; KURGAN; MUSILEK, 2006), the model proposed by Fayyad, Piatetsky-shapiro and Smyth (1996), depicted in Figure 1, is the main and most complete work which inspired the proposal of the further ones. It is composed by nine abstract steps in which input data processed. Ranging from domain understanding to interpretation and consolidation of knowledge, this roadmap resembles to a scientific model of approaching problems. A set of algorithms and techniques, assigned to each of the abstracts steps, compose a concrete strategy (the KDD process).

Figure 1 – Data flow of the Fayyad et al. KDD model.

Source: Adapted from Fayyad, Piatetsky-shapiro and Smyth (1996).

First step consists of domain understanding, as well as acquiring prior knowledge for further steps. Still regarding this step, it is necessary to establish guidelines, such as the knowledge format of the overall output, avoiding to perform data dredging. Such guidelines are typically related to the needs of the final user (i.e., an expert in the field the data depicts), who is going to benefit from the extracted knowledge.

Secondly, a target subset of the data is defined. Initially raw, data as a whole is not entirely tied to the knowledge expected at the end of the process. The guidelines defined at the previous step may only require a particular portion of the raw data which concerns the knowledge format to be attained. The remaining portion of the raw input is disregarded for the next steps.

Preprocessing and cleansing take place next, at the third step. The target data subset may be restructured in the form of tuples, handling potentially missing data as well. One must not address data dependency (be temporal or correlational) matters at this step, but every behavior of the involved variables must be preserved. Fine adjustments are also performed at this point, such as noise removal, conversion of metric units, rescaling and others.

The fourth step concerns data transformation. In case of large proportions of data due to time dependency or correlated variables, projection and dimensionality reduction techniques or simply descriptive statistics may be considered. These strategies are used for feature extraction (which is detailed in [Section 2.1.2](#) due to the need of big efforts with healthcare data), reducing the number of the involved variables. This step also concerns feature selection. Given a set of features (variables) from the previous step, or iteratively, extracting features already at this step, strategies involving metaheuristics or correlations are typically considered in case of performing feature selection. The output of this step must yield a concise and reduced dataset.

Data mining occurs from fifth to seventh steps. In the fifth step, the guidelines defined in the beginning and the available data are to be associated to a machine learning technique from supervised learning (classification, regression) or unsupervised learning (clustering).

Sixth step concerns choosing an algorithm (or a set of candidate algorithms) to perform data mining. A study of the data space may lead to an algorithm with a compatible family of functions for shattering regions spatially or fitting curves according to the available dataset.

In the seventh step, data mining is performed by executing the previously selected algorithms over the prepared dataset. It involves parameter tuning approaches for each candidate algorithm, as well as sampling strategies (k -fold cross-validation), appointing the candidate that best fits the dataset. This step must yield a model based on rules, or

from a family of functions, capable of assigning classes or clusters to unseen instances, or predicting the behavior of a variable. It is worth to mention that KDD steps constitute a chain in which poor strategies of previous steps propagates ahead noise (bad outputs) towards the end of this chain. Therefore, specially at this step, having algorithms sensitive to changes in data, poor practices may entail unexpected results and compromise the overall output.

The eighth step consists of interpreting data mining results. In addition to the study of the space, the induced model is evaluated in terms of metrics from supervised or unsupervised learning. It is also important to interpret the obtained results from the expert's perspective, which delivers a real evaluation of the analysis, deliberating whether further iteration to previous steps is still needed to be done.

Given the final results obtained and evaluated by experts at the previous step, the ninth step concerns consolidating such extracted knowledge. Ways to incorporate this knowledge are to be discussed, still demanding efforts from the involved expert. It is also possible to consider reiterations of the KDD process by adding the extracted knowledge to more data to be analyzed.

Although KDD has data mining as its core step, the remaining steps are still necessary, due to the order in which they are organized. The KDD roadmap comprise every technique to be applied, turning raw data to fine high-level knowledge.

KDD roadmaps define several steps to handle data. Even so, such steps broadly fall in one of three major types of data processing: pre-processing, data mining or post-processing (GOLDSCHMIDT; PASSOS, 2005). Data mining and post-processing deal with high-level information, whereas pre-processing handles data at its initially rudimentary aspects, where most of the efforts must be applied in order to obtain an input suitable for further meticulous processing. That being said, one can notice that among the existing KDD models, they become similar when it comes to the amount of effort spent at pre-processing steps.

Each major step of KDD models demands efforts in terms of time in different proportions, owing to the fact that substantiation of abstract steps, according to particular datasets, does not denote triviality. Kurgan and Musilek (2006) raised estimates of efforts spent at steps of different KDD models (CABENA *et al.*, 1998; SHEARER, 2000; CIOS; KURGAN, 2005), depicted at Figure 2 as a comparative of importance in terms of efforts and time.

It can be seen that the total effort from all steps exceeds 100%. This occurs due to the estimative nature, which individually approximates the relative efforts per step. Even inspecting subjective estimates, the authors unanimously point pre-processing as the most expensive step of a process.

Figure 2 – Effort spent in KDD steps.

Source: Adapted from [Kurgan and Musilek \(2006\)](#).

2.1.2 Feature extraction and selection

Data volumes originating from healthcare (biomedical and biomechanical data) typically take large proportions ([HOLZINGER; DEHMER; JURISICA, 2014](#)), owing to the fact that organic processes, being sensed at medical exams or monitoring procedures, must be represented or described with precision (more details, more data) for diagnoses to be properly perceived. By *organic process*, one can understand as a gradually-changing phenomenon which successively assumes different states. That being said, it is possible to bind time dependency to the successiveness of such processes, since each state depends on the previous one to occur.

According to a data mining technique to be applied, portions of data must be gathered and interpreted as instances of events and, if needed, features may be extracted from these portions. Thus, a conceptual structure of the expected instances is already built, even before obtaining the actual features which substantializes the dataset for data mining.

Even taking large proportions, healthcare datasets may not be representative enough for what is expected to obtain as a result. Such statement can be better accepted by taking into account the necessary precision to properly describe any organic process, which makes representations of event instances naturally large. It is also important to consider that representativeness is intuitively given by large amounts of instances, as a proportionally-diversified subset of the universe of patterns regarding a particular subject to be exploited. For pattern recognition, time dependency must be eliminated, once it entails instances with variable sizes and unfeasibility for machine learning approaches. Feature extraction, also addressing issues of high dimensionality, culminate in a new set of variables, arising from the raw form of time-dependent instances, constituting a new data-space with significantly lower dimensions.

Features may originate from descriptive statistics, an attribute-wise feature extraction which takes the statistical moments of the distribution of the data (SACHDEVA *et al.*, 2013), turning an array of time-dependent measurements (considered from a single attribute) to a single value (per moment) related to the *bell-shaped* curve which describes data distribution.

Dimensionality reduction is another feature extraction (spatial-wise) strategy, aiming to apply spatial linear transformation in order to represent data in spaces with lower dimensions. Given a d -dimensional data space and an $m \times d$ dataset matrix $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, $\mathbf{x}_i \in \mathbb{R}^d$, there is a $d \times k$ transformation matrix H , such that $XH = B$, where B is an $m \times k$ matrix and $k < d$. Feature extraction methods based on linear transformations consist on finding matrix H .

The most used techniques for unsupervised and supervised learning problems are, respectively, Principal Component Analysis (PCA) (SMITH, 2002) and Linear Discriminant Analysis (LDA) (also known as Fischer Discriminant Analysis - FDA) (WELLING, 2009). PCA leans on the idea that if data can be represented in a subspace, then data has a flattened arrangement at the full space, as illustrated in Figure 3. In practical terms, original data X can be *flattened* (B) in lower dimensional spaces preserving a certain degree of variance from the original data. The lower the value of k , the less variance is preserved in B .

Figure 3 – Examples of data arrangements susceptible to projections in lower dimensions

Source: Elaborated by the author.

For PCA, values from each column of matrix X are centered (mean subtracted) and a covariance (if columns have values in similar scales, otherwise correlation must be used) matrix is calculated over the new state of matrix X . The eigenvectors ϕ_i , $i \in 1, \dots, d$, of the covariance matrix are called principal components, having their respective eigenvalues σ_i , $i \in 1, \dots, d$ (ordered descendingly) explains the percentage of variance covered by each principal component. In order to deliberate regarding the amount of variance to be lost,

the ratio of the variance explained by the first k eigenvectors, by the total percentage of variance, as illustrated in Figure 4.

Figure 4 – Usage of k principal components, according to their cumulative explained variance of the original data. (a) Depicts the cumulative explained variance as a function of the first k principal components and (b) defines the ratio of the function in the graph

Source: Elaborated by the author.

A linear transform from the obtained eigenvectors aims to place each component of the basis of the new space in the directions data spatially disperses the most. Matrix $H = \{\phi_1, \dots, \phi_k\}$ encompasses the first k principal components (the eigenvectors), forming a basis in $\mathbb{R}^k$, in which matrix B exists as the dimensionally-reduced dataset with the extracted features.

The reason LDA is the equivalent approach for supervised learning problems is due to a *limitation* of PCA in this context. PCA does not consider class labels if exists, giving no significant projections concerning data separability. For LDA, also being able to reduce dimensionality (hence extracting features), basis vectors are arranged taking into account class labels. LDA consists of an optimization problem, trying to trace a line in which data projection has the maximum distance between the center (mean) of per-class clusters, as well as minimizing the variance from each of such clusters. For the sake of comparison (see Figure 5), only the intuitive notion of linear projections with LDA is here elucidated; more details can be seen in Welling (2009).

Not only for feature extraction, PCA is also a tool for studying the data space, once accountability of variance from few or many principal components may be an evidence to be pondered in relation to distribution of data, as well as the possibility of the projection itself for feature extraction. Even being associated to unsupervised learning, projections for the visual inspection of decision boundaries from a classifier is also possible, once concerns regarding data separability must be addressed beforehand.

Feature selection also has the purpose of reducing dimensionality by selecting a subset of attributes which presents most relevance in relation to the quality of a model to

Figure 5 – Comparison of projections for supervised learning.

Source – Adapted from <http://www.svcl.ucsd.edu/courses/ece271B-F09/handouts/Dimensionality2.pdf>

be inducted. One must assume that no dependency relations exist between attributes, so selection strategies culminate in a subspace given by a subset of attributes with higher discriminative potential, in relation to the subjacent pattern.

Feature selection simply leaves out spurious variables (attributes, dimensions) in order to obtain a space with less dimensions. Specially at healthcare scenarios, feature selection may be seen as an alternative to feature extraction, due to preservation of variables and its scales, and hence its underlying semantics. Such alternative holds interpretability, which may bind and benefit both computer science and healthcare literatures.

Classifiers such as decision trees (QUINLAN, 1993) may be employed at this point for establishing selection criteria, because of the involving impurity measure (Gini, or Shannon's entropy, for example). Such classifiers yield a hierarchy of the most important features, having at the root the feature which its separation of instances isolates class labels the most. Afterwards, such tree-structure must be pruned and the selected features are the remaining ones near root. Still in supervised learning, the usage of metaheuristics, such as Genetic Algorithms (HOLLAND, 1992), approach, probabilistically, the search through combinations of features, having fitness associated to the performance of a classifier. Correlation-based selection (HALL, 2000) is also commonly considered, where correlations are related to a class label.

On the other hand, projection-based feature extraction involves another basis, yielding another space with different scales, from a linear combination of variables. Even being linear, pondering semantics to, only then, ponder whatever was interpreted by a value from its linearly-combined scale (e.g., whether a particular value indicates a potential disorder at an organic process), may not provide reasonable interpretations for healthcare fields. Consequently, since such kind of scales is semantically unknown, lacking bindings to perceive pathological disorders, further knowledge may be difficult to be consolidated

in healthcare literatures.

Complementarily to projection-based feature extraction (if feasible), feature selection must be performed first, removing spurious variables of low discriminative power, as well as strongly correlated variables. Leaving feature extraction to be done afterwards, less important variables were already removed, avoiding to affect variance of data and hence the arrangement of the new basis.

Besides eliminating spurious variables, high dimensionality is also an issue to be addressed. Classifiers tend to be less accurate, as the number of dimensions increases (TRUNK, 1979), even when every feature is strongly inherent to the involved classes. This effect over accuracy is known as the *Curse of dimensionality* (JAIN; CHANDRASEKARAN, 1982).

2.1.3 Robotic Rehabilitation

In comparison to traditional rehabilitation, robotic rehabilitation affords a significantly higher number of repetitions of motor exercises per treatment session, keeping patients motivated (BOSECKER *et al.*, 2010). Such an approach (KREBS *et al.*, 2008) consists of a robotic device, manipulated by patients, as an interface to a computer game. This virtual environment enhances patient's attention, turning physical effort to an engaging activity. The interaction of patients with robotic devices occurs specifically by a device coupled at the end of a robotic arm, enabling manipulation, which is called an *end-effector*.

Robotic devices generally support the upper-limb to be exercised (at the *end-effector*), being able to reproduce the same motor behavior of particular exercises, from traditional physiotherapy, for which they were designed, having one or more degrees of freedom in activity simultaneously. In this novel rehabilitation paradigm, exercises performed with these devices have passive, active-assisted and active-resisted movements (SANTOS, 2013).

When the robot is the only agent that causes the movements, the exercise consists of passive movements, assigning to the patient a merely passive role, because motor behavior is entirely driven by the robot. When every motor activity involves the efforts of the patient, but the aid provided by the robot is still necessary, movements of the exercise are active-assisted. In contrast, regarding active-resisted movements, devices assume an opposite role, presenting resistance to the movement performed by patients. The combination of these exercises during therapy sessions offers great potential of keeping patients engaged to the treatment, providing assistance in cases of difficulty, or even providing challenges if the aid of the robot is no longer needed.

When it comes to the involved data to describe events during rehabilitation ses-

sions, biomechanical variables, such as position, velocity and force are typically considered, being sensed by the device itself. Still regarding local measurements, techniques such as electromyography (EMG) (WANG *et al.*, 2015) sense the electrical activity arising from muscles, retrieving a signal (represented by a large array) per electrode. Electroencephalogram (EEG) (ANG *et al.*, 2010), as EMG, retrieves the same data format, but sensing brain activity.

These variables describe patient activity (motor or electrical) at sessions, and hence are time-dependent. That being said, strong efforts for preprocessing this raw data with techniques like the ones presented at Section 2.1.2 becomes necessary for data mining. Additional information arising from medical expertise (e.g. diagnosis and its severity) may provide guidelines for data mining strategies, possibly labeling instances for supervised learning, or as a complementary knowledge for interpretation.

2.2 Statistical Learning Theory

Proposed by Vapnik (1995), Statistical Learning Theory (SLT) consists of theoretical foundations to support interpretations over machine learning algorithms regarding their ability to learn. SLT provides mathematical guarantees concerning whether a mechanism properly learned from data, as well as how likely is such mechanism to deviate from the desired behavior, according to the involved universe of patterns; it may be a way to ponder induced models when it comes to confidence. For the sake of simplicity, this section encompasses foundations of SLT for supervised learning, more precisely binary classification. The contents of this section is based on the work of Luxburg and Schölkopf (2011).

Given a space of instances $\mathcal{X}$ and a label space $\mathcal{Y} = \{-1, +1\}$, learning consists of estimating an input-output relationship of the form $f : \mathcal{X} \rightarrow \mathcal{Y}$, i.e., a classifier f , mapping $\mathcal{Y}$ in $\mathcal{X}$, accessing training data $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n) \in \mathcal{X} \times \mathcal{Y}$. It must be assumed, concerning $\mathcal{X} \times \mathcal{Y}$ that there exists a *joint probability distribution* P and training data are independently and identically distributed (iid). Vapnik (1995) postulates five conditions, grounding SLT:

1. **Data instances are obtained iid.** No assumption is made on P .
2. **P does not belong to a specific family of distributions.** Data instances do not arise deterministically, being realistically tied to the error of humans, which supervises datasets of $\mathcal{X} \times \mathcal{Y}$.
3. **Labels may assume non-deterministic values, due to noise and class overlapping.** Datasets may contain ambiguous instances due to iid sampling.

4. **P has a fixed distribution.** No time-dependency cases are considered.
5. **P is unknown at training step.** Otherwise, a study of such distribution would entail the best classifier without inducing models.

When a classifier is obtained, it is necessary to evaluate its prediction ability. The *loss function*

$$l(\mathbf{x}, y, f(\mathbf{x})) = \begin{cases} 1 & , \text{ if } f(\mathbf{x}) \neq y \\ 0 & , \text{ otherwise.} \end{cases} \quad (2.1)$$

is defined for this purpose, given an instance $\mathbf{x}$, whereas the *risk* of a classifier

$$R(f) = E(l(\mathbf{x}, y, f(\mathbf{x}))) \quad (2.2)$$

is measured by the average loss over instances arising from distribution P . In practical terms, $R(f)$ is the expected value of the average loss of f for all $\mathbf{x} \in \mathcal{X}$, measuring how good f was adjusted according to P . That being said, it is possible to evaluate classifiers according to their risks. For example, let f and g be classifiers and $R(f) < R(g)$. Therefore, one can say that classifier f is better than g , being less prone to misclassification. However, $R(f)$ is hypothetical, since expected value requires the infinite points arising from P . On the other hand, the *empirical risk* does exist, being the risk over training data $\mathcal{X} \times \mathcal{Y}$:

$$R_{\text{emp}}(f) = \frac{1}{n} \sum_{i=1}^n l(\mathbf{x}_i, y_i, f(\mathbf{x}_i)). \quad (2.3)$$

From these definitions, generalization, according to [Vapnik \(1995\)](#), is defined by $|R(f) - R_{\text{emp}}(f)|$; the smaller the divergence, the more generalized is f and $R_{\text{emp}}(f)$ is a good estimate of $R(f)$.

Since classifiers have different functional structures (i.e., based on distance, probability, etc), it can be stated that a classification algorithm encompasses certain classes of functions (bias), which is binded to the amount admissible functions (variance). Consider a dataset where $\mathcal{X}$ grows proportionally to its desired output $\mathcal{Y}$, and two classifiers: f being linear, and g being a high-order polynomial. Is it preferable to fit the dataset without (g) errors, or with residual (f) errors? This is known as the *bias-variance dilemma*. It is a dilemma because a complex function like g (more admissible functions) may fit training data properly, but may not be good enough with unseen data, whereas a simpler function like f (less admissible functions) may keep its residual errors even with unseen data, having $R(f) < R(g)$.

Let $\mathcal{F}_{\text{all}}$ be a hypothetical space encompassing all admissible functions and $\mathcal{F} \subset \mathcal{F}_{\text{all}}$ related to the bias of a classification algorithm ([Figure 6a](#)). The larger is $\mathcal{F}$, the larger

is its variance, and the weaker is its bias; the inverse also applies. Denoted by $f_{\mathcal{F}} = \arg \min_{f \in \mathcal{F}} R(f)$, this is the best classifier within the bias of $\mathcal{F}$; $f_n = \arg \min_{f \in \mathcal{F}} R_{\text{emp}}(f)$ is a classifier obtained from $\mathcal{X} \times \mathcal{Y}$; and f_{Bayes} is the best classifier within $\mathcal{F}_{\text{all}}$.

Figure 6 – Hypothetical space $\mathcal{F}_{\text{all}}$, encompassing $\mathcal{F}$. (a) Bias-variance tradeoff represented by $\mathcal{F}$; (b) Consistencies of a classification algorithm; (c) Estimation and approximation errors.

Source: Elaborated by the author.

As a property of a set of functions, *consistency* concerns the behavior of generalization as the sample (n) increases towards infinity. SLT defines two types of consistency a classification algorithm may have: *Consistency with respect to $\mathcal{F}$ and P* and *Bayes consistency with respect to P* (respectively I and II from Figure 6b). The former occurs if $R(f_n)$ converges (in probability) to $R(f_{\mathcal{F}})$ as the sample size increases, for all $\varepsilon > 0$, i.e.:

$$P(R(f_n) - R(f_{\mathcal{F}}) > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty \quad (2.4)$$

The latter occurs if $R(f_n)$ converges (in probability) to $R(f_{\text{Bayes}})$ as the sample size increases, for all $\varepsilon > 0$, i.e.:

$$P(R(f_n) - R(f_{\text{Bayes}}) > \varepsilon) \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (2.5)$$

By decomposing the converging term of Bayes consistency, as illustrates Figure 6c, the concept of *estimation error* (variance) and *approximation error* (bias) arise:

$$R(f_n) - R(f_{\text{Bayes}}) = E + A \quad (2.6)$$

$$E = R(f_n) - R(f_{\mathcal{F}}) \quad (2.7)$$

$$A = R(f_{\mathcal{F}}) - R(f_{\text{Bayes}}) \quad (2.8)$$

Regarding estimation error, it increases as more classes of functions are considered, becoming more difficult for $R(f_n)$ to converge to $R(f_{\mathcal{F}})$. At the same time, with large

Figure 7 – Hypothetical space depicting interpretations of estimation and approximation errors.

Source: Elaborated by the author.

variance, bias becomes weaker, hence approximation error decreases. Large variance is more likely to have $\mathcal{F}$ encompassing f_{Bayes} , but also a memory classifier, which would belong to one of the function classes of $\mathcal{F}$. Therefore, out of an excessively large number of functional behaviors, an algorithm with such adjustments is likely to perform *overfitting*. The opposite idea, when the considered classes of functions are not complex enough, estimation error is smaller, because even $R(f_{\mathcal{F}})$ is large, and it is less likely to have $\mathcal{F}$ encompassing f_{Bayes} . This is the case of *underfitting*. That being said, a balance in the tradeoff must be defined, as depicted in Figure 7.

2.2.1 Empirical Risk Minimization

Owing to the fact that $R(f)$ requires the infinite points arising from P , which is unknown, the possible way to proceed is by inferring a function f from $\mathcal{X} \times \mathcal{Y}$. If training data is large and representative enough, $R_{\text{emp}}(f)$ may be a good estimate of $R(f)$. Such approach is called the Empirical Risk Minimization principle (ERM).

The law of large numbers is the main motivation of this approach. It states that the mean of a variable ξ sampled iid, arising from a fixed probability distribution, converges to the expected value of ξ , as the sample size increases towards infinity. As defined in Equation 2.3, since $R_{\text{emp}}(f)$ is the mean of the loss function, Vapnik adapted the law of large numbers to the ERM context:

$$R_{\text{emp}}(f) = \frac{1}{n} \sum_{i=1}^n l(\mathbf{x}_i, y_i, f(\mathbf{x}_i)) \rightarrow E(l(\mathbf{x}, y, f(\mathbf{x}))) \text{ for } n \rightarrow \infty \quad (2.9)$$

where the loss function is the *variable* to be studied from the perspective of the law of large numbers. Therefore, from a large enough sample, $R_{\text{emp}}(f)$ approximates well $R(f)$, which is the actual object of interest. In order to evaluate how well such approximation is, Chernoff's inequality (CHERNOFF, 1952) (later generalized by Hoeffding (1963)) defines

an upper bound that asserts, probabilistically, whether the divergence between $R_{\text{emp}}(f)$ and $R(f)$ exceeds an arbitrary ε :

$$P(|R_{\text{emp}}(f) - R(f)| > \varepsilon) \leq 2 \exp(-2n\varepsilon^2). \quad (2.10)$$

However, this only holds for a fixed function f , unrelated to the increasing sample. [Luxburg and Schölkopf \(2011\)](#) point out that, since the above upper bound is a probability, exceeding ε is unlikely to happen with a large n and a fixed f , but not impossible. In addition, restrictions over $\mathcal{F}$ must be established, so memory classifier does not arise from such bias, because otherwise $R_{\text{emp}}(f)$ goes to zero as it is minimized, but is definitely not a good estimate of $R(f)$.

By restricting the admissible functions of $\mathcal{F}$, ERM becomes consistent leaning on the *worst behavior* encompassed by $\mathcal{F}$. Such strategy aims to benefit from the law of large numbers uniformly over all $f \in \mathcal{F}$. In practical terms, if it holds for the worst, it also holds for any other function within the same class of functions. That being said, *uniform convergence* is about requiring that, for all $f \in \mathcal{F}$, $|R(f) - R_{\text{emp}}(f)|$ becomes less than ε for a large enough n , or simply

$$\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| \leq \varepsilon. \quad (2.11)$$

If uniform convergence holds for any $f \in \mathcal{F}$, it also holds for a function f_n obtained from training data. Therefore,

$$|R(f_n) - R_{\text{emp}}(f_n)| \leq \sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| \quad (2.12)$$

$$P(|R(f_n) - R_{\text{emp}}(f_n)| > \varepsilon) \leq P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \varepsilon). \quad (2.13)$$

From the right hand side of (2.13), the law of large numbers uniformly holds over $\mathcal{F}$ for all $\varepsilon > 0$, if $P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$. Still from (2.13) and the underlying idea of (2.4) (see [Figure 6b](#)), the derivation below indicates that ERM is consistent with respect to $\mathcal{F}$:

$$\begin{aligned} & |R(f_n) - R(f_{\mathcal{F}})| \\ &= R(f_n) - R(f_{\mathcal{F}}) \\ &= R(f_n) - R_{\text{emp}}(f_n) + R_{\text{emp}}(f_n) - R_{\text{emp}}(f_{\mathcal{F}}) + R_{\text{emp}}(f_{\mathcal{F}}) - R(f_{\mathcal{F}}) \\ &\leq R(f_n) - R_{\text{emp}}(f_n) + R_{\text{emp}}(f_{\mathcal{F}}) - R(f_{\mathcal{F}}) \\ &\leq 2 \sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)|. \end{aligned}$$

Therefore,

$$P(|R(f_n) - R(f_{\mathcal{F}})| > \varepsilon) \leq P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \frac{\varepsilon}{2}). \quad (2.14)$$

So far the law of large numbers holds individually for each function $f \in \mathcal{F}$. In order to bind together the case of every function, so the law of large numbers uniformly holds for all functions, the *union bound* is considered. It must be assumed that the number of admissible functions is finite ($\mathcal{F} = \{f_1, f_2, \dots, f_m\}$), having the law of large numbers holding for each f_i , such that $P(|R(f_i) - R_{\text{emp}}(f_i)| > \varepsilon) \leq 2 \exp(-2n\varepsilon^2)$.

If the supremum culminates in an f (over all infinite possibilities), in which its divergence of the respective risks exceeds ε , that is because at least one case is true. So now the probability is considered over the m cases of $\mathcal{F}$, which is now a finite set. Such new probability may interpreted as “the probability of exceeding ε arising from f_1 *or* the probability of exceeding ε arising from f_2 *or* ... *or* the probability of exceeding ε arising from f_m ”. That is when the supremum is disregarded and the union bound takes place.

The union bound states that the probability of a union of events is upper-bounded by the sum of the probabilities of events individually. Since the events (risk divergences from each f_i , for $i = 1, \dots, m$) are coupled with *or*, the following derivation can be written:

$$P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \varepsilon) \quad (2.15)$$

$$= P(|R(f_1) - R_{\text{emp}}(f_1)| > \varepsilon \text{ or } \dots \text{ or } |R(f_m) - R_{\text{emp}}(f_m)| > \varepsilon) \quad (2.16)$$

$$\leq \sum_{i=1}^m P(|R(f_i) - R_{\text{emp}}(f_i)| > \varepsilon) \quad (2.17)$$

$$\leq 2m \exp(-2n\varepsilon^2) \quad (2.18)$$

Asymptotically, exp behavior from Chernoff bound in (2.18) is preserved, since m is a constant, so ERM over a finite set of functions is still consistent with respect to $\mathcal{F}$. However, in order to assume that $\mathcal{F}$ effectively encompasses finite functions (and hence for union bound to hold), m is defined from the *symmetrization lemma* and the *shattering coefficient*.

Training data $(\mathbf{x}_i, y_i)$, for $i = 1, \dots, n$ is everything available to compute the empirical risk of a classifier. Suppose there is another sample $(\mathbf{x}'_i, y'_i)$, for $i = 1, \dots, n$. Such hypothetical sample is called *ghost sample*, also drawn iid, as the actual one. Therefore, it is also possible to compute the empirical risk of the ghost sample, given by $R'_{\text{emp}}(f)$.

According to symmetrization lemma:

$$P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \varepsilon) \leq 2P(\sup_{f \in \mathcal{F}} |R_{\text{emp}}(f) - R'_{\text{emp}}(f)| \geq \frac{\varepsilon}{2}) . \quad (2.19)$$

[Luxburg and Schölkopf \(2011\)](#) interpret symmetrization lemma plausibly: transitively, if both empirical risks are close to each other (having both samples drawn iid from P), the actual empirical risk is also close to the true risk.

The idea of effectively restricting $\mathcal{F}$ to a finite set of functions is based on the yielded outputs which may be the same between certain functions. Although having different functional definitions, all classifiers in $\mathcal{F}$ that yield the same classification (hence having the same empirical risk) for all data points are reduced to that unique way to classify data points. This reduction idea is considered for every possible classification of n points, namely 2^n . Given both actual and ghost samples, the number of possible classifications goes to 2^{2n} . It is important to point out that such restriction on $\mathcal{F}$ is possible due to symmetrization lemma, because both terms of the supremum (the empirical risks) are computable.

It is still necessary to define how many different classifications are actually possible to be yielded arising from $\mathcal{F}$ (i.e., from the functional complexity of a given classification algorithm), which is expected to be less than 2^{2n} . This issue is addressed by the shattering coefficient.

Let training data be denoted by $Z_n = \mathcal{X} \times \mathcal{Y}$ and $|\mathcal{F}_{Z_n}|$ be the number of functions that yield different classifications over Z_n . The shattering coefficient $\mathcal{N}(\mathcal{F}, n)$ of $\mathcal{F}$ with respect to sample size n is defined as:

$$\mathcal{N}(\mathcal{F}, n) = \max\{|\mathcal{F}_{Z_n}| \mid \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathcal{X}\} . \quad (2.20)$$

Whenever $\mathcal{N}(\mathcal{F}, n) = 2^n$, it is said that $\mathcal{F}$ *shatters* n points. Taking a linear classification algorithm in $\mathbb{R}^2$ as an example, [Figure 8](#) illustrates $\mathcal{F}$ shattering $n = 3$ points, such that $\mathcal{N}(\mathcal{F}, 3) = 2^3 = 8$. Considering $n = 4$, $\mathcal{F}$ no longer shatters the sample, as depicted in [Figure 9](#). Out of $2^4 = 16$ possible classifications of 4 points, $\mathcal{F}$ fails in two cases. Therefore $\mathcal{N}(\mathcal{F}, 4) = 14$. In this regard, it is said that the shattering coefficient is a capacity measure of $\mathcal{F}$, effectively measuring the finite size of admissible functions.

Putting it all together, considering a sample of size $2n$ (due to symmetrization lemma), training data is now denoted by Z_{2n} , where the first n points arise from the actual sample, and the remaining n points hypothetically arise from the ghost sample.

Figure 8 – Possible classifications in $\mathbb{R}^2$ from an algorithm of linear bias and a sample of size 3.

Source: Elaborated by the author.

Figure 9 – Cases in which a linear bias fails to shatter a sample of size 4 in $\mathbb{R}^2$.

Source: Elaborated by the author.

$$P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \varepsilon) \quad (2.21)$$

$$\leq 2P(\sup_{f \in \mathcal{F}} |R_{\text{emp}}(f) - R'_{\text{emp}}(f)| > \frac{\varepsilon}{2}) \quad (2.22)$$

$$= 2P(\sup_{f \in \mathcal{F}_{Z_{2n}}} |R_{\text{emp}}(f) - R'_{\text{emp}}(f)| > \frac{\varepsilon}{2}) \quad (2.23)$$

$$\leq 2\mathcal{N}(\mathcal{F}, 2n) \exp(-n \frac{\varepsilon^2}{4}) \quad (2.24)$$

From the upper bound of (2.19), the supremum now handles a finite set of at most $\mathcal{N}(\mathcal{F}, 2n) \leq 2^{2n}$ functions (2.23). Then, the union bound and Chernoff bound are applied, obtaining (2.24). Therefore:

$$P(\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \varepsilon) \leq 2\mathcal{N}(\mathcal{F}, 2n) \exp(-n \frac{\varepsilon^2}{4}) . \quad (2.25)$$

Until the definition of (2.18) from the union bound, it was observed that ERM is consistent with respect to $\mathcal{F}$. The same is now going to be verified from the right hand side

of (2.25), which encompasses the union bound, symmetrization lemma and the shattering coefficient.

In the case of a strong bias, the shattering coefficient is considerably less than 2^{2n} . Assigning a polynomial behavior for such capacity measure, $\mathcal{N}(\mathcal{F}, 2n) \leq (2n)^k$, for a constant k , the right hand side of (2.25) is

$$2\mathcal{N}(\mathcal{F}, 2n) \exp(-n\frac{\varepsilon^2}{4}) = 2(2n)^k \exp(-n\frac{\varepsilon^2}{4}) \quad (2.26)$$

$$= 2 \exp(\log(2n)^k) \exp(-n\frac{\varepsilon^2}{4}) \quad (2.27)$$

$$= 2 \exp(k \log 2n - n\frac{\varepsilon^2}{4}) \quad (2.28)$$

which tends to zero, as $n \rightarrow \infty$ for all ε . So in this case ERM is consistent with respect to the biased $\mathcal{F}$. On the other hand, assigning the bias of $\mathcal{F}_{\text{all}}$ to $\mathcal{F}$, encompassing every admissible function, such that $\mathcal{N}(\mathcal{F}, 2n) = 2^{2n}$, the exponential behavior plugged to the right hand side of (2.25) is

$$2\mathcal{N}(\mathcal{F}, 2n) \exp(-n\frac{\varepsilon^2}{4}) = 2 \cdot 2^{2n} \exp(-n\frac{\varepsilon^2}{4}) \quad (2.29)$$

$$= 2 \exp(\log 2^{2n}) \exp(-n\frac{\varepsilon^2}{4}) \quad (2.30)$$

$$= 2 \exp(2n \log 2 - n\frac{\varepsilon^2}{4}) \quad (2.31)$$

which does not tend to zero as $n \rightarrow \infty$. Therefore, for $\mathcal{F} = \mathcal{F}_{\text{all}}$, there is no guarantee regarding consistency. From both cases, it is possible to conclude that bias provides grounds for guaranteeing consistence with respect to $\mathcal{F}$. It is important to point out, however, that cases in which no consistency is possible to be guaranteed from SLT must not be interpreted as inconsistent.

2.2.2 Generalization bound

So far the right hand side of (2.25) provides an upper bound for the probability of the divergence between true and empirical risks exceeding ε . In order to inspect such bound for a real case, given a sample size n and the shattering coefficient $\mathcal{N}(\mathcal{F}, 2n)$ arising from an algorithm, ε is also to be defined.

Given that ε is an arbitrary constant, its definition becomes difficult, since each sample of training data entails different risks and hence different estimates of the true risk. In addition, no prior knowledge from data space is available in order to establish a satisfying divergence for the worse case of true risk estimation. In contrast, the definition of a satisfying probability δ , as if the bound of (2.25) is already known, i.e., $\delta = 2\mathcal{N}(\mathcal{F}, 2n) \exp(-n\frac{\varepsilon^2}{4})$, it would be possible to obtain the right ε . Such reformulation introduces the *generalization bound*. The derivation below solves such upper bound for ε .

$$\delta = 2\mathcal{N}(\mathcal{F}, 2n) \exp\left(-n\frac{\varepsilon^2}{4}\right) \quad (2.32)$$

$$\frac{\delta}{2\mathcal{N}(\mathcal{F}, 2n)} = \exp\left(-n\frac{\varepsilon^2}{4}\right) \quad (2.33)$$

$$\log(\delta) - \log(2\mathcal{N}(\mathcal{F}, 2n)) = \log\left(\exp\left(-n\frac{\varepsilon^2}{4}\right)\right) \quad (2.34)$$

$$\log(\delta) - \log(2\mathcal{N}(\mathcal{F}, 2n)) = -n\frac{\varepsilon^2}{4} \quad (2.35)$$

$$\varepsilon^2 = \frac{4}{n} (\log(2\mathcal{N}(\mathcal{F}, 2n)) - \log(\delta)) \quad (2.36)$$

$$\varepsilon = \sqrt{\frac{4}{n} (\log(2\mathcal{N}(\mathcal{F}, 2n)) - \log(\delta))} \quad (2.37)$$

Plugging ε in the event of the left hand side probability of 2.25,

$$\sup_{f \in \mathcal{F}} |R(f) - R_{\text{emp}}(f)| > \sqrt{\frac{4}{n} (\log(2\mathcal{N}(\mathcal{F}, 2n)) - \log(\delta))} \quad (2.38)$$

whenever the worst case exceeds ε , for a given probability δ , it can be interpreted as the case when $R_{\text{emp}}(f)$ is not a good estimate for $R(f)$. The generalization bound (2.40) consists of the opposite idea, concerning the case of having a good estimate of $R(f)$. The inequality is hence inverted:

$$R(f) - R_{\text{emp}}(f) \leq \sqrt{\frac{4}{n} (\log(2\mathcal{N}(\mathcal{F}, 2n)) - \log(\delta))} \quad (2.39)$$

$$R(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{4}{n} (\log(2\mathcal{N}(\mathcal{F}, 2n)) - \log(\delta))} \quad (2.40)$$

where $R_{\text{emp}}(f)$ is less than $R(f)$ because it is the risk over a sample, whereas $R(f)$ is the risk involving every data point, including the ones computed by $R_{\text{emp}}(f)$.

2.2.3 VC Dimension

Both generalization bound (2.40) and the upper bound of ERM (2.25) involve the shattering coefficient. Nevertheless, this capacity measure is difficult to evaluate analytically, which motivates methods for yielding estimations, or even choosing another capacity measure. The *VC dimension* (Vapnik-Chervonenkis) is another capacity measure, which evaluates the growth behavior of the shattering coefficient.

As previously elucidated, if every possible classification is able to be performed on Z_n by functions within $\mathcal{F}$, it is said that $\mathcal{F}$ shatters n points, such that $\mathcal{N}(\mathcal{F}, n) = 2^n$, with

a constantly exponential behavior. Denoted by $\text{VC}(\mathcal{F})$, the VC dimension, defined below, evaluates the growth behavior of the shattering coefficient:

$$\text{VC}(\mathcal{F}) = \max \{n \in \mathbb{N} \mid |\mathcal{F}_{Z_n}| = 2^n \text{ for some } Z_n\}. \quad (2.41)$$

This formal definition retrieves the maximum base-two exponent n , which causes $\mathcal{N}(\mathcal{F}, n)$ to grow exponentially. In practical terms, $\text{VC}(\mathcal{F})$ retrieves the maximum n , such that there exists a sample of size n that $\mathcal{F}$ is able to shatter. If such maximum does not exist, it is said, by definition, that the VC dimension is infinite.

RELATED WORK

This chapter introduces a discussion concerning the main methods found in the literature that aim to extract knowledge over robotic rehabilitation data. As a way to establish reasonable means of overall comparisons between methods, this discussion begins presenting the main robotic devices involved in data acquisition for further analysis. The focus is afterwards narrowed to methods for the assessment of the progress of patients along treatments, as well as which variables were previously sensed and considered for analysis.

Based on the underlying evidences that reside in data, related works are here overviewed from computational perspective, so consolidative bindings to healthcare literature lie beyond the scope of this discussion. It is important to point out that hemiparesis is the common pathology to be considered among the data analysis approaches here elucidated.

3.1 Robotic devices and data acquisition

The herein robotic devices are designed for upper-limb rehabilitation. [Hogan *et al.* \(1995\)](#) and [Krebs *et al.* \(1998\)](#) propose the MIT-MANUS (or InMotion ARM, as its commercial version), designed to perform reaching movements through its end-effector. Based on a SCARA (selective compliance assembly robot arm) mechanism, its planar mode has two degrees of freedom (a three-DOF configuration is also available), and is able to sense and record (at 200Hz) kinematic and dynamic measures such as positions, velocities and forces. This device operates with passive, active-assisted and active movements arising from patients efforts, having a virtual environment displaying feedback to the motor behavior.

As a component of the GENTLE/A rehabilitation system, [Linde and Lammertse \(2003\)](#) designed a device called HapticMaster. Sensing forces, this device also provides

positions and velocities, but those are obtained from a virtual model. It has three Degrees of Freedom (DOF) and also enables patients to perform passive, active-assisted and active movements through the end-effector. A virtual reality environment is available to provide feedback to the motor behavior of patients. From the collected measures, [Chemuturi, Amirabdollahian and Dautenhahn \(2013\)](#) also derived the jerk measure, programming the device to follow the minimum jerk trajectory and analyze point-to-point movements.

The T-WREX (commercialized as the Arneo Spring by Hocoma, AG, Switzerland) device ([SANCHEZ *et al.*, 2006](#)) has six DOF and does not have robotic actuators, being proposed as a low-cost solution. It is coupled with sensors that acquire measures of positions and forces at 66Hz. A web-based system provides feedback of patient's efforts through a virtual reality environment. The springs of this exoskeleton provides assistance by reducing or eliminating the effects of gravity, enabling patients to better perform movements actively.

Braccio di ferro ([CASADIO *et al.*, 2006](#)) is another two-DOF device based on the SCARA mechanism for reaching movement exercises, which is possible to be adjusted to operate vertically. At 100Hz, this device senses and collects position and force measures exerted in the end-effector, also being coupled with a virtual environment for visual feedback of motor behavior. Patients may perform active-assisted, active, or active-resisted movements.

Designed for reaching movement exercises, the KINARM ([SINGH; SCOTT, 2003](#)) device also has two DOF and minimizes effect of gravity by supporting each segment of the limb. Its sensing capacity is limited enables to collect position, velocity and angle measures at approximately 300Hz. Active-assisted and active-resisted movements are performed by patients. Visual feedback for motor behavior is similar to one of the virtual environments of the MIT-MANUS device.

There is also several other devices for upper-limb rehabilitation, such as MIME ([LUM *et al.*, 2002](#)), ARM Guide ([REINKENSMEYER *et al.*, 2000](#)), PFM ([GOMI; KAWATO, 1997](#)) and others. [Table 1](#) summarizes the main features of the robotic devices and [Figure 10](#) illustrates the most used ones in the related works here discussed. The aforementioned devices were chosen to be elucidated, rather than the ones just here enumerated, because of similarities of devices, as well as works using these devices for data analysis methods, in relation to the materials and methods presented in [Chapter 6](#).

Despite not every related work involves machine learning techniques, it is necessary to preprocess and organize the raw input for analyzing data properly. That being said, the next section discusses the used methods for such preliminary task, having the aforementioned devices of [Table 1](#) as part of the used materials, as well as which kind of data were extracted from the raw input arising from the robotic devices. Those same methods are also discussed in regards to their analysis goals and afterwards categorized, for finer

Figure 10 – Illustrations of some robotic devices from Table 1.

(a) MIT MANUS device

Source : Adapted from Krebs *et al.* (1998) and https://www.lifetecinc.com/images/products/Company_177/Image_8637.jpg.

(b) Braccio di Ferro device

Source : Adapted from Casadio *et al.* (2006) and <http://physioassistant.com/horizontal.gif>.

Semi-transparent mirror   Fiberglass braces

(c) KINARM device

Source: Adapted from Singh and Scott (2003).

comparison, concerning existence of human labor behind the obtained results, dependency of data analysis in relation to clinical scales, or whether the analysis is purely based on the evidences that underlie data.

3.2 Feature extraction

As previously highlighted in Section 2.1.2, both feature selection and extraction techniques are possible to perform over raw data. For the sake of interpretability, regarding

Table 1 – Summary of the main features of the robotic devices.

Device	DOF	Acq. rate	Data	Movements
MIT-MANUS Hogan et al. (1995) Krebs et al. (1998)	2	200 Hz	Positions, velocities, forces	Passive, active-assisted, active
HapticMaster Linde and Lammertse (2003)	3	N/A	Position*, velocities*, forces	Passive, active-assisted, active
T-WREX Sanchez et al. (2006)	6	66 Hz	Positions, forces	Active
Braccio di Ferro Casadio et al. (2006)	2	100 Hz	Positions, forces	Active-assisted, active, active-resisted
KINARM Singh and Scott (2003)	2	300 Hz	Positions, velocities, angles	Active-assisted, active-resisted
MIME Lum et al. (2002)	6	200 Hz	Positions, forces	Passive, active-assisted, active-resisted
ARM Guide Reinkensmeyer et al. (2000)	3	N/A	Positions, forces	Passive, active-assisted, active, active-resisted
PFM Gomi and Kawato (1997)	2	N/A	Positions, forces	Active-resisted

Source: Elaborated by the author.

feature extraction, in most of the cases it is preferred to keep the underlying semantics from the raw input ([MORETTI et al., 2016](#); [DIPIETRO et al., 2012](#); [COLOMBO et al., 2012](#); [JUNG](#); [GLASGOW](#); [SCOTT, 2008](#)) avoiding spatial transformations. However, it is sometimes the case that more than one technique, including spatial transformations ([BOSECKER et al., 2010](#)) for dimensionality reduction, is required. Then, if necessary, feature selection techniques are also possible to be performed.

It is worth pointing out once more that the final set of features must be inherent to hemiparesis, so that data analysis culminates in expected results. Overall information of each related work regarding analysis goals and the involved robotic device are elucidated in [Section 3.3](#).

[Bosecker et al. \(2010\)](#) and [Dipietro et al. \(2012\)](#) quantify the effort of patients during sessions through macro and micro kinematic metrics through descriptive statistics. Deviation from the straight line connecting targets, aiming, movement mean, peak velocity, movement smoothness (ratio between mean and peak velocities) and movement duration are examples of macro metrics; number of submovements, submovement duration, overlap,

peak, interpeak interval and shape metrics (skewness and kurtosis) are examples of micro metrics. From both works, these features were extracted only from eight center-out point-to-point movements, out of the 16 sequentially performed movements which compose a star-like pattern (see more details of the exercises in [Section 3.3](#)), so the involved descriptive statistics quantify motor behavior with more specificity.

In [Bosecker et al. \(2010\)](#), a total of 20 extracted features was considered large. After properly normalizing the scale of the involved features, PCA was conducted for dimensionality reduction for each analysis strategy, respectively elucidated in [Section 3.3.2](#). According to [Bosecker et al. \(2010\)](#), fewer metrics, besides reducing model complexity, would perhaps reduce the number of tasks performed by patients, thereby reducing the overall time required for patient evaluation.

Also based on descriptive statistics, [Jung, Glasgow and Scott \(2008\)](#) extract eight kinematic measures similarly to [Dipietro et al. \(2012\)](#) and [Bosecker et al. \(2010\)](#), although a different device is used in this case. [Jung, Glasgow and Scott \(2008\)](#) describe the extracted features as mean velocity of the first 500 milliseconds between the end of the previous trial and the beginning of the next trial; reaction time, as the time interval until the subject starts to move after a target is indicated; local maximum velocity of the first reaching movement; global maximum velocity; first movement distance error and first directional error, respectively as errors in distance and angular direction in relation to the straight optimal path (length and direction) to the target at the first movement; duration of a reaching movement and path length ratio, as the ratio between actual path length performed by the subject and the optimal distance between the center and the target.

The extracted features in [Colombo et al. \(2012\)](#) are defined as active movement index, as the percentage of trajectory travelled by means of the patient's voluntary activity; mean velocities; mean distances (current position from absolute target position) and normalized path length, as the ratio between path length of patient trajectory and the straight path.

In previous work ([MORETTI et al., 2016](#)), yielding 24 features purely based on descriptive statistics, we calculated the statistical moments of the distribution of kinematic and dynamic measures, such as positions, velocities and forces. Different from [Dipietro et al. \(2012\)](#) and [Bosecker et al. \(2010\)](#), in which the same device was used, we considered every movement (backwards or towards center) of the star-like pattern, so every feature is a descriptor of the entire cycle of movements composing such pattern.

3.3 Analysis strategies and goals

The robotic device MIT-MANUS was used in ([MORETTI et al., 2016](#); [DIPIETRO et al., 2012](#); [BOSECKER et al., 2010](#); [KREBS et al., 2003](#)), also collecting data from

the same gaming environment. Displaying eight targets around the edge of a clock-like interface, having its positions arranged at every 45 degrees, and a central target, the goal is to perform reaching movements towards every indicated target, and backwards, towards center. The first indicated target is at “twelve o’clock”, then central target is indicated, and the next targets are indicated clockwise. More details from this device, which is used in our work, can be seen at [Chapter 6](#).

In relation to MIT-MANUS, the KINARM device, used in ([JUNG; GLASGOW; SCOTT, 2008](#)), presents the exact same arrangement of targets around a circular edge and the central target. Also with the goal of performing reaching movements towards the indicated targets, the sequence in which targets are indicated is random, having the same probability of being picked, so after a total of repetitions every target would be reached the same number of times.

Used in ([COLOMBO *et al.*, 2012](#)), the device Braccio di Ferro presented a different visual interface for patients to perform reaching movements. The goal is to perform point-to-point movements in sequences that shape geometrical figures.

Table 2 – Demographics from data samples of related work

	Subjects SA/Ch/Co	Gender M/F	Lesion L/R	Age mean±sd
Dipietro <i>et al.</i> (2012)	42/116/0	97/61	63/95	60 ± 1.5
Bosecker <i>et al.</i> (2010)	0/111/0	70/41	52/59	59.9 ± 12.6
Moretti <i>et al.</i> (2016)	0/16/0	8/8	9/7	57 ± 11
Colombo <i>et al.</i> (2012)	0/22/0	15/7	N/A	53 ± 11
Jung, Glasgow and Scott (2008)*	0/105/84	N/A	52/53*	> 20

Source: Elaborated by the author.

Regarding representativeness of data samples, [Table 2](#) encompasses some demographics of the data from the related works under discussion. Subjects column organizes the number of sub-acute (SA) and chronic (Ch) stroke patients, as well as control (Co) subjects; Gender organizes male/female patients (not subjects); Lesion organizes patients diagnosed with left or right hemiparesis; Age encompasses the average age (at enrollment) and the standard deviation. Although in ([JUNG; GLASGOW; SCOTT, 2008](#)) the patients were diagnosed with hemiplegia, this work was considered due to their raised evidences of rehabilitation conditions. Still in ([JUNG; GLASGOW; SCOTT, 2008](#)), all subjects (control and patients) are right-handed.

Several strategies take place for the purpose of assessing the progress of patients enrolled in rehabilitation treatments. [Dipietro *et al.* \(2012\)](#) presented a study about motor learning, where patients recover motor abilities in a generalized way. From a group of exercise movements performed during rehabilitation, patients were afterwards able to learn

how to perform other movements (different from the ones previously exercised), which were initially unable to be performed. [Colombo *et al.* \(2012\)](#) assess patient conditions indirectly from their proposed algorithm called Progressive Task Regulation. Such algorithm is used to tune parameters that model difficulty in the activities performed by patients.

A machine learning strategy ([JUNG; GLASGOW; SCOTT, 2008](#)) aims to propose a measure based on the outputs of a hierarchical ensemble classifier, considering performance data from both healthy and impaired limbs. Also using machine learning, in ([MORETTI *et al.*, 2016](#)), as a preliminary work, we aimed the feature extraction for the spatial separability of features, inherent to hemiparetic sides, for further studies in the obtained space for proposing a way to quantify progress. [Bosecker *et al.* \(2010\)](#) assess patients conditions through estimates of clinical scales as a function of patient behavior.

3.3.1 Human labor

Analysis strategies take place either within sessions, or at a wider scope, over the course of entire treatments. Human labor may take place in the overall scope, by inspecting the behavior of metrics over time, as a way to make decisions for the purpose of optimizing such performance behavior along treatments.

Reducing the scope to sessions, [Krebs *et al.* \(2003\)](#) proposed some performance measures which are calculated at the end of a session by the MIT-MANUS device itself. Considering the clock-like gaming interface explained before, these metrics assess the ability to initiate movements, ability to move from the starting position to the target, ability to aim their movement along the target axis, and the ability to reach the target position. From the recorded kinematic data, the device is also able to plot trajectories performed by patients for visual inspection.

Considering such measures, the human labor in an overall analysis takes place by inspecting the calculated values over time, at the end of sessions. More intuitively, trajectory plots may also be inspected, expecting trajectories to resemble a star pattern along the treatment.

3.3.2 Clinical scales and supervised learning

In machine learning terminology, supervised learning concerns assigning the due functional behavior to instances of a certain domain, so that algorithms may yield models that properly fit the set of labeled instances that constitute previous knowledge. Such assignment (the supervision) typically arise from experts of the studied domain. Taking the same connotation, clinical scales were used ([BOSECKER *et al.*, 2010](#); [PATEL *et al.*, 2010](#); [DIJCK; VAERENBERGH; HULLE, 2009](#)) for the supervision of events arising from patient performance.

Clinical scales are used by clinicians for the purpose of quantifying motor ability of patients under rehabilitation processes. Among the existing scales (Fugl-Meyer Assessment, Motor Status Scale, Modified Ashworth Scale, etc), the Fugl-Meyer Assessment (FMA) is considered one of the most comprehensive quantitative measures of motor impairment following stroke (GLADSTONE; DANELLS; BLACK, 2002).

Supervision based on clinical scales are used either for yielding a continuous value for assessment (BOSECKER *et al.*, 2010), or for interpreting scale values as categories (PATEL *et al.*, 2010), considering discrete scales.

Bosecker *et al.* (2010) tested the performance of linear regression models in order to predict scores from FMA, Motor Status Score, Motor Power and Modified Ashworth Scale. Using wearable sensors, Patel *et al.* (2010) interpreted scale values as classes, using a random forest for classification. Dijck, Vaerenbergh and Hulle (2009) focus on the classification between normal and impaired subjects using a naive bayes classifier. Although supervision in (DIJCK; VAERENBERGH; HULLE, 2009) concerns actual nominal categories, the posterior probabilities from the bayesian classifier were correlated to FMA scores as a way to validate results.

Taking FMA as an example, comprising five assessment domains, as elucidated in (GLADSTONE; DANELLS; BLACK, 2002), FMA has each domain composed by items related to specific tasks to be evaluated. For each evaluation, a clinician may assign a value within a 3-point ordinal scale, as an assessment of whether the patient cannot perform, or is able to perform partially, or to perform completely a particular task. Such cases respectively receive grades of 0, 1 and 2.

Nevertheless, as the means of supervising performance data, the usage of this instrument brings some concerns to be taken into account, such as the subjectivity that arise from the assessment of clinicians. It may be sometimes the case that, from a given task, patient performance is not good enough to be graded as 3, but it is too good to be graded as 2. The grading decision is then left entirely to the clinician, whose interpretation may be optimistic (ceiling the grade) or pessimistic (flooring the grade), whereas other clinicians may provide different grades. Therefore, the supervision of performance data is likely to be biased with respect to clinician's interpretation. Adaptations of FMA, by increasing the range of the ordinal scale (GLADSTONE; DANELLS; BLACK, 2002), may reduce uncertainty, but the scale remains discrete and hence still prone to a noisy supervision.

In supervised learning and SLT contexts, even under mild conditions in which instances must be drawn iid, such relaxation must be as slight as possible, by considering a representative set of instances. Assuming it is important to provide learning guarantees (as elucidated in Section 2.2), for the supervision based on clinical scales to become feasible, the number of clinicians involved in the assessment of patient performance must be large

enough, in order to ensure representativity, so that bias in supervision can be sufficiently diminished.

It is worth to mention that regardless the supervision on data, possible biases arising from procurement of instances must also be addressed. More details regarding this issue is discussed in [Section 6.6.4](#).

3.3.3 Analysis based on evidences in data

Fetching evidences in the collected data, [Dipietro et al. \(2012\)](#), [Colombo et al. \(2012\)](#) and [Jung, Glasgow and Scott \(2008\)](#) may have complementary results with clinical scales, but such instruments were not used for overall validation.

Overall change trends in motor performance were analyzed in ([DIPIETRO et al., 2012](#)) using two-tailed t-tests. For each subject, with a level of significance $p < 0.05$, performance data from trained and untrained exercises at admission and discharge were compared in a one-time analysis. [Colombo et al. \(2012\)](#) assess performance data indirectly using the Progressive Task Regulation algorithm. It continuously analyzes patient performance by predefined rules associated to adjustments of parameters that model difficulty in the proposed exercises.

In regard to supervised learning for performance assessment, the expected outputs binded to data instances must be either numerical or nominal. The provision of a continuous output from an induced model may arise from a direct estimate, whose training data were supervised with values from the variable itself that is desired to be functionally approximated.

There is also the possibility of raising a metric, by considering the continuous output of classifiers, from the indirect supervision with nominal values which, in this case, must not establish any relationship between admissible values other than equality. Otherwise, having other relationships that lean on an underlying semantics would aggregate noise, due to uncertainty. Hence, it leads to a subjective assignment of class labels, so that supervision is again likely to be biased, as the case of clinical scales elucidated in the previous section.

From binary classification, [Jung, Glasgow and Scott \(2008\)](#) labeled instances with the format of features described at [Section 3.2](#) as the motor behavior of control or non-control subjects, as 0 and 1 respectively. A Multi-Layer Perceptron (MLP) was used for each subclassifier of a hierarchical ensemble. For each subclassifier, data instances were equally sampled, without replacement, and three, out of the the eight extracted features, were randomly selected, defining the actual inputs of the model. The ensembler also consists of a neural network, which is fed by the continuous outputs of each subclassifier, yielding also, at the output, the classification between control and non-control subjects. As mentioned earlier in [Section 3.3](#), they collected data from both healthy and impaired

upper limbs; out of 16 subclassifiers, each half of this total was induced over performance data from an upper limb.

For progress assessment, [Jung, Glasgow and Scott \(2008\)](#) proposed a score measure to be yielded at the end of sessions. Given an instance describing the performance of a session, the score is defined as the sum of outputs from each subclassifier, related to each limb. The total from each side is represented as a dimension in $\mathbb{R}^2$, ranging from 0 to 8. Values closer to 8 indicate good motor conditions, whereas values closer to 0 indicate impairment. Performance scores from control subjects fall near 8 in both dimensions.

KINEMATIC AND KINETIC DATA ANALYSIS FRAMEWORK

In contrast to traditional methods, the use of robots for stroke rehabilitation (KREBS *et al.*, 1998; SANCHEZ *et al.*, 2006; CASADIO *et al.*, 2006) affords a significantly higher number of exercise repetitions per treatment session, as well as means to quantitatively evaluate patient's performance (DIPIETRO *et al.*, 2012; BOSECKER *et al.*, 2010; COLOMBO *et al.*, 2012). Robotic therapy employs common principles of motor learning (DIPIETRO *et al.*, 2012). It focuses patient's attention in performing the physical activity while engaged in a series of serious games. Hence it turns an onerous effort into an engaging activity, which may be key to achieving ultimate superior recovery outcomes.

Robotic devices can precisely acquire and register subjects' motor performance over the course of a study, entailing great potential for data analysis from calculations of quantitative performance metrics, as well as statistics. Metrics of such nature bring consistency to statistical analyses due to its determinism, overcoming subjectivity drawbacks of clinical scales that may lead to imprecise evaluations (LINCOLN *et al.*, 1991; KREBS *et al.*, 2002; SEMRAU *et al.*, 2013).

The volume of collected data in its raw form may be of large size, demanding proper preprocessing. Timewise, this phase is indeed the most expensive in terms of data analysis (KURGAN; MUSILEK, 2006). To address this issue, here we present a modular framework (MORETTI; DELBEM; KREBS, 2020) for organizing programs for data preprocessing, visualization, and analysis, as well as the preprocessed data itself within a structure that can be exported in a portable format.

4.1 Related work

In Yao *et al.* (2015) a medical big data processing system was proposed, intended to offer *the bigger picture* behind the healthcare data. Likewise, Hadoop-based approaches have been considered a gold-standard for dealing with considerably large amounts of data (see, for example, Greene *et al.* (2014) and Manogaran and Lopez (2017)). Although studies of human kinematics are sizeable enough to afford a statistical test (DIPIETRO *et al.*, 2012; COLOMBO *et al.*, 2012), their volume is static (i.e., data does not arise from a stream) and unlikely to attain order of magnitude of terabytes. In such a limited scenario, “Big Data” tools are potentially not pragmatic and/or a burden.

Database Management Systems (DBMS), relational or not (MORETTI; ANDRADE; CAURIN, 2013), may be a feasible way to organize both raw and preprocessed data (in terms of metrics and statistics) (TSUJI *et al.*, 2013). It has advantages in terms of gathering data from different sources, as well as the application of engines for retrieving chunks of data as needed. On the other hand, a centralized system may become unfeasible for redistributing data for other platforms as well as for supporting additional robotic devices with different measurements.

Similar to our work, Mandery *et al.* (2016) proposed a database structure for the storage of raw movement data for the whole human body. However both for Mandery *et al.* (2016) and other discontinued work (HWANG; KIM; LEE, 2006; GROSS; SHI, 2001), the purpose of the data structure is to serve as a repository of raw data, not necessarily focused on aggregating additional pieces of software for the extraction and analysis of higher level features.

4.2 Development

The framework presented here was entirely developed using MATLAB® (Mathworks, Natick, MA) although the data structure itself can be exported to the JavaScript Object Notation (JSON).

This section is divided into two parts. The first part depicts the hierarchical organization of data, which can be queried in either MATLAB® or other programming language. The second part presents the framework behind the software implementations for statistical analysis and preprocessing in terms of metrics or features.

4.2.1 Data structure

The arrangement of preprocessed data is depicted in Figure 11. We chose to organize our data structure to take advantage of large amounts of data collected with the MIT MANUS (Figure 10a) robotic gym based on its directory tree (Figure 12) that stores the

raw data (KREBS *et al.*, 1998). It is important to point out that this structure carries the summarized description of the data collected by a robotic device in terms of features or performance measures. Hence, it does not serve as a database/repository for raw data.

The structure carries the date of when the structure was obtained, as well as the descriptive fields `title` and `desc`, provided by the user (e.g., data analyst), with information about all the metrics within `patients` array. Identified by their respective `id`, each patient has his/her `demographics` and `evaluations` to be carried out on a regular basis along a study or treatment. Since such data arise from human input (e.g., a clinician), our framework is hardcoded to a specific spreadsheet template, which is the commonly used tool to organize data. In such a tabular template, rows must be assigned to patients (value at column `id`), whereas the remaining columns make up the parameter-value pairs within `demographics` and `evaluations` fields in Fig 11.

The field names in gray (Figure 11) are placeholders, indicating where single or multiple user-defined instances of that kind of field can exist in a particular application context. For example, using MIT MANUS robot, `device` becomes `planar`; `scale` becomes `fm` (Fugl-Meyer, and other instances for other scales) for `phase1` through `phaseN`, which depends on the treatment protocol (e.g., `admission`, `midpoint`, `discharge`, `followup`); two instances of `mode` array (in which each element is a session): `eval` and `therapy`; multiple instances of `task`, such as `p2p` (point-to-point), and so on.

Sometimes there are observations to be taken into account for interpretation or even additional preprocessing of data from a specific movement (or set of movements)

Figure 11 – JSON structure of preprocessed data. Left and right sides of nodes are respectively the field name and data type.

Source: Elaborated by the author.

Figure 12 – Directory tree of MIT MANUS device

Source: Elaborated by the author.

Figure 13 – UML Sequence diagram of the execution of tasks, modes and metrics dynamically added to the framework for a single subject. Dashed rectangles are placeholders for programs to be dynamically located according to the naming format with words in gray.

Source: Elaborated by the author.

performed at a specific session. At **device** level, the array **note** carries observations written by the operator/therapist.

After programmatically traversing the entire directory tree (Figure 12), the calculated metrics are aggregated in terms of mean, standard deviation, and median. The summarizing values are placed at the same level as the **mode** array (which encompasses metrics of every movement), for each task and metric, as indicated in Figure 11. The criteria to automatically discern between metadata (non-aggregated) and metrics is the presence of an underscore character at the end of the field name.

Ultimately, the structure is exported to JSON format. Our framework also allows us to merge files obtained from another source (aggregating the metrics) in case of involving data collected at different hospitals or devices (of the same or different kind) in a single

study. Every common field of a subject within both files is matched in order to organize instances of `device`, `mode` and `task`.

4.2.2 Software organization

The framework structure was designed in a modular fashion so that novel devices – as well as their respective modes, tasks, and metrics – can be easily implemented. The Unified Modeling Language (UML) sequence diagram from [Figure 13](#) depicts the sequence of steps taken by the framework to locate the selected programs, given the name of the device related to the data to be preprocessed.

Given the device related to the data, when the main program `struct_get` starts to iterate through the list of subjects, it uses `device_tasks` program to retrieve all the modes and tasks related to that device. By joining this information according to the naming convention specified in the dashed squares of [Figure 13](#), both the files with collected data and the programs that define all of the set of metrics to be calculated within every specific task are located and executed. Following this rationale, new programs should be appended along with the existing ones and their information must be defined in the hardcoded list of devices, modes, and tasks within `device_tasks` program.

Once all programs have been properly implemented, the framework is able to provide visualization and report statistical analysis results aiming to verify whether there was a significant performance change, with respect to metrics, among the subjects involved in a particular study. From a list of desired statistical tests to be carried out (according to the availability from MATLAB® language), a table is generated with all the metrics against p-values of each test, be it a normality, paired or unpaired test, combining the phases specified in the structure. In case of unpaired statistical tests, it is also possible to establish a separation criterion from demographics information to divide subjects into two groups.

4.3 Application

We applied this framework in the rehabilitation robotics context using the MIT MANUS robotic gym (shoulder-and-elbow and wrist robots) for preprocessing the collected data of two stroke patients. We considered the point-to-point task, which consists of reaching movements from the origin towards eight different targets presented sequentially.

The application process is illustrated by [Figure 14](#). Describing patient’s motor activity, the raw data is organized by single movements in separate text files. Such files are separately provided as input to a metric function, which performs calculations in the sense of a biomarker, i.e., quantifying severity on a particular feature of the patient’s motor performance. That said, the severity *spectrum* of a metric must range from the worst to

the best level. In [Figure 14 a](#), the rationale of the *deviation* metric (elucidated ahead) was implemented in `metric_deviation` function.

In order to implement all the developed metrics, they must be associated to one or more tasks. In [Figure 14 b](#), for instance, metrics *deviation* and *aim* are associated to the *p2p* (point-to-point) task from the *eval* mode of the *planar* device. Filling the placeholders, [Figure 14](#) provides a concrete case of the sequence diagram in [Figure 13](#): when the framework is looking at files from *planar* device, `task_planar_eval_p2p` will be automatically called, which will now call `metric_deviation` and `metric_aim`, previously implemented by the user.

We considered here two metrics for the *point-to-point* task: The first metric is “Deviation from the straight line,” consisting of the root-mean square of the distance between every collected point over the course of the movement and the straight line connecting the origin to the target. Better movement quality leads to a lower deviation metric. The second metric is “Aim to target,” which is the angle between end-effector’s position (when 10% of peak speed is exceeded for the first time) and the straight line of the task. As in *deviation*, lower values of *aim* metric indicate better movement quality. More details on the implementation of those (and other) metrics are available in [Chapter 5](#), along with their rationalia, which can also be found in the literature ([DIPIETRO et al., 2012](#); [BOSECKER et al., 2010](#); [COLOMBO et al., 2012](#)).

In terms of the aforementioned metrics, [Figure 15](#) traces the assessment of the two patients in four evaluation phases over the course of the treatment: prior to treatment, admission, discharge, and a follow-up six months after the treatment. It is noteworthy that a metric is extracted from a point-to-point movement, whereas the task evaluation consists of aggregating all metric values from all movements of that task. The points in [Figure 15](#) are therefore mean values of the movements performed at each phase.

One can observe that Patient A has recovered motor function along treatment, since lower values for both metrics are indicative of better performance. In practical terms, the metrics indicate that Patient A has improved his/her accuracy and steadiness. In contrast, the assessments of Patient B indicate the opposite situation, in which motor function performance is deteriorating.

Group wise, tabular reports can be generated to track an overall progress of patients between two phases, given a list of the desired statistical tests to be conducted, following the format of [Table 3](#):

Customizations such as sample filtering/splitting criteria (namely demographic data, or a threshold on a clinical score) and type of statistical test (e.g., paired, unpaired) can also be taken into account.

Figure 14 – Example of development and implementation processes of metrics in the framework. After developing a metric that does calculations from raw data (a), it must be allocated within a task and such task must be appended in the task list of the respective device (b).

(a) Development of *deviation* metric

(b) Implementing *deviation* and *aim* metrics in the framework by associating them to *point-to-point* task

Source: Elaborated by the author.

4.4 Discussion

Although we interpreted data from two real de-identified stroke patients (Figure 15) in a *toy* case, our purpose is to demonstrate the concept of the analytic trait of our tool for processing raw data towards final results. One must consider more robot-based metrics for the proper interpretations (as the ones discussed in Dipietro *et al.* (2012), Bosecker *et al.* (2010), Colombo *et al.* (2012)), as well as a meaningful number of patients for proper statistical testing in a particular study.

Our framework as a whole is functional, easy to use, scalable, requiring no deep

Figure 15 – Progress of Patients A (recovering) and B (not recovering) described by Deviation and Aim robot metrics along evaluation sessions.

Source: Elaborated by the author.

Table 3 – Format of an overall report of statistical tests (per metric) over a patient sample. P-values of the provided statistical tests (e.g., t-test and Wilcoxon signed rank test) are reported towards comparing changes in metric values between phases like admission (adm), discharge (dc) and follow-up (fu).

Metric	t-test adm-dc	t-test adm-fu	t-test dc-fu	signrank adm-dc	signrank adm-fu	signrank dc-fu
aim (p2p)	0.3573	0.6982	0.8435	0.4921	0.8278	0.8574
deviation (p2p)	0.2243	0.0843	0.2346	0.1267	0.0201	0.1232

Source: Elaborated by the author.

understanding of its architecture when integrating novel devices and pieces of software. If it is possible to follow the same file structure as depicted in Figure 12, this framework becomes versatile to contexts other than rehabilitation to study human movements.

The modular structure of our framework has scalability with respect to the number of subjects, devices, and every other component that can be dynamically implemented. In addition, such a modularity is useful to increment future functionalities involving machine learning to extract features in the same sense as metrics were calculated. It opens multiple opportunities from pattern recognition to time series prediction.

It is worth noticing that it is unlikely that the data volume attained in rehabilitation

robotics studies will amount to terabytes of data. Hence, our framework affords processing chunks of data that can be mergeable and easily accessed/edited because of the portability of the JSON format. Furthermore, the framework is quite convenient because of its simplicity, not depending on environments that centralize storage, such as Hadoop, or a DBMS, which would require prior installation and configuration.

Although the data structure gives a summary of the raw data, especially the aggregated metrics, it also preserves the metrics of each movement separately. This versatility is important for having information both at a macro-and-micro scale. The latter case enables us to extract micro-metrics: decompose a movement into its building blocks in the sense of submovements, as proposed by [Woodworth \(1899\)](#) and exploited by [Krebs et al. \(1999\)](#) in the context of brain injury and rehabilitation.

Besides the MATLAB®-only compatibility of the core software, our framework has other limitations. It does not serve as a raw data repository, since it provides a summarized version of the data from the perspective of robot-based metrics. Devices outside of MIT MANUS family may need specific mapping for interfacing with raw data.

Apart from several opportunities that entail room for enhancements in future work, we focused our efforts on dedicated studies to verify: the impact in robot-based metrics caused by tDCS on chronic stroke patients receiving robotic therapy ([Chapter 7](#)), the correlation between clinical scales and robot-based metrics ([Chapter 8](#)) and the ability to predict patient's outcome from robot-based metrics at early phases of therapy ([Chapter 9](#)).

4.5 Conclusions

We presented a framework for the organization of data preprocessing, visualization, and analysis. It uses a data structure that carries the preprocessed data itself, which can be exported in JSON, a portable format, allowing us to access or edit the preprocessed data using other programming languages. Because it requires no prior installation or setup, and due to its modularity, its usage simplifies the implementation of the extraction of metrics related to robotic tasks. It is therefore a convenient and straightforward tool to be used in studies involving the analysis of human kinematic and kinetic data collected by robotic devices of any sort.

ROBOT-BASED METRICS

This chapter encompasses the rationalia and implementation details of the robot-based metrics used in the projects covered ahead in Chapters 7, 8 and 9.

As denominated in the robotic rehabilitation literature, let *metric*, or *feature* (a way to see a metric, from machine learning perspective), be a way to evaluate movement quality in the sense of a biomarker, i.e., quantifying severity on a particular trait of the patient's motor performance. That said, the severity *spectrum* of a metric must range from the worst to the best level monotonically.

We here categorize metrics in macrometrics and micrometrics, following the way they were proposed and interpreted in the literature (KREBS *et al.*, 1999; DIPIETRO *et al.*, 2012; BOSECKER *et al.*, 2010; COLOMBO *et al.*, 2012). *Macro* and *micro* refer to the thoroughness with respect to the motor functions of the arm. The rationalia behind the implementation details elucidated in the following sections are consistent with related work on the evaluation of patient's performance, where the same metrics are used.

5.1 Tasks

Our data analysis framework (Chapter 4) performs metric calculations from individual movements that belong to particular tasks. The metrics explained ahead in this chapter are associated to the tasks defined by MIT MANUS devices for shoulder/elbow (S/E) and wrist.

5.1.1 Point to Point and One Way Record

Essentially interchangeable, Point to Point (p2p) and One Way Record (owr) tasks consists in performing cycles of reaching (point to point) movements towards peripheral targets in 8 directions (N, NE, E, SE, S, SW, W and NW), and backwards, as illustrated

in Figure 16. No assistance/resistance from the robotic device is provided during such tasks.

Figure 16 – The virtual environment (Clock game) used by patients for rehabilitation exercises. On the left, the yellow circle indicates the end-effector position to be moved towards the current target (red circle). On the right, the flow of a cycle of 16 movements. All the targets (black circles) are 14cm away from the center.

Source: Elaborated by the author.

Point to Point task is performed in evaluation sessions, repeating five times the cycle of movements towards all targets. One Way Record task is performed in therapy sessions, repeating four times the cycle of movements towards all targets. Both tasks are available for S/E and wrist devices.

5.1.2 Circle drawing

Divided into four subtasks, it consists in drawing circles with the end-effector, combining different starting positions with motion directions: 3 o'clock clockwise (cw3), 3 o'clock counterclockwise (ccw3), 9 o'clock clockwise (cw9) and 9 o'clock counterclockwise (ccw9). The circle pattern is traced five times for each subtask. No assistance/resistance from the robotic device is provided during such tasks.

Although cw3, ccw3, cw9 and ccw9 are actually distinct tasks from MIT MANUS perspective, we are additionally aggregating them all as one task, *circle*, having a higher number of circular motion repetitions within one task. Available for the S/E device only, circle drawing tasks are performed in evaluation sessions.

5.1.3 Round dynamic

While the robot holds the end-effector in the central target of the clock game (Figure 16, on the left), in the Round dynamic (rdyn) task the patient must resist the force applied by the device, attempting to move towards the peripheral targets, one at a

time, indicated in red. All targets are marked only once in this task, which is available for S/E and wrist devices.

5.1.4 *Playback static*

As a variant of the Round dynamic task, in Playback static (pstatic) the roles are switched. The patient attempts to hold the end-effector steady at the central target, while the robot applies 28N of force towards the peripheral targets, one at a time, indicated in red. All targets are marked only once in this task, which is available for S/E and devices.

5.1.5 *Shoulder*

Divided into four subtasks, it consists in moving the end-effector vertically, combining different arm positions with motion directions: extension (sext), pushing down, arm forward; flexion (sflex), pulling up, arm forward; abduction (sabd), pulling up, arm at side; adduction (sadd), pushing down, arm at side.

Although sext, sflex, sabd and sadd are actually distinct tasks from MIT MANUS perspective, we are additionally aggregating them all as one task, *shoulder*, having a higher number of repetitions within one task. Available for the S/E device only, shoulder tasks are performed in evaluation sessions.

5.2 Preprocessing

Because some of the metrics are more sensitive to the noise from the raw data, we considered two candidate filtering/smoothing approaches to carry out as needed before the actual metric calculations: Savitzky-Golay (SG) filter and Radial-Basis Functions (RBF). For both approaches we used the MATLAB® native implementations (`sgolayfilt` and `newrb`, respectively).

5.2.1 *Savitzky-Golay filter*

The kinematic data that describe patient's motor function arise from the robot's encoder, which naturally aggregates noise. We used the Savitzky-Golay filter (ORFANIDIS, 1996) in order to smooth such kinematics, since noise could be propagated when taking higher order derivatives (e.g. Jerk) and hence distort the true behavior of patient's performance during therapy.

Using MATLAB® implementation in `sgolayfilt` function, we considered a third degree polynomial to be fit into a 11-point window ($m = 5$) that goes through the curve. When the number of points of the curve is less than the window size ($2m + 1$), the signal is not smoothed.

Because it preserves high frequencies (ORFANIDIS, 1996), SG filter becomes suitable to smooth this kind of data, specially in cases of big *jumps* in the end-effector's displacement, arising from sudden movements performed by the patient. A caveat to take into account, however, is that it is necessary apply the SG filter whenever the data is differentiated in order to eliminate even higher frequencies, such as the range not attainable by human motor function (DEUSCHL; BAIN; BRIN, 1998), hence considered as noise. Otherwise, jerk-based metrics, for example, would not be reliable due to the propagated high noise. That said, SG smoothing after every differentiation eliminates high frequency components non inherent to human motor functions and hence does not distort its true tendency kinematics-wise.

5.2.2 Radial-Basis Functions

When it comes to preserving the behavior of a smoothed curve along multiple differentiations, Radial-Basis Functions can be an alternative to the SG filter, requiring no more than a single preprocessing step.

It is important to point out that the building blocks (width of the gaussian functions) must be small enough to keep the true behavior of the signal and, at the same time, filter very high frequencies (avoiding to overfit the curve with very narrow gaussians) that may lead to the same problem of the SG filter. In order to attain such a balance, we considered gaussian functions with a spread of 0.2. There is, however, a tradeoff to take into account when dealing with this kind of smoothing. Because no higher frequencies are considered, RBF models will yield a poor fit of curves with big *jumps* (that result in large and narrow peaks in the speed profile), which is not a recurrent case in this context.

5.3 Macrometrics

This category encompasses metrics that provide an overall (macro) perspective on the quality of the arm motor functions.

5.3.1 Aim

Measured in radians, aim refers to the angle difference between the true point-to-point path (straight line) towards a target and the current end-effector position at the moment the patient starts to take action. For the latter, *action* refers to when the patient exceeds 10% of peak speed for the first time, as indicated by an X mark in Figure 17. Aim metric is derived from p2p and ovr tasks.

Equivalently to S/E, for the wrist device, aim metric follows the same definition for Flexion/Extension (FE) and Adduction/Abduction (AA) movements (in 2D). For

Figure 17 – Aim metric rationale. On the left, aim is the angle difference θ between the true path (black solid line) towards NE target and the point of when the patient starts to take action along the movement trajectory (blue dotted line). The X mark indicates such moment both in the speed profile (right) and in the position space (left).

Source: Elaborated by the author.

Pronation/Supination (PS), aim corresponds to the signal of angular position (direction) with respect to target at the time the patient exceeds 10% of peak speed for the first time. Lower values of the aim metric indicate better movement quality.

5.3.2 Duration

Measured in seconds, duration is the time window between exceeding 10% of peak speed for first and last times, as illustrated in Figure 18. In practical terms, duration refers to how long the patient takes true action on performing the movement entirely. Duration metric is derived from p2p and ovr tasks. Lower values of duration metric indicate better movement quality.

It is noteworthy that the patient's action is mostly part of the movement data, possibly having additional idle states before and after the active part. That being said, points within the active time window are the only ones to be considered not only for deviation metric, but also for mean speed, peak speed, speed shape and jerk metrics towards accurate assessments on movement quality.

5.3.3 Deviation

Measured in meters, deviation metric consists in the root-mean-square distance (perpendicular line, as on the right side of Figure 19) between collected positions and the reference line which describes the ideal (straight) path towards a target in a point-to-point

Figure 18 – Duration metric rationale. On the left, the blue dotted line traces the patient’s true movement trajectory (with the respective blue part on the right side) and the black solid line indicates the straight path towards NE target. On the right, the gray part of the speed profile indicates idle state, whereas the blue part (exceeding 10% of peak speed) indicates true action on performing the movement and its time length as the metric itself.

Source: Elaborated by the author.

movement, defined as follows:

$$\text{deviation} = \sqrt{\frac{1}{N} \sum_{i=1}^N \text{dist}(ax + by + c = 0, (x_i, y_i))^2} \quad (5.1)$$

where N is the total number of points within a movement, (x_i, y_i) is the i -th point and a , b and c are the coefficients that define the straight path towards a target in the form $ax + by + c = 0$. Moreover, the distance between the i -th point and the line is defined below:

$$\text{dist}(ax + by + c = 0, (x_i, y_i)) = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}} \quad (5.2)$$

Equivalently to S/E, for the wrist device, deviation metric follows the same definition for flexion/extension (FE) and adduction/abduction (AA) movements. Deviation metric is derived from p2p and ovr tasks. Lower values of deviation metric indicate better movement quality.

5.3.4 Dwell time

Measured in seconds, dwell time plays the opposite role with respect to duration metric. Illustrated in Figure 20, it refers to the time length of the patient’s idle state,

Figure 19 – Deviation metric rationale. On the left, the blue dotted line traces the patient’s movement trajectory and the black solid line indicates the straight path towards NE target in the form $-x + y = 0$. The squared area, magnified on the right side, indicates the distances between each point and the straight path, aggregated in Equation 5.1.

Source: Elaborated by the author.

i.e., the reaction time after a target is displayed in the game prompting patient’s action. The idle state ends when the patient exceeds 10% of peak speed for the first time. Dwell time is derived from p2p and ovr tasks. Lower values of dwell time metric indicate better movement quality.

Figure 20 – Dwell time metric rationale. On the left, the blue dotted line traces the patient’s behavior during idle state. Respectively, on the right, the blue part of the speed profile indicates the idle state. Indicating 10% of peak speed, the X mark indicates when idle state ends, so that dwell time is defined by its time length.

Source: Elaborated by the author.

5.3.5 Mean speed, Peak speed and Speed shape

Measured in meters per second, mean speed and peak speed are self explanatory, illustrated in Figure 21. Mean speed, peak speed and speed shape are derived from p2p and ovr tasks. Higher values of mean speed and peak speed suggest better movement quality.

Figure 21 – Mean speed and peak speed rationale. Peak speed is denoted by the star mark and the mean speed is denoted by the dashed line.

Source: Elaborated by the author.

Speed shape consists in the $\frac{\text{mean speed}}{\text{peak speed}}$ ratio as a way to assess movement smoothness. Such rationale arises from the minimum jerk trajectory principle (HOGAN, 1984), in which a fifth-order polynomial $x(t)$ that describes the smoothest path from point A to point B has the minimum hand jerk (derivative order $n = 3$).

Psychophysical experiments suggest that human reaching movements have a speed shape ratio of about 1.75, which best approaches the minimum jerk trajectory, that has a ratio of 1.875 (RICHARDSON; FLASH, 2002; FLASH; HOGAN, 1985). Speed shape becomes larger as the derivative order increases and $n = 3$ is the order that yield the closest value to 1.75: the minimum-acceleration trajectory ($n = 2$) has a ratio of 1.5, the minimum-jerk trajectory ($n = 3$) has a ratio of 1.875, the minimum-snap trajectory ($n = 4$) has a ratio of 2.186, and so on (SHADMEHR; WISE, 2005). Hence, better movement quality leads to speed shape values closer to 1.875.

5.3.6 Jerk metrics

Jerk is the third derivative of position. Following Krebs *et al.* (2013), besides shape speed, we considered two additional ways to assess movement smoothness based on jerk. The first one is the mean magnitude of jerk normalized by the peak speed:

$$\|\mathbf{j}\| = \sqrt{\ddot{\mathbf{x}}^2 + \ddot{\mathbf{y}}^2} \quad (5.3)$$

$$\text{jerk}_{discrete} = \frac{\frac{1}{N-3} \sum_{i=1}^{N-3} \|\mathbf{j}\|_i}{\text{peak speed}} \quad (5.4)$$

where $\|\mathbf{j}\|$ is an array of jerk magnitudes for all (x_i, y_i) points along a movement and N is the total number of points. Metric $\text{jerk}_{discrete}$ is more suitable for assessing discrete movements (FLASH; HOGAN, 1985).

The second variant consists in the root mean square of jerk normalized by duration:

$$\text{jerk}_{rhythmic} = \frac{\sqrt{\frac{1}{N-3} \sum_{i=1}^{N-3} \|\mathbf{j}\|_i^2}}{\text{duration}} \quad (5.5)$$

which is best for rhythmic movements (HOGAN; STERNAD, 2009).

It is important to point out that differentiation here takes place numerically, since it arises from the robot data, which leads to filtering issues to take into account. As mentioned earlier, SG filters are prone to propagate noise along differentiations of higher orders, because it preserves high frequencies. In contrast, and different from the other metrics, we considered here RBF for smoothing position data, to then numerically differentiate it towards jerk. In comparison, we also experimented the analytical way from the model obtained from position data. Precision-wise, it suffices to stick to the numerical approach, since both ways deliver the true behavior of jerk equivalently for calculating the metrics.

Both $\text{jerk}_{discrete}$ and $\text{jerk}_{rhythmic}$ are derived from p2p and ovr tasks. For both metrics, lower values suggest better movement quality.

5.3.7 Maximum displacement

Derived from rdyn task, and measured in meters (S/E) or radians (wrist), maximum displacement consists in the maximum magnitude of position data during a movement. Higher values of maximum displacement suggest better movement quality.

Equivalently to S/E, for the wrist device, maximum displacement metric follows the same definition for FE and AA (2D) movements. For PS, it consists in the maximum value.

5.3.8 Overall aim

Derived from rdyn task, and measured in radians, overall aim follows the same rationale as in aim metric, but the angle is calculated at the moment maximum displacement

is reached. Lower values of overall aim suggest better movement quality.

Equivalently to S/E, for the wrist device, overall aim metric follows the same definition for FE and AA (2D) movements.

5.3.9 Scatter and offset

Derived from pstatic task, and measured in meters (S/E) or radians (wrist), scatter and offset consist in the standard deviation and mean of position data along the entire cycle of movements. For both metrics, lower values suggest better movement quality, as a way to measure how well the patient resisted to the force, keeping the end-effector at the center.

Equivalently to S/E, for the wrist device, scatter and offset metrics follows the same definition for FE and AA (2D) movements and PS.

5.3.10 Major axis, Minor axis and Axes ratio

Major axis and minor axis, measured in meters, and axes ratio, are derived from circle drawing tasks (cw3, ccw3, cw9, ccw9 and circle).

Major axis and minor axis correspond to the radii of an ellipse fitted to the movement data (tracing a circular pattern), as illustrated in [Figure 22](#). Axes ratio is the $\frac{\text{minor axis}}{\text{major axis}}$ ratio that indicates the *goodness* in patient's attempt to draw a circle, so the closer to one axes ratio is, the better. For major axis and minor axis, higher values suggest better movement quality.

Different from [Dipietro et al. \(2007\)](#) that obtained major and minor axes using PCA, as described in [Oliveira, Simpson and Nadal \(1996\)](#), we followed [Formica et al. \(2012\)](#), that used a least-squares optimization-based approach, as described in [Fitzgibbon, Pilu and Fisher \(1999\)](#). The latter strategy presented better adaptation to the data in comparison to the former, that scales eigenvalues (major and minor axes) by a constant value.

5.3.11 Joint independence

Proposed by [Dipietro et al. \(2007\)](#), joint independence (also derived from circle-drawing tasks) indicates how independent shoulder and elbow are while performing movements. Such an independence is quantified by the correlation between shoulder and elbow angles θ_1 and θ_2 , respectively, as indicated in the two-link mechanism that models the human arm in [Figure 23a](#).

Using inverse kinematics of the human arm ([YOSHIKAWA, 1990](#)), θ_1 and θ_2 are defined as follows:

Figure 22 – Rationale for major axis and minor axis metrics. The blue dotted line traces patient’s movement trajectory (robot data) for ccw3 task. The black ellipse is the ellipse fitted to the robot data and dashed and dotted lines are, respectively, major and minor axes of the fitted ellipse.

Source: Elaborated by the author.

$$k = \sqrt{(x^2 + y^2 + l_1^2 + l_2^2)^2 - 2((x^2 + y^2)^2 + l_1^4 + l_2^4)} \quad (5.6)$$

$$\theta_1 = \text{atan2}(y, x) - \text{atan2}(k, x^2 + y^2 + l_1^2 - l_2^2) \quad (5.7)$$

$$\theta_2 = \text{atan2}(k, x^2 + y^2 - l_1^2 - l_2^2) + \theta_1 \quad (5.8)$$

where $-\pi \leq \text{atan2} \leq \pi$, l_1 and l_2 are upper and lower arm, respectively. For left-arm movements, position data in the x axis are flipped with respect to shoulder position (multiplying by -1), located at the origin. Hence, the same rationale and equations for the right arm can be used. Values for l_1 , l_2 , d_1 and d_2 are defined and visualized in [Figure 23b](#). More details can be seen in [Dipietro *et al.* \(2007\)](#).

In [Dipietro *et al.* \(2007\)](#) subject’s shoulder elevation was considered constant, also constantly projecting upper arm length l_1 in 45° vertically (e.g., $l_{1_{male}} = 0.282 \cos \frac{\pi}{4}$). Different from the rationale from Dipietro and colleagues, we considered that the angle changes according to subject’s hand position:

$$f(d) = -\frac{20^\circ}{r}d + 60^\circ \quad (5.9)$$

where $r = 0.14\text{m}$ is the distance between the center of S/E device workspace and one of the peripheral targets (see [Figure 16](#)) and d is the distance between the end-effector and the closest target to the arm. For example, $d = 0$ if the right arm places the end-effector at SE target, $d = r$ if the end-effector is at the central target and $d = 2r$ if the end-effector

Figure 23 – Joint independence rationale. (a) Arm model (top view) of the subject (gray) performing a cw9 task, where l_1 and l_2 are upper and lower arm, respectively, θ_1 and θ_2 are shoulder and elbow angles, respectively and $(x, y) = (X - d_1, Y + d_2)$. (b) Female (F) and male (M) parameters and body measurements extracted from [Dipietro et al. \(2007\)](#): d'_2 is the distance between device's table border and subject's body midline (0.114m for male, 0.107m for female), $dist$ is the distance between device's table border and the center of robot's workspace (0.23m), so that $d_2 = d'_2 + dist$; d_1 is the distance between shoulder and subject's body midline (0.175m male, 0.157m female); l'_2 is the forearm length (0.254m male, 0.234m female), wgd is the distance from the wrist to the grip line (0.076m male, 0.071m female) and rh is the robot handle radius (0.03m), so that $l_2 = l'_2 + wgd + rh$; l_1 is variable, so that by plugging the upper arm length (0.282m male, 0.264m female) into [Equation 5.10](#), it yields a projection ranging from l'_1 (0.141m male, 0.132m female) to l''_1 (0.265m male, 0.248m female).

Source: Elaborated by the author.

is at NW target; for the left arm, $d = 0$ at SW and $d = 2r$ at NE. [Equation 5.9](#) is plotted in [Figure 24a](#).

Let l_{actual} be the actual upper arm length with no projection. The projected l_1 is defined as:

$$l_1 = l_{actual} \cos f(d) \quad (5.10)$$

where l_1 is a function of the distance d and male or female upper arm measurements are to be plugged in l_{actual} . [Equation 5.10](#) is plotted in [Figure 24b](#).

Given θ_1 and θ_2 for all (x, y) points along the movement trajectory, joint indepen-

Figure 24 – Definition of l_1 . (a) upper arm angle (vertically) as a function of the distance d with respect to the closest target to the arm, defined in Equation 5.9. (b) Projection of the upper arm length as a function of d (Equation 5.10).

Source: Elaborated by the author.

dence is defined as:

$$jointind = \frac{C_{1,2}}{\sqrt{C_{1,1} \cdot C_{2,2}}} \quad (5.11)$$

where C is the covariance matrix of θ_1 and θ_2 arrays with all (x, y) points. From Equation 5.11, joint independence metric consists in the correlation between shoulder and elbow angles while performing a movement. Lower values indicate poor correlation, i.e., independence of the joints when performing movements, which suggest better movement quality.

It is noteworthy that the arm model from Figure 23a provides an estimate of θ_1 and θ_2 values, given a point (x, y) . This must be taken into account because the arm model parameters, universally used for both genders, are also estimates (DIFFRIENT, 1974). In addition, although subjects wear a safety restraint to diminish body/trunk compensation, it does not avoid such movements completely.

5.3.12 Orientation

Also following Dipietro *et al.* (2007), orientation metric concerns the direction along which most points are distributed in a circular movement. For that to be quantified, the slope of a linear model, fitted to the position data of a movement, is used.

Orientation metric is derived from circle-drawing tasks. Lower values suggest better movement quality.

5.3.13 *Delta z*

Delta z, measured in newtons, consists in the difference between peak z-axis force and the minimum force before the peak. This metric is derived from shoulder tasks (sflex, sext, sabd, sadd, shoulder), available for S/E device only.

5.4 Micrometrics

Apart from the stroke rehabilitation context for upper limbs, point-to-point movements have been studied for over a century, towards identifying the primitive motor components that make up the movement of limbs as a whole (WOODWORTH, 1899). Krebs *et al.* (1999) pointed out many studies that introduce the idea of submovements as the building blocks of point-to-point movements. In simple terms, let *submovement* be a bell-shaped impulse in the speed profile, in the sense that the limb starts to take action by gradually increasing the speed towards a peak and then gradually decreasing it afterwards.

For the decomposition of point-to-point movements into submovements, we adapted and implemented a greedy algorithm, proposed by Krebs (1997), as described in Algorithm 1.

The iterative process of Algorithm 1 is illustrated in Figure 25. At each iteration, given the current (local) peak speed sm_{peak} (the same as the global peak speed v_{peak} at the first iteration), SUBMOVEMENTRANGE walks around the peak (the l^* -th value in array $\mathbf{v}$), until the points at both directions no longer go down symmetrically at each step, i.e., diverging by more than 5%. Function SUBMOVEMENTRANGE yields the locations of such points r_0 and r around the peak. If the current peak is located at the beginning or end of the speed profile, the half curve is mirrored for the next steps, even though it is going to be considered an invalid submovement (see the discarded submovement from Figure 25 and line 20 from Algorithm 1). The half curve is also mirrored (and also considered invalid) when the walking process hits the beginning or the end of the speed profile before violating the symmetry condition.

A gaussian function is then fitted from the points in the range $[r_0, r]$ and their respective time offsets, provided that points here are spaced in time by $\Delta t = \frac{1}{200}$ seconds (200 Hz). FITGAUSSIAN returns the parameters of a gaussian function in the form defined at line 11 of Algorithm 1, where a is the amplitude, b is the spread and c is the center of the gaussian curve.

Different from Krebs (1997), we adapted FITGAUSSIAN, so that b is the only

---

Algorithm 1 – Submovement decomposition algorithm. Adapted from Krebs (1997).

Input: $\mathbf{x}$ and $\mathbf{y}$ ▷ From a raw movement file
Output: $\mathbf{SM}$ ▷ Extracted submovements and their properties

```

1: function SUBMOVEMENTS( $\mathbf{x}, \mathbf{y}$ ) ▷  $\mathbf{x}$  and  $\mathbf{y}$  from a raw file
2:    $\mathbf{SM} \leftarrow \{\}$ 
3:    $\mathbf{v} \leftarrow \sqrt{\dot{\mathbf{x}}^2 + \dot{\mathbf{y}}^2}$  ▷ Speed profile (meters per second)
4:    $v_{peak} \leftarrow \max(\mathbf{v})$  ▷ Global peak speed
5:    $sm_{peak} \leftarrow v_{peak}$  ▷ Local peak speed
6:    $l^* \leftarrow \underset{l}{\operatorname{argmax}}(v_l)$ 
7:    $\mathbf{t} \leftarrow 0..(|\mathbf{v}| - 1)\Delta t$ 
8:   while  $sm_{peak} > 0.05 \cdot v_{peak}$  do
9:      $r_0, r \leftarrow \text{SUBMOVEMENTRANGE}(\mathbf{v}, l^*)$  ▷  $r_0 \leq l^* \leq r$ 
10:     $a, b, c \leftarrow \text{FITGAUSSIAN}(\{t_i\}_{i=r_0}^r, \{v_i\}_{i=r_0}^r)$ 
11:     $\mathbf{s} \leftarrow a \exp\left(\frac{-(\mathbf{t} - c)^2}{b^2}\right)$  ▷ Submovement from fitted gaussian
12:     $\mathbf{i} \leftarrow \{i | s_i > 0.01 \cdot v_{peak}\}$  ▷ Array of indices
13:    if  $|\mathbf{i}| = 0$  then ▷ If submovement's peak is less than 1% of global peak
14:      break ▷ All submovements have been caught. Halt.
15:    end if
16:     $sm_{begin} \leftarrow \min(\mathbf{i})$  ▷ Index
17:     $sm_{end} \leftarrow \max(\mathbf{i})$  ▷ Index
18:     $sm_{peakttime} \leftarrow (l^* - 1)\Delta t$  ▷ Time offset (seconds)
19:     $sm_{duration} \leftarrow (m_{end} - m_{begin})\Delta t$  ▷ Interval length (seconds)
20:    if  $sm_{duration} > 0.05 \ \& \ 0 < r_0, r < |\mathbf{v}|$  then ▷ Filtering valid submovements
21:       $\mathbf{sm} \leftarrow \{sm_{begin}, sm_{end}, sm_{peakttime}, sm_{peak}, sm_{duration}\}$ 
22:       $\text{APPENDTO}(\mathbf{SM}, \mathbf{sm})$  ▷ Appending submovement  $\mathbf{sm}$  to  $\mathbf{SM}$ 
23:    end if
24:     $\mathbf{v} \leftarrow \mathbf{v} - \mathbf{s}$  ▷ Removing current submovement from speed profile
25:     $sm_{peak} \leftarrow \max(\mathbf{v})$  ▷ Peak speed of next iteration's submovement
26:     $l^* \leftarrow \underset{l}{\operatorname{argmax}}(v_l)$ 
27:  end while
28:  return  $\text{SORT}(\mathbf{SM})$  ▷ Sorting list of submovements by  $sm_{peakttime}$ 
29: end function

```

---

parameter to be tuned, whereas a and c remain fixed as sm_{peak} and $l^*\Delta t$, respectively. Through gradient-descent optimization, we considered the quadratic error function as the cost function. At each iteration, b was adjusted as follows:

$$g(a, b, c, t) = a \exp\left(\frac{-(t - c)^2}{b^2}\right) \quad \text{Algorithm 1, line 11} \quad (5.12)$$

$$E = (y - g)^2 \quad \text{Cost function} \quad (5.13)$$

$$b(t' + 1) = b(t') - \eta \left(\frac{\partial E}{\partial b}\right) \quad \text{Adjustment at iteration } t' + 1 \quad (5.14)$$

where y is the expected output from the speed profile within $[r_0, r]$, η is the size of the

Figure 25 – Iterative process described in [Algorithm 1](#). The green curve indicates the current fitted gaussian function $\mathbf{s}$, given sm_{peak} (star mark) and the points within r_0 and r range (left and right X marks, respectively). The solid blue curve is the speed profile $\mathbf{v}$, whereas dotted blue is $\mathbf{v} - \mathbf{s}$, i.e., dotted blue curve from the i -th iteration is the solid blue curve of the $(i + 1)$ -th iteration. The red curve indicates an invalid submovement.

Source: Elaborated by the author.

gradient step and $\frac{\partial E}{\partial b}$ is obtained as follows:

$$\begin{aligned}
 \frac{\partial E}{\partial b} &= (y - g) \frac{\partial g}{\partial b} \\
 \frac{\partial E}{\partial b} &= (y - g) \frac{2a \exp\left(\frac{-(t - c)^2}{b^2}\right) (-c + t)^2}{b^3} \\
 \frac{\partial E}{\partial b} &= (y - g) \frac{2g (-c + t)^2}{b^3}
 \end{aligned} \tag{5.15}$$

The adjustments in b take place for at most 10^5 iterations, or when E is lower than $\varepsilon = 10^{-30}$. In addition, b is randomly initialized within the range $[0, 0.25]$ and $\eta = 1$.

After extracting the properties of the current submovement (lines 16 to 19), they are stored in case the submovement is valid and lasts more than 50 milliseconds. The fitted curve $\mathbf{s}$ must then be subtracted from $\mathbf{v}$, so the next submovement's peak and location can be updated (lines 25 and 26) for the next iteration. The iterative process halts when the current sm_{peak} is less than 1% of v_{peak} (line 13). Once all submovements have been found and stored in $\mathbf{SM}$ array (Figure 26a), they are sorted by $sm_{peaktime}$ (Figure 26b) so the micrometrics can now be calculated.

Although there are more accurate ways to decompose point-to-point movements into submovements (ROHRER *et al.*, 2004), such approaches are computationally more expensive, which may be a burden as the data volume increases. The greedy approach here considered becomes convenient on that matter, also delivering equivalent trends from statistical perspective (as reported in Chapter 7).

5.4.1 Submovement number

Given the extracted submovements in the array $\mathbf{SM}$, submovement number consists in its length:

$$sm_{number} = |\mathbf{SM}| \quad (5.16)$$

Figure 26 – Submovements from the iterative process illustrated in Figure 25 in (a) the order they were extracted and (b) sorted by $sm_{peaktime}$ (line 28).

Source: Elaborated by the author.

5.4.2 Submovement duration

As defined in line 19 of Algorithm 1, submovement duration is defined as:

$$sm_{duration} = \frac{1}{sm_{number}} \sum_{i=1}^{sm_{number}} sm_{i_{duration}} \quad (5.17)$$

where $sm_{i_{duration}}$ is the *duration* property of the i -th submovement and $sm_{duration}$ aggregates all submovements by averaging such property.

5.4.3 Submovement peak

As defined in line 25 of Algorithm 1, submovement peak is defined as:

$$sm_{peak} = \frac{1}{sm_{number}} \sum_{i=1}^{sm_{number}} sm_{i_{peak}} \quad (5.18)$$

where $sm_{i_{peak}}$ is the *peak* property of the i -th submovement and sm_{peak} aggregates all submovements by averaging such property.

5.4.4 Submovement overlap

Overlap concerns the time length that two submovements overlap each other, as illustrated in Figure 27:

Figure 27 – Submovement overlap rationale. The time length that the 4th submovement overlaps the third is defined as $sm_{4_{overlap}}$.

Source: Elaborated by the author.

Described in Algorithm 2, the overlap between i -th and $(i - 1)$ -th submovements is calculated for all submovements and averaged in $sm_{overlap}$:

Algorithm 2 – Submovement overlap calculation**Input:** SM ▷ List of submovements from SUBMOVEMENTS**Output:** $sm_{overlap}$

```

1: function OVERLAP(SM)
2:    $sm_{overlap} \leftarrow 0$ 
3:   if |SM| > 1 then
4:     for  $i$  in 2..|SM| do
5:        $sm_{i_{overlap}} \leftarrow 0$ 
6:       if  $sm_{i_{begin}} < sm_{i-1_{end}}$  then
7:          $sm_{i_{overlap}} \leftarrow \min(sm_{i-1_{end}} - sm_{i_{begin}}, sm_{i-1_{end}} - sm_{i-1_{begin}})\Delta t$ 
8:       end if
9:     end for
10:     $sm_{overlap} \leftarrow \frac{1}{|SM|-1} \sum_{i=2}^{|SM|} sm_{i_{overlap}}$ 
11:  end if
12:  return  $sm_{overlap}$ 
13: end function

```

If the i -th submovement overlaps the $(i-1)$ -th completely, or even beyond, $sm_{i_{overlap}}$ saturates at $sm_{i-1_{duration}}$, i.e., the $(i-1)$ -th submovement duration (line 7). If there is no overlapping between i -th and $(i-1)$ -th submovements, then $sm_{i_{overlap}} = 0$ (line 5).

5.4.5 Submovement interpeak

Interpeak concerns the time-offset difference between two submovement peaks, as illustrated in Figure 28:

Figure 28 – Submovement interpeak rationale. The time-offset difference between 4th and 3rd submovement peaks, defined as $sm_{4_{interpeak}}$.

Source: Elaborated by the author.

Described in [Algorithm 3](#), the interpeak between i -th and $(i - 1)$ -th submovements is calculated for all submovements and averaged in $sm_{interpeak}$:

Algorithm 3 – Submovement interpeak calculation

Input: $\mathbf{SM}$ $\triangleright$ List of submovements from SUBMOVEMENTS

Output: $sm_{interpeak}$

```

1: function INTERPEAK( $\mathbf{SM}$ )
2:    $sm_{interpeak} \leftarrow 0$ 
3:   if  $|\mathbf{SM}| > 1$  then
4:      $sm_{interpeak} \leftarrow \{sm_{i_{peaktime}} - sm_{i-1_{peaktime}} \mid i = 2..|\mathbf{SM}|\}$ 
5:      $sm_{interpeak} \leftarrow \frac{1}{|\mathbf{SM}|-1} \sum_{i=2}^{|\mathbf{SM}|} sm_{i_{interpeak}}$ 
6:   end if
7:   return  $sm_{interpeak}$ 
8: end function

```

DEFINITION OF THE HEMIPARESIS-INHERENT DATA SPACE TOWARDS PATIENT'S PROGRESS ASSESSMENT

This chapter presents a KDD strategy based on the model proposed by [Fayyad, Piatetsky-shapiro and Smyth \(1996\)](#). A preliminary goal is the definition of a data space, whose components, arising from kinematic and dynamic measures, are inherent to hemiparesis. Since outcomes from data analyses are sensitive to the underlying representativity of the involved samples, the idea is to focus on the method itself, although an instance of analysis is also here elucidated. We consider such data space as the subject of study for [Section 6.7](#), regarding time dependency issues to be addressed for the definition of a performance measure.

For every KDD strategy, as elucidated in [Section 2.1.1](#), data must be collected, so a target set is defined as the actual input of the analysis process. In this context, data arise from the efforts of patients during physical therapy sessions, where a robotic device, providing motor assistance when needed, also senses and collect the data that properly describe motor behavior.

Two InMotion Arm^{TM} robotic devices (the clinical version of MIT-MANUS ([HOGAN et al., 1995](#); [KREBS et al., 1998](#)) from Interactive Motion Technologies¹, illustrated in [Figure 29a](#)) were used for the interaction with patients, as well as the sensing and collection of data. In the course of rehabilitation sessions, patients were oriented by a virtual environment, in which interaction takes place by manipulating the end-effector of the device.

---

¹ More details available at <https://www.bioniklabs.com/products/inmotion-arm>. Accessed: 1 jan. 2021.

When the robotic device is in operation, at every sensory acquisition event for data collection, the following measures are yielded in relation to the end-effector: position (meters) in x and y axes; velocity (meters per second) in x and y axes; force (newtons) in x , y and z axes. As elucidated in Table 1, the acquisition of these measures occur at every five milliseconds (200Hz), along with a time offset (tick), which is also collected.

For this work, the virtual environment displayed in rehabilitation sessions is called Clock game, as depicted in Figure 29b. Displaying a clock-like circle with eight black targets arranged around the edge, and a central target, the goal is to perform reaching movements by moving the yellow circle, which represents the patient's hand (the end-effector), towards the current target to be reached, indicated in red. After a peripheral target is reached, the same movement is performed backwards, reaching center. The distance from the central to any peripheral target measures 14cm. Beginning from "twelve o'clock" position and proceeding clockwise, a star-pattern is traced, having the flow of this cycle of movements indicated in the right side of Figure 29b.

The cycle of movements is repeated several times in the course of rehabilitation sessions. At the end of a session, the aforementioned measures are stored, such that each reaching movement (towards a target) is organized in a single text file. Figure 29c plots position data of one cycle of movements performed by a chronic patient. Black circles indicate the actual position of targets.

The InMotion Arm^{TM} devices (A and B) are located at Physical and Rehabilitation Medicine Institute (São Paulo, SP, Brazil) of the University of São Paulo, from Lucy Montoro Rehabilitation Network. These devices collected data from the motor performance of nineteen chronic stroke patients diagnosed with left or right hemiparesis (whose demographics are shown in Table 4), along treatments of upper-limb rehabilitation. No data regarding non-paretic sides were considered. It is important to point out that the identity of patients were preserved, complying to requirements of an ethics committee (Annex A).

Regarding structure of Table 4, Age column indicates the age (years) of patients at enrollment; Idle column indicates in months the period without treatment since the stroke event; Cycles_{Dev} column indicates the overall number of cycles of movements performed, as well as whether data were collected from device A, B or both.

Enhancing our previous results (MORETTI *et al.*, 2016), this strategy, aiming the definition of the hemiparesis-inherent data space, is organized in the following sections, similarly as established in the KDD roadmap (Figure 1).

Figure 29 – The robotic device used in physical therapy sessions. (a) InMotionArm^{TM} robotic device for the robot-aided rehabilitation (adapted from <http://imt.lynxdesign.com/wp-content/uploads/2012/11/InMotionARM240.jpg>). (b) At the left, the virtual environment (Clock game) used for rehabilitation exercises from which data were collected in this work; at the right, the flow of cycles of movements performed by patients. (c) Visualization of position data where a cycle of movements is performed by a chronic patient; black circles are the actual target positions.

Source: Elaborated by the author.

6.1 Domain understanding and definition of KDD goals

Although each side of the brain is associated to distinct functional abilities, and that paresis may entail different and characteristic symptoms, according to which cerebral hemisphere is injured (SPERRY, 1974), the goals for data analysis in this KDD strategy retains to the scope of motor control abilities, as kinematic and dynamic measures of upper limbs are the only available input. Issues of cognitive order do not concern this work.

For the definition of a data space, the KDD goal is to induce a model for the purpose of distinguishing hemiparetic sides, given descriptors of motor performance (inherent to

Table 4 – Demographics from the collected data.

Left hemiparesis					Right hemiparesis				
#	Gender	Age	Idle	Cycles _{Dev}	#	Gender	Age	Idle	Cycles _{Dev}
1	F	64	8	51 _{AB}	13	F	67	7	1 _A
2	F	48	28	1 _A	14	F	39	21	51 _{AB}
3	M	55	15	2 _A	15	M	65	13	40 _{AB}
4	M	59	12	42 _{AB}	16	M	58	20	36 _{AB}
5	M	69	12	10 _B	17	M	60	18	51 _{AB}
6	F	32	32	32 _B	18	M	49	11	43 _{AB}
7	F	33	13	43 _{AB}	19	F	61	28	55 _{AB}
8	F	74	21	51 _{AB}	-	-	-	-	-
9	M	69	19	15 _A	-	-	-	-	-
10	F	60	17	2 _A	-	-	-	-	-
11	M	48	12	48 _{AB}	-	-	-	-	-
12	M	61	7	37 _B	-	-	-	-	-
12 Patients: 6 Male and 6 Female Average age: 56 ± 13.5 Average idleness: 16.3 ± 7.6 334 Cycles					7 Patients: 4 Male and 3 Female Average age: 57 ± 9.8 Average idleness: 16.9 ± 7.1 277 Cycles				
19 Patients: 10 Male and 9 Female Average age: 56 ± 12 Average idleness: 16.5 ± 7.2 611 Cycles									

Source: Elaborated by the author.

the pathology) from the efforts of a patient.

6.2 Selection of the target set

At its raw form, the collection of data encompasses measures in text format. Describing single point-to-point movements, each text file is tabularly structured, where each line, acquired at 200Hz, is composed by the following space-separated measures: time offset (tick), positions (p_x and p_y), velocities (v_x and v_y) and forces (f_x , f_y and f_z). Tick was disregarded from this target set of data, as well as f_z , since only planar measures were considered for this analysis.

Different from our last strategy (MORETTI *et al.*, 2016), which considered only device A, data from both devices were merged for analysis, even taking into account potential wear of mechanical parts and uncertainty in sensing. Otherwise, it would lead to a biased analysis, as every data used to induce a model would arise from the only device involved in the experiment. Therefore, the proneness to such bias must be avoided, once performance data from patients typically arise from different sources, such as different devices and different rehabilitation centers. Every arrangement of data may raise a different

bias to be diminished or avoided. These issues are later discussed in [Section 6.6.4](#).

So far, the target set of data is composed by movement files (acquired from the efforts of the nineteen patients from [Table 4](#)), in which only p_x , p_y , v_x , v_y , f_x and f_y measures are considered for the preprocessing step.

6.3 Preprocessing

Since the KDD goal defined at [Section 6.1](#) designates a supervised learning problem, for the purpose of building an initial format of instances, portions of the target set of data must be aggregated, so that these can be labeled as left or right hemiparesis.

Observing patterns in the repetitions of movements, we decided to consider instances as the motor behavior of patients along cycles of movements, as depicted in [Figure 29b](#), regardless of whether some instances belong to the same session or not. So far, each instance consists of a concatenation of sixteen files (eight targets, two movements per target: towards and backwards) in the same order they were acquired, so an unbroken cohesive trajectory, in terms of positions, velocities and forces, is preserved. An example of instance, considering only position data for visualization, is plotted in [Figure 29c](#).

One can observe that each patient has a particular condition to perform the same exercise. That being said, movements take different time intervals until they are finished, and hence the length of data instances become variable. However, such data format is not suitable for further machine learning techniques, since dimensionality is expected to be fixed. The next KDD step addresses this issue by extracting the due descriptors (features), allowing to build an actual dataset.

6.4 Feature extraction

Not only because of the variable length of instances, feature extraction must also take place because of the precision in data acquisition. It is known that six measures are collected 200 times every second. Considering, in average, that single movements last seconds, and that cycles of movements last sixteen times more, data instances naturally take large proportions, which is not suitable for machine learning techniques as well, due to the aforementioned curse of dimensionality.

One can understand *inherence* of extracted features as their proneness to present separability of instances in space. In this case, from a set of features inherent to hemiparesis, instances of each paretic side, by coordinates in the defined dimensions, are spatially gathered, such that it is possible to define boundaries for the purpose of shattering regions associated to each hemiparetic side.

Based on descriptive statistics, our strategy (MORETTI *et al.*, 2016), given an instance whose format arises from preprocessing step, consists of calculating for each measure (elucidated in Section 6.2) the four statistical moments of the distribution: mean ($\bar{X}$), standard deviation (σ), skewness (γ) and kurtosis (κ). This strategy yields a raw set of 24 features. In practical terms, so far, cycles of movements are now represented as points in a space of 24 dimensions.

$$D^{\text{raw}} = \left\{ \begin{array}{cccccc} \bar{p}_x, & \bar{p}_y, & \bar{v}_x, & \bar{v}_y, & \bar{f}_x, & \bar{f}_y, \\ \sigma_{p_x}, & \sigma_{p_y}, & \sigma_{v_x}, & \sigma_{v_y}, & \sigma_{f_x}, & \sigma_{f_y}, \\ \gamma_{p_x}, & \gamma_{p_y}, & \gamma_{v_x}, & \gamma_{v_y}, & \gamma_{f_x}, & \gamma_{f_y}, \\ \kappa_{p_x}, & \kappa_{p_y}, & \kappa_{v_x}, & \kappa_{v_y}, & \kappa_{f_x}, & \kappa_{f_y} \end{array} \right\} \quad (6.1)$$

Although dimensionality from these features are greatly lower in relation to the previous format, they are still considered raw, for the purpose of attempting to reduce such set through feature selection. Out of 276 different two-dimensional subspaces, from a not-so-thorough visual inspection, we previously perceived a proneness to separability in some cases (MORETTI *et al.*, 2016), as illustrates Figure 30, which involves data from both devices.

Figure 30 – Visual inspection of subspaces where the proneness to separability is more evident.

(a) $\bar{f}_x \times \sigma_{f_y}$ subspace. (b) $\gamma_{f_x} \times \gamma_{f_y}$ subspace.

Source: Elaborated by the author.

In pilot essays (MORETTI *et al.*, 2014; MORETTI *et al.*, 2016) and also from Figure 30, it can be observed that force attributes are more inherent to hemiparesis. It is worth to recall that each point in these subspaces is a descriptor of the motor behavior of a patient who performed a cycle of movements (see Figures 29b and 29c). Yellow and blue points are respectively cycles from patients diagnosed with left and right hemiparesis.

In some subspaces arising from (6.1), specially in $\bar{f}_x \times \sigma_{f_y}$ and $\gamma_{f_x} \times \gamma_{f_y}$, despite the potential separability, one can observe an intersection between both classes, which can be interpreted as uncertainty regions, where no decision boundaries would sharply shatter the due regions for a neat classification. Thus, it is necessary to consider more dimensions for the proper definition of the space for further steps.

For the sake of interpretability (see Section 2.1.2), we decided not to consider projection-based techniques for dimensionality reduction, such as PCA and LDA, over the twenty-four-dimensional space (6.1). Feature selection took place instead, along with data mining step, so that a combination of features may be attained for a good separability between classes.

6.5 Data mining

Following our previous findings (MORETTI *et al.*, 2016), this step encompasses the means for the definition of the final set of features that compose a data space, as well as the involved machine learning algorithms to perform the distinction between hemiparetic sides.

6.5.1 Definition of techniques

According to the KDD task established in Section 6.1, and given data instances associated to its respective labels (the hemiparetic sides, arising from patient conditions), the distinction between hemiparetic sides is given by a classification task.

From a grid search in (MORETTI *et al.*, 2016), ranges of values were assigned to the parameters of classification algorithms, such as k-Nearest Neighbors, J48 Decision Tree, Random Forest and Multi-Layer Perceptron (MLP) with backpropagation training algorithm. An MLP was the best candidate, which fitted data with an accuracy of 94%. Despite the considerable (at first) outcome, our previous findings lie in the number of hyperplanes. Such accuracy was obtained with a single hyperplane, which is a strong evidence of linear separability. Therefore, we will retain $\mathcal{F}$ to a restricted bias, such that the selection of algorithms in the next section may only yield linear classifiers.

6.5.2 Machine learning and metaheuristics

Before submitting data to actual classification, a final descriptor (with less than 24 features) must be defined. Based on Genetic Algorithms (GA), we aim to yield a combination of features which culminates in high classification accuracies. From D^{raw} , defined in (6.1), a binary structure (the *chromosome*) of the same size is defined for the GA, in which values 1 and 0 indicate, respectively, to exclude or retain features. Each

position in the binary structure refers to the respective position of a feature in D^{raw} . The implementation of the GA used in our method is from *genalg* R package.

Table 5 – Grid search parameter configuration.

Algorithm	Parameter	Range	Steps
MLP (RProp)	Error function	SSE, CE	—
SVM	kernel	linear, poly	—
	gamma	[0.06, 0.125]	14
	coef0	[0, 1]	2
	degree	[2, 3]	2
	cost	[0.333, 1.6]	20

Source: Elaborated by the author.

The fitness of each chromosome is evaluated by the accuracy of an SVM classifier. A grid search (Table 5) was performed on SVM parameters over the twenty-four dimensional space, so that the best configuration will be used to induce a model at each evaluation of a chromosome. For comparison, still retaining $\mathcal{F}$ to a restricted bias, we also considered an MLP with Resilient Propagation (RProp) training algorithm, whose topological configuration consists of a single hyperplane shattering the space. Both SVM and RProp algorithms used in our method are respectively from *e1071* and *neuralnet* R packages. Parameters and technical data from Table 5 concern such packages.

Algorithm 4 – Fitness function

```

1: function FITNESS( $\mathbf{x}, dim$ )
2:   if  $dim > -1$  and  $\langle \mathbf{1} - \mathbf{x}, \mathbf{1} \rangle \geq dim$  then
3:     return 0
4:   end if
5:    $D \leftarrow \{D_i^{\text{raw}} : x_i = 0\}$  ▷ Selecting features according to  $\mathbf{x}$ 
6:   return PATIENTCROSSVALIDATIONSVM( $D$ ) ▷ Accuracy from cross-validation
7: end function

```

The best configuration was used to build the fitness function, which is defined in Algorithm 4. We found traditional 10-fold cross-validation inconsistent as a measure of the generalization ability of a classifier for this problem. A random sampling would take instances from different patients at every fold, leaving instances of these patients both in training and test sets. Such sampling gives bias to the induced classifier, which learns about a patient condition that is also presented in test step, entailing high accuracy that gives a false notion of generalization. Instead, each fold $\mathbf{X}^p$ comprise all instances from each patient p (the number of folds is equal to the number of patients), so that its size $|\mathbf{X}^p|$ weights its accuracy in classification, having a weighted average yielding the final

accuracy:

$$\text{PATIENTCROSSVALIDATIONSVM} = \frac{1}{|\mathbf{X}|} \sum_{p \in P} |\mathbf{X}^p| \cdot f(\mathbf{X}^p) \quad (6.2)$$

where $\mathbf{X}$ is the entire dataset and f is the SVM classifier, induced with training data $\mathbf{X} \setminus \mathbf{X}^p$, which yields the final accuracy in the classification of instances from fold $\mathbf{X}^p$. This approach, using SVM classifiers, is referred to as `PATIENTCROSSVALIDATIONSVM` in [Algorithm 4](#).

Yielding the chromosome with the highest fitness, and referred to as `GENETICALGORITHM`, an elitist GA is defined to run along 100 generations with a population size of 150 and mutation probability of 10%.

[Algorithm 5](#) consists of suggesting chromosomes for the initial population of GAs. Such suggested chromosomes are variations of the best solution found on a previous execution. Given a threshold of *loss* (in accuracy) on the first execution of `FEATURESELECTION`, it is verified (at [line 4](#)) whether the other parameters were provided, which is always false at this point. In this case, a GA is executed with no suggestions of chromosomes, defining in $\mathbf{x}$ an initial combination of features to be improved. Henceforth, as also the case of the subsequent executions, suggestions are provided as variations of $\mathbf{x}$, with one dimension less. The *highest* fitness is kept as reference regardless of whether it arises from the first execution. Then, the fitness function starts to constraint (at [line 2](#) of [Algorithm 4](#)) the dimensionality of feature combinations, yielded on subsequent executions of the GA, so that accuracies are considered as non-zero only if dimensionality is lower. If the accuracy (*current*) of the next found combination $\mathbf{x}(t+1)$ diverges (from *highest*) less than the admissible *loss*, this process continues recursively until such divergence is no longer admissible.

6.6 KDD results and interpretation

This section encompasses the final results of the KDD strategy for the definition of a data space, as well as interpretation of the obtained results and its confrontation with SLT is also discussed.

6.6.1 Model configuration for fitness function

Within the ranges from [Table 5](#), the best configuration of an SVM model fitted data in the twenty-four dimensional space with an accuracy of 76.6% (F1 score of 0.74). For comparison, the best configuration of an MLP model presented an accuracy of 70.4%. The SVM parameter configuration consists of a *linear kernel* and *cost* = 0.87.

Algorithm 5 – Recursive population initialization for feature selection with genetic algorithms

---

```

1: function FEATURESELECTION(loss,x,highest)           ▷ x and highest are optional
2:   S ← {}
3:   dim ← -1
4:   if |x| = 0 then                                     ▷ Default for x is {}
5:     x ← GENETICALGORITHM(dim,S)
6:   end if
7:   for each i, such that xi = 0 do
8:     x' ← x
9:     x'i ← 1
10:    append(x', S)                                     ▷ Appending x' to S
11:  end for
12:  current ← FITNESS(x,dim)                             ▷ From Algorithm 4
13:  if highest = 0 then                                   ▷ Default for highest is 0
14:    highest ← current
15:  end if
16:  dim ← ⟨1 - x, 1⟩                                     ▷ Attempting to fetch less than dim features
17:  x(t + 1) ← GENETICALGORITHM(dim,S)
18:  current ← FITNESS(x(t + 1),dim)                     ▷ From Algorithm 4
19:  if current > highest then
20:    highest ← current
21:  end if
22:  if highest - current ≥ loss then return x           ▷ Ceasing recursion
23:  else
24:    x(t + 1) ← FEATURESELECTION(loss,x(t + 1),highest) ▷ Recursion
25:  end if
26:  return x(t + 1)
27: end function

```

Substantial guarantees of better performance from SVM, in regard to the best separability, are only valid if the feature space is reconstructed with the ideal kernel. Nevertheless, the strong evidences of linear separability raised in (MORETTI *et al.*, 2016) led to the use of the SVM algorithm. As can be seen in Table 5, no higher orders (given by *degree* parameter) of polynomial kernels were considered for the grid search, even culminating in higher accuracies, since high-dimensional spaces entail high shattering coefficients, and hence weaker guarantees of actual learning ability.

6.6.2 Final descriptor for the data space

Admitting at most 2% of loss in accuracy, features from D^{raw} were selected through Algorithm 5, yielding a twelve-dimensional descriptor:

$$D = \left\{ \bar{p}_x, \bar{f}_x, \bar{f}_y, \sigma_{v_y}, \sigma_{f_x}, \sigma_{f_y}, \gamma_{v_x}, \gamma_{v_y}, \gamma_{f_x}, \gamma_{f_y}, \kappa_{v_x}, \kappa_{f_x} \right\} \quad (6.3)$$

In the space defined by D , instances from the available dataset (of patients from Table 4) were classified with an accuracy of 81%.

Figure 31 – Accuracies and dimensionalities of reduced sets of features along the feature selection strategy of Algorithm 5.

Source: Elaborated by the author.

The results along feature selection process are shown in Figure 31. The curve indicates the *current* accuracy from the descriptor found in the t^{th} execution of FEATURESELECTION (line 18), whereas circles quantify the dimensionality of such descriptor. It can be observed that, along the process, a higher accuracy reference was obtained (in this case when $t = 2$, due to line 19), as well as a significant decrease in dimensionality. Such improvement can be explained by the suggestions provided before the execution of GENETICALGORITHM, taking place only when $t \geq 2$, in which variations of the previous solution (the suggestions) may be contiguous to local minima of the search space.

6.6.3 Learning guarantees from SLT

Due to the lack of closely-related work, as well as discrepancies with regard to the underlying representativity of the collected data (given by sample size, criteria for selection of patients, etc.), we decided to confront our results with foundations of the Statistical Learning Theory. Generalization Bound (defined in (2.40)) provides an upper bound on the true risk $R(f)$, given δ , as the dataset size n (number of cycles of movements) tends to infinity. With $R_{\text{emp}}(f) = 0.19$ and $\delta = 0.05$, i.e., with a probability of 95% of the divergence between empirical and expected risks not exceed ε (now as function of δ), the worst case of $R(f)$ is upper-bounded by the right hand side of (2.40) as a function of the sample size n , illustrated in Figure 32.

Since it is not trivial to obtain the shattering coefficient $\mathcal{N}(\mathcal{F}, 2n)$ analytically, in practical terms, we plugged an approximation based on estimates of possible classifications

Figure 32 – Generalization bound of the SVM classifier f , with $\delta = 0.05$ and $R_{\text{emp}}(f) = 0.19$, indicating the bound with respect to the sample from Table 4.

Source: Elaborated by the author.

of n points, with n ranging from 1 to 30. For all $n > 13$, $\mathcal{N}(\mathcal{F}, 2n)$ grows polynomially, having this growth behavior approximated in the form of a quadratic polynomial.

Such approximation was yielded as follows: let $d = 12$ be the dimensionality of the data space. For each n , ranging from 1 to 30, n instances in $\mathbb{R}^d$ are randomly generated arising from a normal distribution of mean 0 ± 1 . For each i^{th} out of 10000 iterations (still for each n), a hyperplane is randomly placed in space in terms of $\mathbf{w}$ and b , such that $|\mathbf{w}| = d$. The inner product between every random instance and the hyperplane classifies positively (1), for $\langle \mathbf{w}, \mathbf{x} \rangle + b \geq 0$, and negatively (-1) for $\langle \mathbf{w}, \mathbf{x} \rangle + b < 0$, so that a combination of classifications for all instances is stored. At the end of the 10000 iterations, the the total of different combinations of outputs is associated to n . This process repeats for the next values of n .

Figure 33 – Estimation of $\mathcal{N}(\mathcal{F}, 2n)$ via coefficients of a quadratic polynomial.

Source: Elaborated by the author.

So far, a discrete estimated behavior of $\mathcal{N}(\mathcal{F}, 2n)$ is given by every number of classifications associated to each n , as can be seen by blue points in Figure 33. A quadratic

polynomial is then fit to this behavior (the red curve of Figure 33), so that $\mathcal{N}(\mathcal{F}, 2n)$ can be used to obtain the generalization bound curve from Figure 32. A *warm-up* period of 9 was considered, so only points from $n \geq 10$ were considered to fit the curve.

From another perspective, defining an admissible divergence of (at most) $\varepsilon = 0.05$ between empirical and expected risks, Figure 34 shows the upper bound of Chernoff/Hoeffding inequality (2.18), plugging $\mathcal{N}(\mathcal{F}, 2n)$ in m , as a function of n . As can be seen more thoroughly in Figure 34b, it is necessary at least, say, 6000 instances drawn iid for the probability of the divergence between $R_{\text{emp}}(f)$ and $R(f)$ to exceed ε to be less than 0.1. In practical terms, this means that with a probability less than 0.1, the accuracy of our classifier, induced with, say, 6000 iid instances, would have an error of at most 5% in relation to the expected value of the accuracy of the same classifier.

Figure 34 – Upper bound of Chernoff/Hoeffding inequality as a function of the sample size. (a) shows the curve and points out the size of the available sample (Table 4). (b) shows the same curve within the range of interest.

Source: Elaborated by the author.

It is important to mention that SLT provides guarantees of learning once some grounding conditions (enumerated in the beginning of Section 2.2) are considered, as well as the outcomes properly raise tighter bounds of risks, entailing generalization. However, such conditions are not necessary for learning to occur. It is also worth to point out that every bound from SLT are related to the worse case within $\mathcal{F}$.

From the perspectives illustrated by Figures 32 and 34, one may conclude that more iid instances are necessary, so that tighter bounds of risks and higher guarantees of learning can be attained. A larger sample may also influence in the definition of D , and hence the data space itself.

Regarding the generalization bound of the obtained classifier f , it can be observed that such bound attenuates its variation along n from a relatively low number of instances (say, $n = 2000$). Being sensitive to the shattering coefficient, the capacity term (the square root term) in (2.40) produces this behavior, whereas $R_{\text{emp}}(f)$ shifts the curve to converge to higher or lower bounds. Since instances are expected to be drawn iid, a larger sample

would bring more representativity, so that the step of [Section 6.5](#) may culminate in higher accuracies. A classifier with a higher accuracy would then have a tighter generalization bound, since its empirical risk is lower.

In order to preserve representativity, it is also important to consider biases (in sampling) to be avoided in the procurement of instances that constitute the whole dataset, especially when the sample takes large proportions. Such issue is discussed in the next section.

In comparison to the related work from [Chapter 3](#), from SLT perspective, the bias $\mathcal{F}$ of induced models following our strategy is more strict. In ([JUNG; GLASGOW; SCOTT, 2008](#)), for example, there are sixteen neural network models with five hidden nodes each, feeding an ensemble network. Although these models are in three-dimensional spaces, the bias is weaker due to the five hyperplanes, causing the shattering coefficient to grow more abruptly and hence attenuate the convergence of a generalization bound. The same case of models with higher complexity occur in ([PATEL *et al.*, 2010](#)), in which random forest is used also for ensembling, as well as in ([DIJCK; VAERENBERGH; HULLE, 2009](#)) and ([BOSECKER *et al.*, 2010](#)).

6.6.4 Avoidance of bias in sampling

In order to give the due relevance to the obtained results, it is important to point out possible biases to be avoided. We raised three categories of bias in sampling that are necessary to be diminished, arising from patients, robotic devices and implementations of the treatment protocol.

As mentioned before in the context of cross-validation, patient bias concerns the false notion of generalization when instances from the same patient appear in both training and test steps. It entails higher accuracies even in cross-validation approaches, because the induced model already learned about the conditions of particular patients. In order to diminish such bias, classifiers must be tested with unseen (in training) patients (the set of their associated instances), rather than unseen instances regardless of the associated patient.

Device bias is related to the sensing ability of a particular robotic device. Due to mechanical wear of the parts over time, uncertainty in sensing becomes a core concern to be dealt with, once data arising from devices of different (and discrepant) usage time must be merged for overall analyses. Devices with major wear (i.e., prone to repair) must not be considered.

Protocol bias addresses strictness in the treatment protocol implemented at a particular rehabilitation center. It concerns whether any unrelated activity (be motor or cognitive), in the course of the exercises, affects the motor behavior being sensed during

sessions.

From the available dataset, patient bias issues were addressed with the PATIENTCROSSVALIDATIONSVM function from Algorithm 5, defined in (6.2). Device bias was attenuated by considering instances from both robotic devices (A and B). Since the available dataset was originated from a single rehabilitation center, it was not possible to address protocol bias.

In this work we attained an accuracy of 81% (F1 score of 0.79) with twelve features, having biases from patients and devices addressed. Considering only device bias to be attenuated, such accuracy goes to 91% with fourteen features, with a 10-fold cross-validation instead of the patient-wise approach. In our previous results (MORETTI *et al.*, 2016), in comparison, although attaining 94% of accuracy with eight features, no concerns regarding any of the aforementioned biases were taken into account.

6.6.5 Testing on data with different protocol bias

Additionally, in an exploratory fashion, we also tested the SVM model with the data from the study described in Chapter 7. In summary, such data comprise the efforts of 82 chronic stroke patients with right hemiparesis. More details can be found in Figure 39 and Table 8.

The same process on the dataset from Table 4 was conducted to prepare this dataset for testing. It carries 5304 cycles of movements described in terms of D^{raw} . Before feature selection (considering D^{raw} , Equation 6.1), the model yielded an F1 score of 0.928 (86.6% accuracy), whereas after feature selection (considering descriptor D), the model yielded an F1 score of 0.929 (86.8% accuracy).

Although this is not a proper way to evaluate the model, since there is only one of the two classes, we here ponder the reported results from an exploratory perspective. That being said, it is noteworthy that right hemiparesis is the minority class in the dataset used to train the model (Table 4), which makes the model less prone to be biased when outputting this classification.

It is also key to interpret and ponder these results from clinical perspective. One must take into account that different regions of the brain affected by stroke entail different motor deficits, and hence different patterns of recovery, especially when it comes to lesions on the left hemisphere of the brain (right hemiparesis), which plays a key role on the motor function for both ipsilesional and contralesional upper limbs (MUTHA; HAALAND; SAINBURG, 2012).

6.7 Progress assessment based on uncertainty in classification

This section presents initial considerations regarding the data space defined along previous sections. The discussion then focuses on the strategy we developed so far, as well as on the results for progress assessment of patients.

6.7.1 State of rehabilitation

Inspecting the defined data space, as well as its subspaces (such as the ones from Figure 30), one can observe an intersection between instances of different classes, which leads to uncertainty in classification and may explain the attenuated accuracy obtained so far.

Efforts towards sharpening the transition between classes, along the defined data space, would bring no considerable benefits. In fact, we interpreted such nebulous intersection as the case when a patient is recovered enough, that it is no longer wise to assure regarding paresis impairments. We hereafter call this hypothetical region in space, given by the intersection of points from different classes, as the state of rehabilitation (SOR). More precisely, SOR is considered as the region within an ε -neighbourhood with respect to the decision boundary defined beforehand.

The definition of SOR is chained to the representativity of the available data, since the definition of the data space, where SOR is mapped, arises from data itself. Therefore, SOR must be pondered by the underlying accuracy with respect to the placement of the hyperplane in space by the SVM classification algorithm. In practical terms, the definition of SOR becomes feasible from a sample with a sufficiently diminished bias, used to induce a classifier, which must have reasonable learning guarantees from SLT as well.

6.7.2 Time dependency

In the context of the assessment of patient's progress, we associate recovery to when patient's performance, over time, would culminate in instances of cycles of movements spatially closer to the decision boundary. Such convergence in distance would substantiate the underlying semantics of the spatial representation of a state of rehabilitation.

Grouping instances by patients, as well as taking into account the time offset of each instance, there can be observed that time dependency takes place. The time offset indicates session date and the order in which each cycle of movements was performed. For such time series, we aim to associate the detection of improvements in the motor behavior to the study of the convergence of such instances towards the hyperplane. We tie this idea to the situation in which motor deficit symptoms attenuate with the exercise

of rehabilitation. As an example, Figure 35 illustrates the order in which 32 instances emerged over time from the efforts of a single patient.

Figure 35 – Order of emergence of instances from a single patient (#15, from Table 4) over time in $\bar{f}_x \times \sigma_{f_y}$ subspace. Black and gray points arose respectively from devices A and B. (a) Region in $\bar{f}_x \times \sigma_{f_y}$ to be enlarged in **b** and **c**. (b) The first 15 instances. (c) Instances from the 15th to the 32nd.

Source: Elaborated by the author.

6.7.2.1 Attendance and regularity of time series

Time-dependent variables, arising from particular processes, are expected to comprise values acquired at regularly spaced time intervals. In the context of rehabilitation, these intervals are not regular, especially regarding the convergence in distance of instances from the defined space. Instances may originate consecutively, when cycles of movements are performed in a single session. Actual spaces in time are periods between rehabilitation sessions on, say, a weekly basis.

Figure 36a illustrates the attendance of a patient (the same from Figure 35) along 107 days of treatment. Horizontal axis measures time (in days) and vertical axis measures how many cycles were performed in a session. This is the case of a poor time-dependent behavior, with a large gap of 71 days of idleness in rehabilitation. The performance assessment would be poor for two reasons in this case. First, due to the poor behavior described by data; second, due to the long time without physical activity, which would naturally postpone improvements in motor abilities. As can be seen in the simulation from Figure 36b, even when attendance is ideal, instances are still not equally spaced. This occurs due to the consecutive emergence of instances, arising from each session, having such amount measured by the vertical axis.

Figure 36 – Emergence of instances (vertical axis) over time (horizontal axis). Black and gray bars indicate instances which arose respectively on devices A and B. (a) Attendance of patient #15 in rehabilitation sessions. (b) Ideal attendance with random numbers of cycles per session.

Source: Elaborated by the author.

6.7.3 Results and discussion

Given a set of instances from a particular patient, for each instance $\mathbf{x}_i$, the i^{th} term of the time series is given by the absolute value of $\langle \mathbf{w}, \mathbf{x}_i \rangle + b$, i.e., the minimum distance to the hyperplane of the SVM classifier. It is important to point out that such hyperplane was defined considering every instance available from the dataset. Figure 37 illustrates a time series arising from patient #15.

Figure 37 – Outputs of the SVM model, for each instance from patient #15, over time. Blue curve is a linear model fitting the series.

Source: Elaborated by the author.

A linear regression was performed for each time series, as illustrated by the blue curve in Figure 37, so that its angular coefficient is considered as an initial evidence of the convergence towards SOR. Still regarding Figure 37, for example, the angular coefficient is positive, indicating a growth behavior, and hence there is no convergence towards SOR (for patient #15). No issues with respect to the attendance were addressed.

Figure 38 – Distribution of the angular coefficients from each time series related to patient performance. The mean is indicated in blue.

Source: Elaborated by the author.

In contrast, when the angular coefficient is negative, the model behavior decreases over time, which may be an evidence of convergence towards SOR. These coefficients were yielded from data of patients who performed more than 10 instances of cycles of

movements, according to the collected data. Otherwise, considering few instances, such as the case of patients #3 and #10 from Table 4, the slope of the linear model becomes excessively steep, which can be considered as noise, since the underlying representativity is poor. The remaining cases compose a total of 14 coefficients, whose distribution is indicated in Figure 38. For the ε -neighbourhood with respect to the hyperplane from the defined SVM model, alluding to SOR, we considered $\varepsilon = 0$.

For an overall analysis of the progress of patients, with respect to the convergence towards SOR, a left-tailed t-test was performed over the distribution of Figure 38, with 95% of confidence. The alternative hypothesis is defined as $\mu < 0$. Although the mean slope of the linear models is -0.0096 , from the available data, and a p-value of 0.27, the test failed to reject the null hypothesis $\mu \geq 0$.

In regard to the distribution from Figure 38, so far we can conclude that the available data is insufficient for consistent outcomes. Even if there were evidences, given by the t-test, of an overall convergence towards SOR, more instances from more patients (with regular attendance) must be involved for more representativity in the final distribution of the angular coefficients. In relation to the available data described in Table 4, as a comparison, it can be seen a discrepancy in Table 2 concerning the number of patients involved in the data from related work. In such cases, consequently, data present higher volume, which contributes to more expressive outcomes.

THE IMPACT OF TDCS PRECEDING INTENSIVE ARM TRAINING IN CHRONIC STROKE

Upper Extremity (UE) hemiparesis is the most common impairment, with the prevalence as high as 80% in the acute and 40% in the chronic phase post-stroke (LAWRENCE *et al.*, 2001). Patients often report dissatisfaction with their UE recovery and long-lasting, costly restrictions to participation in Activities of Daily Living (ADL) and quality of life (LAI *et al.*, 2002). More effective, individualized, and targeted therapy interventions are vital to improve patient outcomes and address the demand for rehabilitation (OVBIAGELE *et al.*, 2013; LOHSE; LANG; BOYD, 2014; BERNHARDT *et al.*, 2019).

Robotic therapy for UE rehabilitation has shown promising results, but the benefits do not extend to all patients at all stages of their recovery (LO *et al.*, 2010; MILOT *et al.*, 2014; MAZZOLENI *et al.*, 2013; HSIEH *et al.*, 2018). Several studies have attempted to augment the benefits of robotic therapy by combining interventions (BOLOGNINI *et al.*, 2011; EDWARDS *et al.*, 2019). Edwards *et al.* (2019) investigated the effectiveness of combining intensive robotic therapy for the UE with anodal Transcranial Direct Current Stimulation (tDCS) compared to sham-tDCS (groups *Robot_{tDCS}* and *Robot_{Sham}*, respectively), in a chronic stroke population. Results showed that thirty-six therapy sessions produced significant and clinically meaningful improvements in motor recovery (as measured by the Upper Extremity Fugl-Meyer Assessment (FMA)) of both the *Robot_{tDCS}* and *Robot_{Sham}* groups. However, the application of anodal tDCS did not confer further advantage as measured by clinical outcome measures Edwards *et al.* (2019).

Combined therapy studies using tDCS have reported variability in effectiveness, even when adopting similar methodology (EDWARDS *et al.*, 2019; GIACOBBE *et al.*, 2013; LEFEBVRE; LIEW, 2017; STRAUDI *et al.*, 2016). One reason for the lack of

reproducibility could be the subjectivity and insensitivity of clinical outcome measures (KREBS *et al.*, 2002; BOSECKER *et al.*, 2010; SEMRAU *et al.*, 2013). Literature in this field has advocated using kinematics to study motor recovery post-stroke (BERNHARDT *et al.*, 2019; SEMRAU *et al.*, 2013; DIPIETRO *et al.*, 2012; DUKELOW, 2017; SCOTT; DUKELOW, 2011; DUKELOW *et al.*, 2012; COLOMBO *et al.*, 2005; REINKENSMEYER; EMKEN; CRAMER, 2004). Studies to date suggest kinematics (collected during a robotic evaluation) can provide a standardized and objective measure of a patient’s motor control, correlating with well-known clinical measures (BOSECKER *et al.*, 2010; SEMRAU *et al.*, 2013; DIPIETRO *et al.*, 2012; DUKELOW, 2017; KREBS *et al.*, 2013), and have the potential to enhance our understanding of stroke recovery Bernhardt *et al.* (2019). Particularly, smoothness is a measure showing good potential for quantifying motor recovery in patients post-stroke Rohrer *et al.* (2002). There is no consensus on the best single approach to quantify smoothness, but most studies have built on the work by Flash and Hogan (1985), who define smoothness as a measure of jerk (DIPIETRO *et al.*, 2012; ROHRER *et al.*, 2002; FLASH; HOGAN, 1985). Subsequent studies explain changes in smoothness as a gradual blending of discrete submovements (ROHRER *et al.*, 2002; KREBS *et al.*, 1999).

Edwards *et al.* (2019) included, but did not report the results of, robotic evaluations which generated kinematic and kinetic data. Here we explore if there was a difference in kinematic or kinetic measures of motor control between patients in *Robot_{tDCS}* and *Robot_{Sham}*. Further analysis of kinematics investigates if the severity of motor impairment was a factor in determining if intervention outcomes can be optimized by targeting a subgroup of patients. Lastly we examine trends in tDCS treatment effects, so as to contribute to the understanding of UE motor recovery in the chronic post-stroke population.

7.1 Study Overview

Subjects were enrolled from two sites, Burke Neurological Institute (White Plains, NY, USA) and Feinstein Institute for Medical Research (Manhasset, NY, USA), to participate in a double-blind, sham-controlled, repeated measures study¹. The study was approved by the Institutional Review Board at each center (Burke Rehabilitation Hospital (BRC 426), MIT-COUHES (Committee on the Use of Humans as Experimental Subjects, 1203004976), and Feinstein Institute for Medical Research IRB (12-102B)), and participants provided written informed consent. Candidates were eligible if they presented with a single ischemic stroke (over six months) and right hemiparesis. The robotic intervention involved 1024 movement repetitions per session, alternating Shoulder/Elbow (S/E) and wrist-forearm robot therapy (using planar and wrist robots, respectively) on separate days,

¹ ClinicalTrials.gov Identifier: NCT03562663. More details available at <<https://www.clinicaltrials.gov/ct2/show/NCT03562663>>.

performed three times a week (for a total of 36 sessions). Following [Giacobbe et al. \(2013\)](#), from an M1/SO montage, delivering 2mA of anodal-tDCS to the affected hemisphere, both the $Robot_{tDCS}$ and $Robot_{sham}$ interventions were administered at rest for 20 minutes immediately prior to each robotic session. Baseline clinical and robotic evaluations were performed twice within 6 weeks prior to commencing the intervention (separated by at least one week to ensure a stable baseline), at completion of the 12-week intervention, and after six months post training. More details on the study design are presented in [Edwards et al. \(2019\)](#).

Of the 82 subjects enrolled, 76 were available for the discharge evaluation (an attrition rate of 7.32%, see the Consolidated Standards of Reporting Trials (CONSORT) diagram in [Figure 39](#)). One subject in the $Robot_{sham}$ and two in the $Robot_{tDCS}$ group experienced an unrelated illness, two subjects had corrupted robot data, and one subject was excluded when s/he underwent Botulinum Toxin A therapy. Seventy-two subjects were included in the follow-up analysis due to unrelated illness of one subject, death/illness of a spouse, and transportation issues.

It is noteworthy that there is a small discrepancy between the number of subjects included in the analysis reported in [Edwards et al. \(2019\)](#) and the number of subjects analyzed in the robotic metric analysis due to different evaluation methods being conducted during different sessions, by different assessors.

Towards defining a proper sample size, this study's sample size was based on the primary clinical outcome measure (FMA) and not on kinematic or kinetic metrics ([EDWARDS et al., 2019](#)). According to [Lo et al. \(2010\)](#), 12-weeks of UE robotic training was expected to result in a 2.88-point FMA change in the chronic stage of rehabilitation. For a combined tDCS and robotic intervention, a FMA change of 4.33 points was estimated. Using a two-sided alpha and a standard deviation of approximately 1.5, enrolling 56 subjects (28 per group) would result in 90% power to detect a difference of 1.45 in the FMA between the intervention groups. An enrollment goal of 66 subjects (33 per group) was selected to allow for a 15% subject attrition rate.

7.2 Data Analysis

The macro and micrometrics from [Chapter 5](#) were derived from S/E and wrist evaluation tasks ([Section 5.1](#)). Additional details about the S/E robotic evaluations have been published previously ([BOSECKER et al., 2010](#); [KREBS et al., 1999](#)).

7.2.1 Analysis of kinematic and kinetic data

Described in [Chapter 4](#), our data analysis framework ([MORETTI; DELBEM; KREBS, 2020](#)) was used to preprocess the raw data arising from the robotic devices.

Figure 39 – CONSORT flow diagram

Source: Adapted from Edwards *et al.* (2019).

The rationalia behind the robot metrics were based on formerly developed and tested calculations (BOSECKER *et al.*, 2010; DIPIETRO *et al.*, 2007) to automatically process the robotic kinematic and kinetic measures of all study subjects.

Statistical significance between *Robot_{tDCS}* and *Robot_{sham}* was determined for each robotic measure (both macro and micrometrics) by first conducting a Jarque-Bera test ($\alpha = 5\%$) to choose between a parametric or non-parametric test. As samples did not follow a normal distribution, we used the Wilcoxon-Mann-Whitney (one-tailed) unpaired two-sample test.

For the overall paired analysis of both patient groups (*Robot_{tDCS}* and *Robot_{sham}* combined) from admission to discharge, admission to follow-up and discharge to follow-up, we used the Wilcoxon signed rank (one-tailed) test.

All statistical analyses were performed using MATLAB^{TM} version R2019b and

results were deemed significant if p-values were less than 0.05.

7.2.2 Post-hoc exploratory analysis

The severity of patient's UE impairment was defined by his/her baseline FMA result. A brute force search was conducted to establish the range of FMA scores in which tDCS had a significant effect on the greatest number of robotic measures. Its process is depicted in Figure 40.

Figure 40 – Brute-force search process by reducing and offsetting a window along the FMA scale. The patients that fall in such window, i.e., with an FMA score within the window, were divided into $Robot_{tDCS}$ and $Robot_{Sham}$ groups and submitted to the unpaired test for each metric. Then, the first quartile (Q1) is obtained, as the criterion for finding the window in which metrics are impacted the most.

Source: Elaborated by the author.

The brute force search first analyzed the largest window of FMA scores (1 to 66, all study participants), identified $Robot_{tDCS}$ from $Robot_{Sham}$ subjects, and then applied the Wilcoxon-Mann-Whitney test for each robot metric. Next the FMA score window was reduced, from 1 to 65 then 2 to 66, and so on. The process of narrowing and offsetting the window along the scale was carried out for every case, generating lists of p-values for each combination. The subgroup with significant p-values in the lowest quartile demonstrated the largest difference between $Robot_{tDCS}$ and $Robot_{Sham}$ outcomes.

7.3 Results

Analyzing kinematic and kinetic data on the effectiveness of combined robotic and tDCS therapy, completed robotic evaluation results were available for 76 (out of 82) enrolled subjects at discharge and 72 at six-month follow-up (Figure 39). There were significant improvements in the kinematic and kinetic measures of motor control for the combined $Robot_{tDCS}$ and $Robot_{Sham}$ study population from admission to discharge (Table 6, Figures 41 and 42).

Figure 41 – Combined *Robot_{tDCS}* and *Robot_{Sham}* mean and standard error of kinematic and kinetic outcome metrics for S/E device. Significant changes ($p < 0.05$) between admission (ad), discharge (dc), and follow-up (fu) are marked with an * mark.

Source: Elaborated by the author.

For the S/E, most of kinematic and kinetic metrics significantly improved (Table 6). For p2p task, these included: deviation, mean speed, speed shape, jerk and submovement

Figure 42 – Combined $Robot_{DCS}$ and $Robot_{Sham}$ mean and standard error of kinematic and kinetic outcome metrics for wrist device. Significant changes ($p < 0.05$) between admission (ad), discharge (dc), and follow-up (fu) are marked with an * mark.

Source: Elaborated by the author.

number, duration, and overlap. Examples of patients at different levels of severity, who improved their tracing pattern, can be seen in Figures 43, 44 and 45). Both metrics of the reaching against resistance task (maximum displacement and overall aim) showed significant gains. Similarly, all metrics of the circle drawing task and both measures of the S/E isometric stabilization task (scatter and offset) demonstrated a significant change. The mean change in shoulder peak force (delta z) improved significantly from admission to discharge.

For the unconstrained wrist pointing task, deviation and speed shape improved

considerably. Submovement number trended down, overlap increased, and interpeak interval decreased, but there was no improvement in the measure of wrist jerk, worsening in -21.456 (95% CI $[-46.063, -1.117]$, $p = 0.89$). Maximum displacement and offset improved significantly for the wrist pointing against resistance and isometric stabilization task, respectively. All other metrics did not demonstrate a significant change (Figures 41 and 42).

On the impact of tDCS, there was no noteworthy difference in any metric between $Robot_{tDCS}$ and $Robot_{Sham}$ groups at admission, discharge, or follow-up for the S/E or wrist data of all patients.

As for the post-hoc analysis on stroke severity, for the S/E, only patients with severe motor impairments, i.e., 18 patients ($n = 18$) within $[7, 9]$ range of FMA scale, demonstrated a significant difference in kinematic metrics of motor control between groups, showing greater improvement for patients receiving $Robot_{Sham}$ ($n = 10$) compared to $Robot_{tDCS}$ ($n = 8$). Table 7 outlines the metrics that improved in the severe $Robot_{Sham}$ sub-group.

Table 6 – Significant changes from admission to discharge for the combined analysis of kinematic and kinetic data.

Device	Task	Metric	Change	95% CI	p-value
S/E	p2p	Deviation	0.007	[0.004, 0.01]	<0.001
		Mean Speed	0.009	[0.0002, 0.019]	0.03
		Speed shape	0.035	[0.022, 0.048]	0.001
		Jerk	45.907	[6.811, 99.157]	0.006
		SM number	0.804	[0.033, 1.595]	0.02
		SM duration	0.05	[0.024, 0.078]	<0.001
		SM overlap	0.024	[0.01, 0.04]	<0.001
	rdyn	Max displacement	0.008	[0.004, 0.012]	<0.001
		Overall aim	0.068	[0.026, 0.11]	<0.001
	circle	Axes ratio	0.085	[0.056, 0.114]	<0.001
		Minor axis	0.015	[0.01, 0.02]	<0.001
		Joint independence	0.091	[0.063, 0.122]	<0.001
	pstatic	Scatter	0.002	[-0.00006, 0.004]	<0.001
		Offset	0.004	[0.002, 0.007]	<0.001
	shoulder	Delta z	5.075	[2.08, 8.092]	0.001
Wrist	p2p	Deviation	0.018	[0.004, 0.033]	0.01
		Speed shape	0.043	[0.029, 0.059]	<0.001
		SM number	0.625	[-0.873, 2.175]	0.224
		SM overlap	0.013	[-0.024, 0.049]	0.074
		SM interpeak	0.012	[-0.033, 0.057]	0.411
	rdyn	Max displacement	0.018	[0.00004, 0.037]	0.01
	pstatic	Offset	0.002	[-0.044, 0.038]	0.006

Source: Elaborated by the author.

Table 7 – Post-hoc analysis results: significant changes seen in the *Robot_{Sham}* group in the severe patient sub-group.

S/E - FMA [7, 9], $n = 10$										
Metric	Admission to discharge			Admission to follow-up			Discharge to follow-up			p
	Change	95% CI	p	Change	95% CI	p	Change	95% CI	p	
SM duration (p2p)	0.165	[0.027,0.296]	0.03				0.154	[0.009,0.291]		0.02
SM overlap (p2p)	0.081	[0.014,0.149]	0.03							
Jerk (p2p)										
Overall aim (rdyn)				226.26	[-15.02,587.58]	0.02				
Axes ratio (circle)	0.125	[0.03,0.226]	0.03	0.252	[0.039,0.476]	0.02				
Joint indep. (circle)	0.141	[0.04,0.24]	0.01	0.197	[0.109,0.290]	<0.001				
Minor axis (circle)	0.035	[0.014,0.057]	0.004							
				0.43	[0.025,0.063]	0.001				
Wrist - FMA [8, 12], $n = 8$										
Metric	Admission to discharge			Admission to follow-up			Discharge to follow-up			p
	Change	95% CI	p	Change	95% CI	p	Change	95% CI	p	
SM interpeak (p2p)				0.1	[0.032,0.2]	0.007				
Overall aim (rdyn)				0.288	[0.013,0.554]	0.01				

Source: Elaborated by the author.

Figure 43 – Example movement plots of p2p (S/E and wrist) and circle (S/E) tasks for Patient #42627 (FMA of 9 at admission and 16 at follow-up).

Source: Elaborated by the author.

Figure 44 – Example movement plots of p2p (S/E and wrist) and circle (S/E) tasks for Patient #42658 (FMA of 15 at admission and 25 at follow-up).

Source: Elaborated by the author.

Figure 45 – Example movement plots of p2p (S/E and wrist) and circle (S/E) tasks for Patient #42633 (FMA of 24 at admission and 34 at follow-up).

Source: Elaborated by the author.

Similarly for the wrist, only patients with severe motor impairments (FMA [8, 12], $n = 16$) in the *Robot_{Sham}* group ($n = 8$) showed a significant difference in kinematic metrics (Table 7) from admission to follow-up. The remaining patient subgroups did not differ.

7.4 Discussion

Edwards *et al.* (2019) reported a mean FMA improvement of 7.36 points from baseline over the 12-week intervention. The kinematic and kinetic measures showed a significant improvement from admission to discharge. Our findings support previous literature that a robotic evaluation and the metrics it generates are consistent with clinical measures and may provide a reliable, standardized, and objective tool for the quantification of motor recovery (BOSECKER *et al.*, 2010; SEMRAU *et al.*, 2013; DIPIETRO *et al.*, 2012) that may have the potential advantage of less interrater variability and subjectivity (KREBS *et al.*, 2002).

The S/E kinematic results were consistent with previous investigations, showing a progressive recovery in movement smoothness (speed shape, jerk, and submovement metrics) (DIPIETRO *et al.*, 2012; ROHRER *et al.*, 2004), shoulder strength, movement against resistance (BOSECKER *et al.*, 2010), isometric stabilization, and coordination

(circle) (DIPIETRO *et al.*, 2007). Dipietro *et al.* (2012) and (ROHRER *et al.*, 2004) used the same robotic evaluation to investigate the pattern of recovery of trained and untrained (circle drawing) movements in acute and chronic patients post-stroke by analyzing changes in smoothness and submovements. Both studies reported acute motor recovery of inpatients to be characterized by reduced movement duration, increased mean and peak speed, but a concurrent worsening of jerk. The submovement profile of the acute patients demonstrated fewer, larger movements being blended together (submovement number reduced and duration, interpeak interval, and overlap increased). A similar pattern was evident for patients with chronic stroke, but smaller magnitude of change compared to acute. However, the jerk metric was seen to improve in the outpatient population. Our results replicate this pattern of chronic-stage stroke recovery, therefore supporting the value of both macro- and micro-kinematic measures to characterize motor recovery of the S/E. Taken together, these findings support kinematic and kinetic metrics as sensitive and valid tools for assessing recovery of the UE during both the rapid and marked spontaneous recovery early after stroke, as well as intervention-related improvement during the stable recovery plateau.

Improvements in the circle drawing task indicate that training point to point movements generalize motor gains to circle drawing and coordination of S/E movements away from the body. Dipietro *et al.* (2012), Dipietro *et al.* (2007) also reported an increase in circle ratio and kinematic changes for the circle drawing task. Consistent with our results, Dipietro *et al.* (2007) reported improvements in the circle ratio predominately due to changes in the minor axes, not the major axes, of the circle drawing task. The minor axis corresponds primarily to independent elbow extension and shoulder abduction movements, suggesting that patients improve their ability to move out of the UE flexor synergy (elbow flexion and shoulder abduction) and gain greater joint independence (see Figure 4). Circle drawing was not trained in this study or in Dipietro *et al.* (2012), Dipietro *et al.* (2007) and gains were sustained (six months after training), indicating generalizability of training outcomes. This supports the use of motor learning as a better model, not adaptation (usually characterized by short-term changes), for motor recovery of the UE (DIPIETRO *et al.*, 2012).

Wrist kinematic results generated from the combined analysis of all participants revealed a novel insight in motor recovery. In our chronic patient population, one measure of smoothness (speed shape) improved significantly but the jerk metric did not. A superficial interpretation suggests the intervention did not impact wrist smoothness. Nonetheless, wrist micrometrics and submovement characteristics provide clarification on the nature of the smoothness changes. The number of submovements decreased, overlap increased, and there was a trend for reduced interpeak interval. These results indicate that submovements grew closer together in the chronic stage of recovery. Interestingly, this pattern of motor recovery replicates the S/E submovement characteristics of the sub-acute inpatient population reported in Rohrer *et al.* (2004) and Dipietro *et al.* (2012). Although the chronic population

Figure 46 – Improvements in joint independence over time. (a) and (b): individual values of joint independence metric, for all patients, sorted by their performance at admission (squares) in comparison to discharge (diamonds), and follow-up (circles). A lower number indicates greater/improving joint independence. (c) and (d): changes in the joint independence metric, where filled circles indicate significance ($p < 0.05$) at discharge (c) and follow-up (d).

Source: Elaborated by the author.

within the same studies shared many trends in submovement characteristics, changes to interpeak interval appeared to be exclusive to the acute population for the S/E. It is possible that both the proximal and distal segments of the UE share the same characteristics of recovery (as shown by a submovement analysis), but the time-course of the wrist recovery profile is relatively delayed compared to the S/E. This conjecture may begin to explain

findings in the literature such as the superior recovery of the wrist compared to the S/E in chronic patients post-stroke. [Hsieh et al. \(2018\)](#) compared the effectiveness of distal versus proximal robotic therapy in patients six months or more after stroke onset. Results revealed participants receiving wrist robotic therapy had greater improvements in muscle strength and quality of movement in ADLs than patients receiving S/E therapy ([HSIEH et al., 2018](#)). Although further studies are required to investigate this conjecture, our wrist submovement analysis suggests there may be potential for patients with distal impairment to continue to improve well into the chronic stage of their recovery. Our results attest to the importance and value of kinematic measures and a submovement analysis to enrich our understanding of UE motor recovery post-stroke.

The analysis comparing *Robot_{tDCS}* and *Robot_{Sham}* kinematic and kinetic metric outcomes confirms the findings of the previously reported clinical outcomes ([EDWARDS et al., 2019](#)). Although significant motor gains were demonstrated between admission and discharge and admission and follow-up in both groups, the tDCS intervention was not shown to provide additional benefit to the motor recovery of the UE than robotic therapy alone in the chronic stroke population.

This result is consistent with numerous studies investigating combined robotic and tDCS therapy ([GIACOBBE et al., 2013](#); [STRAUDI et al., 2016](#)). [Giacobbe et al. \(2013\)](#) suggest tDCS may not have an additive or augmentation effect to the already significant motor gains induced by robotic therapy, but may instead change the nature of the training effect. It is possible that robotic training induces a ceiling effect whereby motor recovery is maximized, and any additional benefit from tDCS cannot be identified ([STRAUDI et al., 2016](#)). Comparing our study to other studies demonstrating either an additional positive effect of tDCS ([BOLOGNINI et al., 2011](#)) or a difference in the kinematic measures of the tDCS group ([GIACOBBE et al., 2013](#)) reveals that factors such as sample size, tDCS method, lesion site, motor task, stage of recovery, and type of stroke may contribute to the variance seen in patient outcomes ([LEFEBVRE; LIEW, 2017](#)).

This study is the first to our knowledge to investigate the impact of UE impairment severity on the effectiveness of tDCS. Most studies implementing tDCS recruit patients with a large range of impairment ([BOLOGNINI et al., 2011](#); [GIACOBBE et al., 2013](#)), despite neural recovery depending on the severity of stroke ([COUPAR et al., 2012](#); [KIM; WINSTEIN, 2017](#)). Our results revealed the only patient subgroup with a significant difference in kinematic measures of motor control between *Robot_{tDCS}* and *Robot_{Sham}* had very severe motor impairment in favor of the *Robot_{Sham}* group. This result is interesting for two reasons.

Firstly, most studies investigating stroke prognosis have concluded that patients with less burden of disease (measured by the functional integrity of the corticospinal tract or the structural integrity of descending white matter pathways) have better functional

outcomes (COUPAR *et al.*, 2012; KIM; WINSTEIN, 2017). Nonetheless, other studies have also reported more significant improvements in the chronic patient population or those with lower levels of cortical activity and greater UE impairment. Straudi *et al.* (2016) showed significant and positive therapy outcomes for a combined tDCS and robotics intervention that was dependent on stage of recovery (patients with chronic stroke improved more than the acute group) and lesion location (subcortical was superior to cortical). (MILOT *et al.*, 2014) investigated predictors of functional gain for patients receiving robotic therapy and found that lower baseline Motor Evoked Potential (MEP) activation correlated with better functional outcomes of the affected UE. The explanation offered for these findings was: the lower baseline cortical activity may represent underused cortical potential, therefore patients with higher cortical activation and baseline measures of motor control may have less reserve to augment either cortical activity or behavioral changes with therapy (MILOT *et al.*, 2014; STRAUDI *et al.*, 2016).

Secondly, the *Robot_{Sham}* group was shown to have better kinematic outcomes than *Robot_{tDCS}* in the severe subgroup of patients. Edwards *et al.* (2019) also reported significant clinical findings of improved FMA response rate, in terms of Minimum Clinically Important Difference (MCID), greater in the *Robot_{Sham}* group. The mechanisms for this result are unknown, although we speculate that increases in abnormal tone could have been a factor. Data on tone was not collected in this study to inform this theory, but we believe this warrants further investigation. As a difference between groups was only observed in patients with very severe UE impairments, this speaks to the diversity of interactions of tDCS between different deficits, cortical activity levels, and individual patients (KRAUSE; KADOSH, 2014).

Significant improvements in motor control were detected with kinematic and kinetic measures for a combined UE robot and tDCS intervention in chronic stroke. However, no difference was found between *Robot_{Sham}* and *Robot_{tDCS}* groups. This confirmed the results of previous clinical measures (EDWARDS *et al.*, 2019). Submovement characteristics provided insight into improvements in smoothness and highlighted the importance of kinematic micrometrics for enhancing our understanding of motor recovery post-stroke. Novel wrist kinematic data suggested an interesting pattern of delayed wrist motor recovery in the chronic stage compared to the timeline of the S/E. Our subgroup analysis revealed that patients with very severe motor impairment in the *Robot_{Sham}* group benefited the most from the intervention. Our study demonstrated limited benefit for adding anodal tDCS prior to robotic training in chronic stroke and future studies are needed to investigate the interaction of intensive motor training with other forms of neuro-stimulation and patient populations.

7.5 Conclusion

Consistent with confirmed results of clinical outcome measures, our study and analysis on the robotic data show that intensive robotic therapy is beneficial. However, no additional motor gain is obtained by administering anodal transcranial direct current stimulation. Patients with severe impairment did not benefit from the combined intervention. Submovement characteristics of wrist smoothness showed a delayed pattern of motor recovery compared to the shoulder/elbow, of relevance for intensive intervention-related recovery of upper extremity function in chronic stroke.

ESTIMATION OF CLINICAL SCALES BASED ON KINEMATIC MEASURES FOR CHRONIC STROKE

Stroke continues to be a leading cause of chronic disability and health-related costs (LOPEZ *et al.*, 2006) in the United States and across the globe (JAUCH *et al.*, 2013). Limited functional recovery of the UE is often reported following stroke and is the cause of significant long-term restriction to activities of daily life and community reintegration (LAWRENCE *et al.*, 2001; LAI *et al.*, 2002). More standardized and targeted measures of UE function are key to prescribing more effective and individualized therapy, with the aim of improving patient outcomes and reducing length of stay, the costs, and the demand for rehabilitation particularly in the chronic phase of stroke recovery (BERNHARDT *et al.*, 2019; HARCUM *et al.*, 2019).

A detailed evaluation of sensorimotor impairment and function is essential in planning appropriate and effective therapy interventions for the individual patient (HARCUM *et al.*, 2019; RECH *et al.*, 2020). Clinical outcome measures, such as the widely used Fugl-Meyer Assessment of Upper Extremity Motor Recovery after Stroke (FMA-UE), offer a convenient and low cost method of assessing motor function. These measures have been used as the gold standard for evaluation of rehabilitation outcomes across the continuum of care and in clinical research. Nonetheless, many clinical scales are limited by subjectivity and insensitivity to detect changes in motor control, particularly for patients with severe motor impairment (KREBS *et al.*, 2002; BOSECKER *et al.*, 2010; SEMRAU *et al.*, 2013; SEE *et al.*, 2013). Even for highly recommended clinical outcome measures with good psychometric properties, such as the FMA-UE, previous studies have highlighted that without adequate training and standardization of assessment procedures, the variability in inter-rater scores can be equivalent to the minimally clinically important difference of the scale (KREBS *et al.*, 2002; LEE *et al.*, 2013; SEE *et al.*, 2013). It is not well understood

that the FMA-UE is not a linear scale, meaning gains on the items at the beginning of the measure are not equivalent to and have a different meaning to items in the sections at the end of the measure. In addition, there are varied reports and uncertainty in the literature regarding the extent that the FMA-UE scale can be used to stratify patients at different stages of recovery and the impact of ceiling and floor effects (SEE *et al.*, 2013; GLADSTONE; DANELLS; BLACK, 2002; CRAMER, 2010). With rapid advances and improved accessibility of rehabilitation technology in the developed world, the reliability of clinical measures is now being questioned given that other measurement tools are more available.

Robotic derived kinematic evaluations may be one solution to better understand the recovery of stroke and assist with optimizing the rehabilitation journey for patients. The ability to objectively evaluate movement quality is key for distinguishing between behavioral restoration and behavioral compensation post-stroke (KWAKKEL *et al.*, 2019). Previous studies in this field suggest that kinematic data can provide a standardized and objective measure of a patient's motor control and movement quality, which correlates with well-known clinical outcome measures (COLOMBO *et al.*, 2005; BOSECKER *et al.*, 2010; SEMRAU *et al.*, 2013; DIPIETRO *et al.*, 2012; DUKELOW, 2017; KREBS *et al.*, 2013), and has the potential to enhance knowledge of treatment effects, clinical reasoning, and our understanding of stroke recovery (BERNHARDT *et al.*, 2019). The existing literature on correlating robotic kinematic data with clinical outcome measures has been limited by small sample sizes (ZOLLO *et al.*, 2011; MURPHY; WILLÉN; SUNNERHAGEN, 2011; KRABBen *et al.*, 2012), enrolling patients with a narrow spectrum of UE impairment (MURPHY; WILLÉN; SUNNERHAGEN, 2011; ZOLLO *et al.*, 2011), only including data on the proximal UE (BOSECKER *et al.*, 2010), or investigating a single modelling technique (BOSECKER *et al.*, 2010; KREBS *et al.*, 2013), potentially impacting the perceived value and strength of the correlation between robotic kinematic evaluations and clinical scales.

We point out here two main contributions we aim to deliver from this study. First by investigating the value of kinematic wrist data (in addition to the previously studied S/E metrics) to improve the correlation between standardized robotic measures and common clinical outcome measures of the UE. Secondly, we will build on previous work (BOSECKER *et al.*, 2010; KREBS *et al.*, 2013) by investigating both linear and nonlinear regression models for estimating clinical outcome measures from robot metrics in the chronic post-stroke population.

This study derives from the randomized controlled trial from Chapter 7 that elucidates the impact of tDCS preceding intensive upper-limb training for stroke patients. The robotic intervention involved 1024 movement repetitions per session, alternating S/E wrist-forearm robot therapy on separate days, performed three times a week for a total

of 36 sessions (KREBS *et al.*, 2007; KREBS *et al.*, 1998). Additional details of the study design and outcome measures are presented in Chapter 7 and Edwards *et al.* (2019).

8.1 Clinical Outcome Measures for the UE

The primary clinical outcome measure employed in this study was the FMA-UE, which uses an ordinal performance-based scale to score the level of UE motor impairment after stroke. The scale comprises 33 items designed to examine reflex activity, motor control, and muscle strength (with a maximum score of 66). The measure is both highly recommended by stroke guidelines and widely used for clinical and research settings in the chronic stage of stroke recovery (Shirley Ryan AbilityLab, 2016b).

Secondary outcome measures included the Wolf Motor Function Test (WMFT), Barthel Index (BI), and the Medical Research Council Motor Power score (MP). The WMFT is a 21-item measure of UE motor ability comprising three parts: time, functional ability, and strength tasks. The maximum score is 75 with lower scores indicating lower functional ability. WMFT is recommended for use in the assessment of stroke across the continuum of rehabilitation care (Shirley Ryan AbilityLab, 2016c). BI is a performance measure of functional independence including mobility, gait, and performance of ADLs. Ten items are rated according to the amount of assistance required to successfully complete the task, with a maximum score of 100 (Shirley Ryan AbilityLab, 2016a). The MP score is a grading system of manual muscle testing used widely in neurological and musculoskeletal clinical and research settings. The power of each muscle is evaluated in relation to the maximum expected for the movement of a single joint and scored from 0-5 (Shirley Ryan AbilityLab, 2020; Medical Research Council, 2020).

8.2 Dataset format and selection of UE Kinematic and Kinetic features

The dataset is based on the macro and micrometrics from Chapter 5, derived from S/E and wrist evaluation tasks (Section 5.1). Our data analysis framework from Chapter 4 (MORETTI; DELBEM; KREBS, 2020) was used to preprocess the raw data arising from the robotic devices, so that such metrics could be obtained from all study subjects for feature selection.

The dataset comprises evaluation sessions (admission, discharge and follow-up) for all patients, where features (predictive attributes) are the mean-aggregated kinematic and kinetic values, having the respective clinical measures of a particular session set as the target attributes.

Due to the high dimensionality of the dataset, we performed a correlation-based feature selection, keeping only features substantially correlated with the target attribute ($R > 0.5$, see Figures 47 and 48). In addition, for every group of features with high dependency among themselves, the one with better interpretability was preserved and the remaining features were discarded in order to reduce collinearity.

Figure 47 – Heatmap of S/E kinematic and kinetic data correlation with clinical scales. Significant correlations ($p < 0.05$) are marked with a triangle at the upper-right corner of the respective element.

Source: Elaborated by the author.

Similarly, correlations among the clinical measures were calculated to determine dependent scales (Figures 47 and 48). If there was a very strong correlation between clinical measures, this suggested redundancy whereby it may be possible to reduce the number of clinical outcome measures performed and therefore the time needed to conduct a clinical evaluation. The strength of the correlations calculated in the analyses (for both R and the proportion of variance, R^2) were interpreted as follows: $[0, 0.3)$ - very weak,

Figure 48 – Heatmap of wrist kinematic data correlation with clinical scales. Significant correlations ($p < 0.05$) are marked with a triangle at the upper-right corner of the respective element. The * mark refers to metrics from PS tasks, whereas the others are from 2D tasks.

Source: Elaborated by the author.

[0.3, 0.5) - weak, [0.5, 0.7) - moderate, and [0.7, 1.0] - strong (MOORE; NOTZ; FLIGNER, 2013).

8.3 Clinical estimator models

Following strategies from previous work, we considered linear (BOSECKER *et al.*, 2010) and nonlinear (KREBS *et al.*, 2013) complexities as candidate models for the correlation of clinical scales. Nonlinear models took the form of an MLP: one hidden layer with two sigmoid nodes and one linear output (KREBS *et al.*, 2013). Features were rescaled by normalizing to their maximum values. Clinical measures were rescaled for the

nonlinear model only.

Every candidate model was submitted to a 4-fold cross-validation. Because one patient was associated with more than one example (i.e., an evaluation session) inside the dataset, we implemented a variation of the traditional cross-validation with patient-wise sampling instead of instance wise (similar to Equation 6.2). Instances from 25% of patients (randomly selected, without replacement) were held as a test set, while the instances from the remaining 75% of patients were used to build a model for training. This process continued for each remaining 25%-75% partition. A Mann-Whitney test was carried out to ensure that instances from training and test partitions were similarly distributed with respect to their expected outputs.

The S/E MP analysis was performed in 3 ways to explore the correlation between the clinical MP measure and the different kinematic and kinetic methods of measuring shoulder force. The first correlation (MP total) was calculated from the model using both kinetic metrics (derived from a force transducer) and relevant kinematic metrics ($R > 0.5$). The second correlation (MP kinematic force metrics) was obtained from the model using only force derived kinematic metrics (from shoulder, rdyn and pstatic tasks). The third correlation (MP force transducer metrics) was obtained from the model using only the shoulder kinetic metrics to determine the value of the different shoulder kinetic metrics (shoulder AA, FE, mean shoulder force) and the use of a force transducer.

8.4 Results

Eighty two participants were enrolled in the study in the chronic stage of recovery following stroke with a wide range of UE impairment. Table 8 summarizes the demographics of study participants and the admission scores for the clinical evaluations.

Table 8 – Participant demographics and admission results of clinical measures

Characteristic	Robot _{tDCS}	Robot _{Sham}	Overall
Participants (n, %)	41 (50%)	41 (50%)	82 (100%)
Age (mean, range)	66.4 [42,87]	69.2 [42,90]	67.8 [42,90]
Days from stroke to start of trial (mean days, range)	1475.2 [226,6935]	1160 [151,6936]	1317.6 [151,6936]
Gender (n Female, %)	16 (39%)	16 (39%)	32 (39%)
Stroke location (n cortical, %)	26 (63.4%)	27 (65.9%)	53 (64.6%)
FMA-UE (mean, range)	25.6 [7,57]	25.4 [7,55]	25.5 [7,57]
WMFT (mean, range)	60 [1,169]	56 [0,167]	58 [0,169]
BI (mean, range)	88.4 [10,100]	85 [15,100]	86.7 [10,100]
MP (mean, range)	46.8 [8,79.5]	44.1 [15,85]	45.5 [8,85]

Source: Elaborated by the author.

8.4.1 Correlation between clinical scales

The majority of the clinical scales selected for this study, the FMA-UE, WMFT, and MP, were highly correlated, except for the BI, as can be seen in the heatmap representations from Figures 47 and 48.

8.4.2 Correlation between robotic data and clinical scales

Table 9 summarizes the correlation values reported in this section between the estimator models and the respective clinical scales using S/E and wrist features as input.

8.4.2.1 Correlation with FMA-UE

Six features with $R > 0.5$ were selected for the model that estimates FMA-UE from S/E data: offset (pstatic), delta z (shoulder), joint independence (circle), maximum displacement (rdyn), deviation (p2p) and speed shape (p2p). As can be seen in Table 9, the linear model fitted to such features showed a significant correlation with the FMA-UE ($R = 0.82$, $p < 0.001$), as well as the nonlinear model ($R = 0.88$, $p < 0.001$).

Five features were selected for the model that estimates FMA-UE from wrist data: deviation (FE/AA p2p), mean speed (FE/AA and PS p2p) and offset (FE/AA and PS pstatic). Both the linear and nonlinear models fitted to such features showed significant correlation with FMA-UE: $R = 0.91$ ($p < 0.001$) and $R = 0.92$ ($p < 0.001$), respectively.

The models fitted to S/E and wrist features combined (six S/E and five wrist and forearm metrics) also correlated well with FMA-UE: $R = 0.91$ ($p < 0.001$) for the linear model and $R = 0.91$ ($p < 0.001$) for the nonlinear model.

8.4.2.2 Correlation with BI

Five features ($R > 0.5$) were selected for the model that estimates BI from S/E data: deviation (p2p), submovement duration (p2p), overall aim (rdyn), minor axis (circle) and offset (pstatic). The linear model fitted to such features yielded a correlation with the BI of $R = 0.68$ ($p < 0.001$), and $R = 0.73$ ($p < 0.001$) for the nonlinear model.

Although no wrist features were found to be substantially correlated ($R > 0.5$) to the BI, four of the strongest, all with R below 0.5, were selected: deviation (FE/AA p2p), mean speed (FE/AA p2p), overall aim (FE/AA rdyn) and offset (PS pstatic). The models fitted to such features yielded lower correlations, with $R = 0.53$ ($p < 0.005$) for the linear model and $R = 0.58$ ($p < 0.005$) for the nonlinear model.

With the nine S/E and wrist features combined, the respective models also yielded lower correlations: $R = 0.67$ ($p < 0.001$) for the linear model and $R = 0.70$ ($p < 0.001$) for the nonlinear model.

8.4.2.3 Correlation with WMFT

Because WMFT is strongly correlated to the FMA-UE, the same six S/E and five wrist features were used to build estimators for WMFT scores.

A high correlation was found with the WMFT for both the linear and nonlinear models using S/E data: $R = 0.85$ ($p < 0.001$) and $R = 0.89$ ($p < 0.001$), respectively. The models using wrist features also correlate well with WMFT, with $R = 0.89$ ($p < 0.001$) for the linear model and $R = 0.9$ ($p < 0.001$) for the nonlinear model.

The models fitted to S/E and wrist features combined also resulted in strong correlations with WMFT: $R = 0.92$ ($p < 0.001$) for the linear model and $R = 0.93$ ($p < 0.001$) for the nonlinear model.

8.4.2.4 Correlation with MP

Five features were selected for the model that estimates MP from S/E data: offset (pstatic), delta z (shoulder), joint independence (circle), maximum displacement (rdyn) and deviation (p2p). The linear model fitted to such features showed a significant correlation with the MP ($R = 0.86$, $p < 0.001$), as well as the nonlinear model ($R = 0.87$, $p < 0.001$). The same five wrist features as used for estimating the FMA-UE were used to estimate MP, with a correlation of $R = 0.88$ ($p < 0.001$) for the linear model and $R = 0.90$ ($p < 0.001$) for the nonlinear model. Combining S/E and wrist features, the linear model yielded $R = 0.91$ ($p < 0.001$) and the nonlinear model yielded $R = 0.91$ ($p < 0.001$).

For the MP models using only force metrics, three features were found to have $R > 0.5$: offset (pstatic), delta z (shoulder) and maximum displacement (rdyn). A strong correlation of $R = 0.86$ ($p < 0.001$) was found for both the linear and nonlinear models. Four wrist features were selected for estimating MP: offset (FE/AA and PS pstatic) and overall aim (FE/AA and PS rdyn). The linear model showed a correlation of $R = 0.86$ ($p < 0.001$) and the nonlinear model $R = 0.90$ ($p < 0.001$). Combining S/E and wrist features, both the linear and nonlinear models correlated well with MP: $R = 0.89$ ($p < 0.001$) and $R = 0.90$ ($P < 0.001$), respectively.

Using only the metrics derived from the force transducer, the mean delta z (mean-aggregated, from sabd, sadd, sext and sflex tasks) produced a correlation of $R = 0.62$ ($p < 0.001$) for the linear model and $R = 0.66$ ($p < 0.001$) for the nonlinear model. Using the shoulder features individually, both the linear and nonlinear models produced a similar result of $R = 0.66$ ($p < 0.001$) and $R = 0.69$ ($p < 0.001$) for the nonlinear model.

Table 9 – Correlation between estimator models and and clinical scales. $**p < 0.001$, $*p < 0.05$.

Scale	Device	Model	R	p	Scale	Device	Model	R	p
FMA UE	S/E	Linear	0.82	**	BI	S/E	Linear	0.68	**
		Nonlinear	0.88	**			Nonlinear	0.73	*
	Wrist	Linear	0.91	**		Wrist	Linear	0.53	*
		Nonlinear	0.92	**			Nonlinear	0.58	**
	Both	Linear	0.91	**		Both	Linear	0.67	**
		Nonlinear	0.91	**			Nonlinear	0.7	**
WMFT	S/E	Linear	0.85	**	MP (Total)	S/E	Linear	0.86	**
		Nonlinear	0.89	**			Nonlinear	0.87	**
	Wrist	Linear	0.89	**		Wrist	Linear	0.88	**
		Nonlinear	0.9	**			Nonlinear	0.9	**
	Both	Linear	0.92	**		Both	Linear	0.91	**
		Nonlinear	0.93	**			Nonlinear	0.92	**
MP (only force metrics)	S/E	Linear	0.86	**	MP (only shoulder metrics)	S/E (mean)	Linear	0.62	**
		Nonlinear	0.86	**			Nonlinear	0.66	**
	Wrist	Linear	0.86	**		S/E (AA/ FE)	Linear	0.66	**
		Nonlinear	0.9	**			Nonlinear	0.69	**
	Both	Linear	0.9	**					
		Nonlinear	0.91	**					

Source: Elaborated by the author.

8.5 Discussion

This study provides both incremental and novel advances to the methodology of outcome measure correlation studies in stroke research. This is the first study to our knowledge to investigate the value of including comprehensive distal (wrist and forearm) kinematic data and compare the linear and nonlinear correlation models in the outpatient, chronic population post-stroke. The methods presented here build on previous work by this group investigating the non-linear correlation model in the inpatient population (KREBS *et al.*, 2013) and the linear correlation model in the chronic outpatient population (BOSECKER *et al.*, 2010).

Of note, our results reveal higher correlation values between the kinematic and kinetic metrics and clinical outcome measures (see [Appendix B](#) for optimized neural network parameters) compared to previous studies (BOSECKER *et al.*, 2010; SEMRAU *et al.*, 2013; KREBS *et al.*, 2013; COLOMBO *et al.*, 2005; DUKELOW, 2017; ZOLLO *et al.*, 2011). Numerous factors may have influenced this outcome. One reason may be the baseline characteristics of the patient population studied here ([Table 8](#)), which represent a broader more distributed spectrum of patient impairment (FMA-UE ranging from 7 to 57) than previous studies, which may also impact from a mathematical perspective ([Appendix A](#)). Bosecker *et al.* (2010) recruited a large sample of 111 patients in the chronic stage of recovery with an admission FMA-UE ranging from 7 to 38. Both the FMA-UE and

MP scores were included in their correlation analysis using a linear regression model, but displayed lower R values (0.802 for FMA-UE and 0.797 for MP) than what was found in this study, even for our linear model. Zollo *et al.* (2011) investigated the correlation of robotic kinematics using data from 15 chronic patients post stroke with an admission FMA-UE ranging from 8 to 36. This group also evaluated the correlation between robotic measures and the FMA-UE and MP using a linear regression method and reported only moderate R values ($R = 0.674$ and $R = 0.773$, respectively). It is well known that the value of a correlation is greater when there is greater variability in the observations studied (GOODWIN; LEECH, 2006).

Consistent with the results of Bosecker *et al.* (2010), submovement characteristics had little influence on the correlation between models using kinematic metrics and the clinical outcome measures. The only submovement metric seen to have a substantial correlation degree ($R > 0.5$) was submovement duration, which was selected for the correlation with respect of the BI (which had the lowest correlation value). Although submovement characteristics had a low impact in this correlation study and Bosecker *et al.* (2010), these metrics have been shown to be very important for enhancing our understanding of the motor recovery process of the UE (DIPIETRO *et al.*, 2012; ROHRER *et al.*, 2002).

In keeping with previous findings, the nonlinear model consistently performed better than the linear correlation model (Table 9) for predicting clinical outcome measures (KREBS *et al.*, 2013; KRAKOVSKA *et al.*, 2019). The variation in nonlinear and linear models of correlation are less dramatic than what is seen in the subacute population, likely due to less variability in the patient presentation (KRAKOVSKA *et al.*, 2019). The use of nonlinearity in the estimation models and a feature selection strategy was likely influential in the superior correlation results described in this study.

The inclusion of wrist kinematic data is novel within this field of study. Although the wrist data showed high correlation with the clinical outcome measures, it did not substantially improve the already strong correlation provided by the S/E data alone. This finding is valuable for numerous reasons. Clinically it is more efficient to only perform one robotic evaluation, therefore the results from this study do not support the need to complete both a wrist and S/E robotic evaluation in order to predict the results of clinical outcome measures. The clinician can select the robot that best suits the needs of the patient's impairments and goals without compromising the ability to generate predictions of clinical measures. We also saw that either robot is reliable at predicting the clinical scales in both high and low functioning patients (Table 8).

Correlation among clinical measures was high, except for the BI. The correlation between the FMA-UE and MP in this study was higher ($R = 0.94$) than what was published to date in the chronic stroke population. Bosecker *et al.* (2010) reported a

moderate correlation between MP and both the FMA-UE ($R = 0.785$) and Motor Status Scale (MSS, $R = 0.774$, a similar measure to the FMA-UE but consisting of a finer grading scale for UE impairment for the subacute patient population). Of note, the study by [Bosecker et al. \(2010\)](#) on a chronic post-stroke population differed from this study by using only a linear regression model. [Krebs et al. \(2013\)](#) also reported high correlation between MP and FMA-UE ($R = 0.933$). Therefore the robot metrics were seen to predict the outcome of FMA-UE, WMFT, and MP consistently.

The analysis of the BI in this study indicates that the measure does not correlate well with either the clinical measures or with robotic metrics, confirming previous reports ([CHEN; WINSTEIN, 2009](#); [MAYO et al., 2002](#)). In a recent large cohort study of 434 subjects in the chronic stage of recovery, [Mayo et al. \(2002\)](#) reported patients had a mean score of 90.6/100 on the BI, yet 65% of this population could not incorporate their affected UE into activities of daily living. [Sivan et al. \(2011\)](#) also reported limited responsiveness of the BI involving chronic patients with severe impairments. It appears there are significant variations within and between the results of a patient's BI score, which implies that a patient's mood and other factors may influence or bias the scoring of the scale.

Although the methodology used in this study led to higher correlations between robotic metrics and clinical outcome measures than previous studies, the results may not apply to all patient populations. It is unclear if the superior correlations generated with the nonlinear model extrapolate to other phases of stroke recovery, which warrants future investigation using our methodology within the acute and subacute population. The same limitation may apply to the strength of wrist kinematic data correlating with clinical measures. In our study the wrist data did not substantially improve the correlation value compared to the S/E correlation result alone, however the wrist data may be more influential in the acute or subacute stages of recovery following stroke.

8.6 Conclusions

The improved method of correlation modeling for robotic kinematics and clinical measures in this study is an important step towards objectively measuring and planning a patient's rehabilitation journey in the chronic stage of recovery following stroke. The often time consuming and subjective nature of clinical outcome measures and the growing access and validity of robotic measures will likely increase the demand for reliable correlation models such as the methods described in this study. Robotic measures offer enhanced objectivity of movement quality, accuracy, and standardization that can improve our understanding of the complex process of motor recovery in chronic stroke. As this study confirms, robotic kinematics are well suited to representing impairments in body function or structure but are perhaps restricted in their ability to translate kinematic data to

performance of daily activities or participation in real-world environments, which warrants further study.

EARLY PREDICTION OF CLINICAL OUTCOMES FOR CHRONIC STROKE PATIENTS

In terms of motor function recovery, tracking stroke patient's performance deterministically, in the same sense traditional clinical scales are currently employed, still remains an underexplored major problem in the literature. This extends to robotic rehabilitation with the usage of machine learning models, due to the lack of an expressive sample that carries the underlying pattern, so that predictive models could have the due statistical significance. In addition, provided that motor recovery takes place at different paces according to the phase of the stroke (e.g., acute, subacute and chronic), the difficulty on this problem scales up even more.

Demographics and clinical/medical data from the beginning of the treatment have been considered relevant on detecting trending clinical improvements for both chronic and acute stroke patients. However, such outcome predictors may not be reliable when applied to patients singularly (TOZLU *et al.*, 2020; COUPAR *et al.*, 2012; KIM; WINSTEIN, 2017; BURKE; CRAMER, 2013; ASADI *et al.*, 2014).

Tozlu *et al.* (2020) used neurophysiological and imaging features, based on Magnetic Resonance Imaging (MRI) data to predict FMA scores at discharge. MRI-based features assess functional and cognitive traits from a pathological perspective, by inspecting alterations in white matter integrity (KUCEYESKI *et al.*, 2013), which lays beyond of our scope here. Hsieh *et al.* (2014) and Duret *et al.* (2019) also used clinical (Box and Block test and FMA, respectively) and demographics data to predict FMA scores at discharge.

All the aforementioned kinds of attributes arise from the admission phase, which is considered the best point in time to yield metrics with the best predictive capabilities (COUPAR *et al.*, 2012). Such a claim is what commonly supports related work in the

literature in the sense of analyzing a static time point (admission phase), not addressing changes along therapy from a time-series perspective.

Following the approach traditionally employed in the literature, as well as addressing time dependency issues, this chapter outlines our efforts on predicting treatment outcomes, i.e., whether the patient responded well enough to the treatment at discharge, from both static and time-dependent perspectives. Towards such goal, this study aims to train models from shallow learning and deep learning and mechanisms, respectively.

9.1 Time dependency

We used the preprocessed data from the same chronic stroke patients from the study described in [Chapter 7](#), since tDCS itself provided no impact on motor performance towards retention or gains of motor function. We based our experiments on a concept proposed one century ago by Hermann Ebbinghaus, regarding the forgetting curve ([MURRE; DROS, 2015](#)), as depicted in [Figure 49a](#). Basically, such curve indicates how information is lost over time when there is no attempt to retain it. Equivalently, we plan to associate the retention ([Figure 49b](#)) of motor recovery over time by observing changes in the robot-based metrics ([Chapter 5](#)) along the treatment.

In order to address the time dependency of the examples, we used a Long Short Term Memory (LSTM) neural network, aiming to capture the overall performance trend in terms of robot-based metrics, as depicted by the dashed curve in [Figure 49a](#), and map it to the patient's recovery potential as the expected output. For comparison, we used a Multi-Layer Perceptron (MLP) neural network for building a predictor based on a static point in time, following the traditional approaches in literature, as mentioned in the aforementioned related work.

9.2 Machine learning

We followed the correlation-based study ([Chapter 8](#)) for feature selection to build the dataset ([Figure 50a](#)) for the LSTM models to be trained on. For such, two strategies were considered.

The first strategy concerns a regression problem, The LSTM model is recurrently fed by robot metrics along the first two therapy sessions as input, whereas the output is the respective FMA score at discharge (option (1) from [Figure 50b](#)), equivalently with the study described in [Chapter 8](#), but now taking into account the time dependency. This model then turns into a classifier by taking the difference between admission assessment and predicted discharge assessment, verifying whether such a difference exceeds an MCID

Figure 49 – Performance of patient #42658 along therapy sessions described by deviation metric. (a) Ebbinghaus’ forgetting curve. Each therapy session is represented by a solid curve. Dashed curves are exponential functions fitted to sessions individually (shorter curves) or to all sessions (longer curve). (b) gains kept (retained) from the previous session, defined as $i - 1$ -th session’s last point minus i -th session’s first point.

Source: Elaborated by the author.

of 3 points. *Responders* are the patients that did exceed that threshold, whereas the others are *non-responders*.

The second strategy consists in a classification problem. It has input examples directly labelled as responder/non-responder classes (option (2) from Figure 50b). As in the first strategy, following the same MCID criterion to define the class labels (before training), the difference here is between actual (not predicted) discharge and admission FMA scores.

Both first and second strategies also involve the training of MLP models, using the same features. Since this kind of model is trained on data from a fixed time point, the dataset consists of mean-aggregated values from the evaluation session at admission only.

9.3 Results

Although it has been stated in the literature the importance of FMA itself at admission to be its own prediction for discharge (TOZLU *et al.*, 2020), more precision

Figure 50 – Dataset format with patients included in the post analysis (size $n = 76$, see Figure 39) for LSTM models. (a) Metric values at movement level are fed at each time step, for all examples. (b) The expected output: FMA score for the first strategy or Responder/non-responder label for the second strategy.

Source: Elaborated by the author.

is needed than just a static prediction. That being said, we built different datasets with different combinations of data to increment the first two therapy sessions at the input: with and without FMA and with and without the evaluation session prior to therapy for S/E and wrist devices (individually and merged/concatenated). Examples with missing data that could not form a proper time sequence (i.e., more than one cycle is missing) were discarded.

For each of these combinations, a grid search was conducted for tuning the hyperparameters for training LSTM models, combining: number of LSTM nodes (5, 10 or 15), number of fully connected nodes (from 0, i.e., no layer, to 16), learning rate (10^{-1} , 10^{-2} or 10^{-3}) and epochs (100, 150 or 200). This process was repeated several times with different seeds for the initial random initialization of weights. The same rationale was applied for MLP models, but with mean-aggregated values of the robot metrics from the evaluation before therapy, combining evaluation from S/E and wrist devices and the FMA at admission. The grid search optimized only the number of nodes at the hidden layer from 1 to 16. The learning rate was automatically adjusted over epochs with the Levenberg-Marquardt training algorithm.

One can observe good predictions both from the time-series perspective (LSTM models) and the static perspective (MLP models) in Tables 10 and 11. The LSTM results refer to models from the second strategy, where examples are directly labeled as *responder*

or *non-responder*. Aiming to estimate the FMA scale, the models from the first strategy resulted in strong correlations ($R^2 > 0.98$ for S/E evaluation and therapy data with FMA at admission, as the best case) with the scale itself. Nevertheless, the conversion of such predictions into a classifier model yielded inferior results.

Table 10 – Best parameter configuration for the LSTM models (grid search results), given different (time-dependent) input data combinations.

S/E		Wrist		FMA	MCC	Best configuration (LSTM nodes, nodes, learning rate)
Eval	Ther	Eval	Ther	(Adm)		
■	■	■	■	■	0.777365	10, 6, 10^{-3}
				■	0.712726	5, 4, 10^{-3}
■	■	■	■	■	0.777365	10, 6, 10^{-3}
				■	0.712726	5, 4, 10^{-3}
■	■	■	■	■	0.721179	10, 2, 10^{-1}
				■	0.688312	5, 4, 10^{-2}
■	■	■	■	■	0.732242	5, 4, 10^{-2}
				■	0.712726	5, 2, 10^{-3}
■	■	■	■	■	0.839924	10, 2, 10^{-1}
				■	0.765992	5, 6, 10^{-2}
■	■	■	■	■	0.886621	10, 6, 10^{-1}
				■	0.777365	15, 4, 10^{-2}

Source: Elaborated by the author.

Table 11 – Best parameter configuration for the MLP models (grid search results), given different (static) input data combinations.

S/E	Wrist	FMA	MCC	Best config. (nodes)
Eval	Eval	(Adm)		
■	■	■	0.888683	4
		■	0.888683	12
■	■	■	0.888683	8
		■	0.888683	12
■	■	■	0.712726	12
		■	0.703526	2
■	■	■	0.858842	2

Source: Elaborated by the author.

Given that the final dataset had 49 responders and 11 non-responders, the classifiers were evaluated using the Matthews Correlation Coefficient (MCC) due to the severe class imbalancing. We considered the MCC because it combines all values from the confusion matrix towards a more reliable outcome with respect to other classification metrics, even though Area Under ROC Curve (AUC) culminated in higher outcomes than the ones from Tables 10 and 11.

9.4 Discussion

Because both strategies for training LSTM models share the same classification task, it is not surprising that the second strategy yielded better results, even though reliable estimations of FMA could be obtained from features inherent to FMA (as also reported in [Chapter 8](#)).

Provided that stroke outpatients will not change dramatically with therapy, this can explain the high MCC values for both types of models, which did perform equivalently. Interestingly, the predictors with the highest performance are MLP models, using data from a static point in time.

From the best models in [Table 11](#), one can observe that model complexity tends to compensate the amount of information provided as input, i.e., the more features provided, the lower the complexity of the model, and vice versa, towards the same performance. This indicates that information about FMA itself is not necessarily key for more reliable predictions, even though it plays a very important role in this scenario. Moreover, the grid search result for the MLP model with FMA as the only input is consistent with our strategy from [Chapter 8](#) and the findings in [Krebs *et al.* \(2013\)](#), regarding model complexity for the nonlinearity trait of the FMA scale.

In contrast, the role of FMA at admission was more substantial on boosting the performance of LSTM models, as can be seen in [Table 10](#). It is also possible to observe that time-static S/E data yielded better results with MLP models, whereas wrist data yielded better results with LSTM models.

Application wise, it becomes more convenient to use the MLP models rather than LSTM models, since the performances between them are equivalent. Such a claim has foundation in the SLT, since the dataset is static with respect to time and samples are not iid. Although many more samples are needed for obtaining the due learning guarantees, the model complexity for, say, the model using four hyperplanes from [Table 11](#), requires an attainable amount of points (patients, in this case), for such scenario to happen, following the rationale from [Figure 34](#) and [Section 2.2](#).

Other than artificial neural networks, the application of other machine learning mechanisms for classification are to be explored in future work. Following [Krebs *et al.* \(2013\)](#), we initially considered MLP models, since it is already familiar in the application context literature. Using accelerometer data, [Wang *et al.* \(2014\)](#) trained an SVM model for regression, predicting FMA scores. Although such a solution may be feasible on the prediction, a kernel must reconstruct the data space, since FMA scale is not linear, as verified in [Chapter 8](#) and in [Krebs *et al.* \(2013\)](#). Therefore, it is important to take into account the interpretability of the data space, so it can be properly consolidated from the clinician's perspective.

Apart from SVM and neural networks, [Tozlu et al. \(2020\)](#) used tree-based algorithms and elastic nets on the prediction of FMA from demographics and clinical data, which may be good guidelines with robot-based metrics in the future.

We also implemented the PREP2 algorithm proposed by [Stinear et al. \(2017\)](#) for comparison. Such an algorithm consists in a decision tree that evaluates patient's age and clinical scores, such as MEP, Shoulder Abduction and Finger Extension (SAFE) and the National Institutes of Health Stroke Scale (NIHSS). The leaves of the decision tree classify patients as *excellent*, *good*, *limited* and *poor*. The first two and last two labels were here considered as *responder* and *non-responder*, respectively. Using data from the same patients that resulted in the outcomes presented in [Table 11](#), the PREP2 algorithm yielded an MCC of 0.21. Such a discrepancy takes place for two reasons. Firstly, such an algorithm was designed for patients recruited early with respect to onset (207 patients, within 3 days of stroke), whereas our models were trained on data from chronic stroke patients; it must be taken into account that the recovery process of patients in different phases of the stroke (e.g., acute, sub-acute, chronic) take place at different paces. Secondly, more data is needed for our models so more consistent results can be obtained.

We are currently investigating why LSTM models did not outperform MLP models, given that more information is being fed, as well as that recovery takes place gradually, and that such a trending feature can be typically detected in time-series analyses. Our first steps towards this goal in the future is to add more complexity to LSTM models in different orders of magnitude, with proper hyperparameter adjustments, and observe its impact.

The importance of an accurate tracking of patient's progress on treatments is key not only for customizing treatment to patients' conditions towards anticipating outcomes. It is also of great importance that the final outcome (discharge) itself can be predicted at early stages of the treatment. That being said, patients that do not present a significant response to the treatment (from a clinical perspective) can then cease the treatment, since pursuing that outcome would not bring solid changes in quality of life.

9.5 Conclusion

From our current outcomes, we conclude that it is possible to improve the decision-making process of clinicians by alerting when patients will potentially not present a clinically significant response to the treatment, which may substantially impact the quality of life of stroke patients. Bringing to debate the possibility of ceasing the treatment, since pursuing that outcome would not enhance patient's motor function, can be beneficial, in the sense of saving resources that can be used other ways towards bringing actual improvements in quality of life of stroke patients.

CONCLUSION

Along this thesis, after introducing the context of robotic rehabilitation and the motivation of this project, given by the rich potential of data analysis, we started reviewing related work concerning the assessment of progress of stroke patients enrolled in rehabilitation programs. Owing to the fact that the definition of biomarkers for a quantitative and deterministic assessment of patient's progress remains as an open problem in the literature, we conducted four different studies in this PhD project, aiming to address such an issue. In parallel, side projects were developed, providing supporting tools to the studies focused on the patient data, as well as clarification of implementation details, which also culminated in contributions for all the fields in the multidisciplinary.

We developed a modular framework for organizing programs for data preprocessing, visualization, and analysis, as well as the preprocessed data itself within a structure that can be exported in a portable format. The standards from our framework's data structure are intended to be used for simplifying the data sharing from The 77 Laboratory at MIT, which has one of the largest data sources in the field. We provided clarification and implemented in our framework the calculations of more than 20 commonly used robot-based metrics, which remain subjectively explained in the literature. The proposed standards from both the framework's data structure and the rationalia behind the implementation of robot-based metrics are to be a good roadmap for conducting data analysis studies for this kind of data.

From a naive perspective (i.e., purely based on data evidences) a study was conducted towards attempting to map a spacial region, in terms of features based on statistical moments. Initially alluding to a state of rehabilitation, such a region in a 12-dimensional space would define a novel biomarker for tracking patient's progress along treatment. Although the results seem to be promising with a bigger data volume, the clinical perspective does not coadunate with its consolidation, given the role of the brain hemispheres and the impact of lesions in different sizes with respect to UE motor function

in general.

Focusing on the clinical importance for consolidating data analysis strategies, we initially provided one more evidence, consistent with related work, from the analysis on the robotic data of 82 patients, that intensive robotic therapy is beneficial. We also found that no additional motor gain is obtained by administering anodal transcranial direct current stimulation. Even patients with severe impairment did not benefit from the combined intervention, as such a debate has been taking place in the field. Submovement characteristics of wrist smoothness showed a delayed pattern of motor recovery compared to the shoulder/elbow, of relevance for intensive intervention-related recovery of upper extremity function in chronic stroke.

Using the same data, we took an important step towards objectively measuring and planning a patient's rehabilitation journey in the chronic stage of recovery following stroke. Such a step took place by modeling clinical score estimators from the robotic kinematic and kinetic measures derived from the robot data. The often time consuming and subjective nature of clinical outcome measures and the growing access and validity of robotic measures will likely increase the demand for reliable estimation models, such as the ones obtained through our studies. Robotic measures offer enhanced objectivity of movement quality, accuracy, and standardization that can improve our understanding of the complex process of motor recovery in chronic stroke. As our study confirms, robotic kinematics are well suited to representing impairments in body function or structure but are perhaps restricted in their ability to translate kinematic data to performance of daily activities or participation in real-world environments, which warrants further study.

Cumulatively with our findings, we devoted our efforts on investigating whether chronic stroke patients could present early evidences of recovery during treatment, after the first two therapy sessions. Both from time-dependent and time-static perspectives, our contributions on the early prediction of clinical outcomes showed promising results, especially from a static point in time, impacting both the clinician's guideline and patient's quality of life. The importance of an accurate tracking of patient's progress on treatments is key not only for customizing treatments to patients' conditions towards anticipating outcomes. It is also of great importance that the final outcome (discharge phase) itself can be predicted at early stages of the treatment. That being said, patients that do not present a significant response to the treatment (from a clinical perspective) can then cease the treatment and redirect their resources, since pursuing that outcome would not bring solid changes in quality of life.

10.1 Publications

Along the development of this research project, our previous and current results culminated in the published papers enumerated below, as well as two more under consideration for publication and one to be submitted.

- MORETTI, C. B.; JOAQUIM, R. C.; CAURIN, G. A. P.; KREBS, H. I.; MARTINS, J. Knowledge discovery, rehabilitation robotics, and serious games: Examining training data. In: **Biomedical Robotics and Biomechatronics (2014 5th IEEE RAS EMBS International Conference on)**. 2014. p. 567–572. ISSN 2155-1774.
- MORETTI, C. B.; JOAQUIM, R. C.; TERRANOVA, T. T.; BATTISTELLA, L. R.; MAZZOLENI, S.; CAURIN, G. A. P. Knowledge discovery strategy over patient performance data towards the extraction of hemiparesis-inherent features: A case study. In: **2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)**. 2016. p. 717–722. ISSN 2155-1782.
- MORETTI, C. B.; DELBEM, A. C. B.; KREBS, H. I. Human-robot interaction: Kinematic and kinetic data analysis framework. In: **2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)**. 2020. p. 235–239. ISSN 2155-1782.

In [Moretti *et al.* \(2014\)](#) we presented an initial attempt to apply Knowledge Discovery techniques over real performance data, from patients enrolled in robotic therapy, in order to explore how to better optimize therapy. Also involving an InMotion Arm device (different from the ones from [Table 4](#)), in this pilot essay we used PCA for dimensionality reduction and an MLP was used to predict performance measures defined in ([KREBS *et al.*, 2003](#)). From a jackknife approach in sampling (before applying PCA) we evaluated the impact of kinematic and dynamic variables in the induction of a model for the prediction of such measures. We observed that force data are the most relevant, whereas positions presented no contributions concerning the performance measures.

In [Moretti *et al.* \(2016\)](#) we raised our first evidences of linear separability in data from hemiparetic patients over a KDD strategy. With the goal of classifying hemiparetic sides, we performed feature extraction based on the statistical moments over the distribution of kinematic and dynamic data from instances of cycles of movements. Only data from device A ([Table 4](#)) were involved. Different classifiers were considered, such as J48 Decision Tree, Random Forest, k-Nearest Neighbors and Multi-Layer Perceptron. We perceived a proneness to linear separability by the best result from a grid search, involving the aforementioned machine learning mechanisms and ranges of parameter values. The best result consists of an MLP with a single neuron at the hidden layer. We concluded that

out of the different biases $\mathcal{F}$ considered, the linear bias was sufficient to deliver the best result in terms of classification accuracy. Such finding motivated the usage of the SVM algorithm in [Chapter 6](#). It was also observed that most of the features arose from force data, as we previously perceived the relevance of force in ([MORETTI *et al.*, 2014](#)).

In [Moretti, Delbem and Krebs \(2020\)](#) we proposed a modular framework ([Chapter 4](#)) for organizing programs for data preprocessing, visualization, and analysis, as well as the preprocessed data itself within a structure that can be exported in a portable format. The standards from our framework's data structure are intended to be used for simplifying the data sharing from The 77 Laboratory at MIT, which has one of the largest data sources in the field.

The projects from [Chapters 7 and 8](#) were submitted to the Neurorehabilitation and Neural Repair (NNR) journal and are currently under consideration. The project from [Chapter 9](#) is to be continued for further publication.

BIBLIOGRAPHY

ANAND, S.; BUCHNER, A. **Decision support using data mining**. London: Financial Times Management, 1998. Citation on page 36.

ANDERSON, D. R.; BURNHAM, K. P. General strategies for the analysis of ringing data. **Bird Study**, v. 46, n. sup001, p. S261–S270, 1999. Available: <<http://dx.doi.org/10.1080/00063659909477253>>. Citation on page 35.

ANG, K. K.; GUAN, C.; CHUA, K. S. G.; ANG, B. T.; KUAH, C.; WANG, C.; PHUA, K. S.; CHIN, Z. Y.; ZHANG, H. Clinical study of neurorehabilitation in stroke using eeg-based motor imagery brain-computer interface with robotic feedback. In: **2010 Annual International Conference of the IEEE Engineering in Medicine and Biology**. [S.l.: s.n.], 2010. p. 5549–5552. ISSN 1094-687X. Citation on page 44.

ASADI, H.; DOWLING, R.; YAN, B.; MITCHELL, P. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. **PLoS One**, v. 9, n. 2, p. e88225, 2014. Citation on page 143.

BARTKE, K. **2D, 3D and High-Dimensional Data and Information Visualization**. 2005. [Http://archiv.iwi.uni-hannover.de/lv/seminar_ss05/bartke/Assets/Paper.pdf](http://archiv.iwi.uni-hannover.de/lv/seminar_ss05/bartke/Assets/Paper.pdf). Accessed 25 mar. 2015. Citation on page 36.

BERNHARDT, J.; BORSCHMANN, K. N.; KWAKKEL, G.; BURRIDGE, J. H.; ENG, J. J.; WALKER, M. F.; BIRD, M. L.; CRAMER, S. C.; HAYWARD, K. S.; O’SULLIVAN, M. J.; CLARKSON, A. N.; CORBETT, D. Setting the scene for the Second Stroke Recovery and Rehabilitation Roundtable. **Int J Stroke**, v. 14, n. 5, p. 450–456, 07 2019. Citations on pages 115, 116, 131, and 132.

BOIAN, R.; DEUTSCH, J.; LEE, C. S.; BURDEA, G.; LEWIS, J. Haptic effects for virtual reality-based post-stroke rehabilitation. In: **Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on**. [S.l.: s.n.], 2003. p. 247–253. Citation on page 32.

BOLOGNINI, N.; VALLAR, G.; CASATI, C.; LATIF, L. A.; EL-NAZER, R.; WILLIAMS, J.; BANCO, E.; MACEA, D. D.; TESIO, L.; CHESSA, C.; FREGNI, F. Neurophysiological and behavioral effects of tDCS combined with constraint-induced movement therapy in poststroke patients. **Neurorehabil Neural Repair**, v. 25, n. 9, p. 819–829, 2011. Citations on pages 115 and 128.

BOSECKER, C.; DIPIETRO, L.; VOLPE, B.; KREBS, H. I. Kinematic robot-based evaluation scales and clinical counterparts to measure upper limb motor performance in patients with chronic stroke. **Neurorehabil Neural Repair**, v. 24, n. 1, p. 62–69, Jan 2010. Citations on pages 43, 58, 59, 60, 61, 62, 65, 70, 71, 75, 108, 116, 117, 118, 125, 131, 132, 135, 139, 140, and 141.

BURKE, E.; CRAMER, S. C. Biomarkers and predictors of restorative therapy effects after stroke. **Curr Neurol Neurosci Rep**, v. 13, n. 2, p. 329, Feb 2013. Citation on page 143.

BURKE, J.; MCNEILL, M.; CHARLES, D.; MORROW, P.; CROSBIE, J.; MCDONOUGH, S. Optimising engagement for stroke rehabilitation using serious games. **The Visual Computer**, Springer-Verlag, v. 25, n. 12, p. 1085–1099, 2009. ISSN 0178-2789. Available: <<http://dx.doi.org/10.1007/s00371-009-0387-4>>. Citation on page 32.

CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; ZANASI, A. **Discovering Data Mining: From Concepts to Implementation**. [S.l.]: Prentice Hall, 1998. Citations on pages 36 and 38.

CASADIO, M.; SANGUINETI, V.; MORASSO, P. G.; ARRICHELLO, V. Braccio di ferro: a new haptic workstation for neuromotor rehabilitation. **Technology And Health Care: Official Journal Of The European Society For Engineering And Medicine**, v. 14, n. 3, p. 123 – 142, 2006. ISSN 0928-7329. Citations on pages 56, 57, 58, and 65.

CHEMUTURI, R.; AMIRABDOLLAHIAN, F.; DAUTENHAHN, K. Adaptive training algorithm for robot-assisted upper-arm rehabilitation, applicable to individualised and therapeutic human-robot interaction. **Journal of NeuroEngineering and Rehabilitation**, v. 10, n. 1, p. 102, 2013. ISSN 1743-0003. Available: <<http://www.jneuroengrehab.com/content/10/1/102>>. Citation on page 56.

CHEN, S. Y.; WINSTEIN, C. J. A systematic review of voluntary arm recovery in hemiparetic stroke: critical predictors for meaningful outcomes using the international classification of functioning, disability, and health. **J Neurol Phys Ther**, v. 33, n. 1, p. 2–13, Mar 2009. Citation on page 141.

CHERNOFF, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sums of observations. **Annals of Mathematical Statistics**, v. 23, p. 409–507, 1952. Citation on page 47.

CIOŚ, K.; TERESINSKA, A.; KONIECZNA, S.; POTOCKA, J.; SHARMA, S. A knowledge discovery approach to diagnosing myocardial perfusion. **Engineering in Medicine and Biology Magazine, IEEE**, v. 19, n. 4, p. 17–25, July 2000. ISSN 0739-5175. Citation on page 36.

CIOŚ, K. J.; KURGAN, L. A. Advances techniques in knowledge discovery and data mining. In: _____. [S.l.]: Springer, 2005. chap. Trends in Data Mining and Knowledge Discovery, p. 1–26. Citation on page 38.

COLOMBO, R.; PISANO, F.; MICERA, S.; MAZZONE, A.; DELCONTE, C.; CARROZZA, M. C.; DARIO, P.; MINUCO, G. Robotic techniques for upper limb evaluation and rehabilitation of stroke patients. **IEEE Trans Neural Syst Rehabil Eng**, v. 13, n. 3, p. 311–324, Sep 2005. Citations on pages 116, 132, and 139.

COLOMBO, R.; STERPI, I.; MAZZONE, A.; DELCONTE, C.; PISANO, F. Taking a lesson from patients' recovery strategies to optimize training during robot-aided rehabilitation. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 20, n. 3, p. 276–285, May 2012. ISSN 1534-4320. Citations on pages 58, 59, 60, 61, 63, 65, 66, 70, 71, and 75.

COUPAR, F.; POLLOCK, A.; ROWE, P.; WEIR, C.; LANGHORNE, P. Predictors of upper limb recovery after stroke: a systematic review and meta-analysis. **Clinical Rehabilitation**, v. 26, n. 4, p. 291–313, 2012. PMID: 22023891. Available: <<https://doi.org/10.1177/0269215511420305>>. Citations on pages 128, 129, and 143.

CRAMER, S. C. Stratifying patients with stroke in trials that target brain repair. **Stroke**, v. 41, n. 10 Suppl, p. S114–116, Oct 2010. Citation on page 132.

Dept. of Veterans Affairs; Dept. of Defense. **VA/DoD Clinical Practice Guideline for the Management of Stroke Rehabilitation, Guideline Summary**. 2010. [Http://www.rehab.research.va.gov/jour/10/479/pdf/VADODclinicalGuidelines479.pdf](http://www.rehab.research.va.gov/jour/10/479/pdf/VADODclinicalGuidelines479.pdf). Accessed 12 fev. 2015. Citation on page 32.

DEUSCHL, G.; BAIN, P.; BRIN, M. Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee. **Mov. Disord.**, v. 13 Suppl 3, p. 2–23, 1998. Citation on page 78.

DIFFRIENT, N. **Humanscale 1/2/3**. Cambridge, MA, USA: MIT Press, 1974. Citation on page 87.

DIJCK, G. V.; VAERENBERGH, J. V.; HULLE, M. M. V. Posterior probability profiles for the automated assessment of the recovery of patients with stroke from activity of daily living tasks. **Artif Intell Med**, v. 46, n. 3, p. 233–249, Jul 2009. Citations on pages 61, 62, and 108.

DIPIETRO, L.; KREBS, H. I.; FASOLI, S. E.; VOLPE, B. T.; STEIN, J.; BEVER, C.; HOGAN, N. Changing motor synergies in chronic stroke. **J Neurophysiol**, v. 98, n. 2, p. 757–768, Aug 2007. Citations on pages 17, 84, 85, 86, 87, 118, and 126.

DIPIETRO, L.; KREBS, H. I.; VOLPE, B. T.; STEIN, J.; BEVER, C.; MERNOFF, S. T.; FASOLI, S. E.; HOGAN, N. Learning, not adaptation, characterizes stroke motor recovery: Evidence from kinematic changes induced by robot-assisted therapy in trained and untrained task in the same workspace. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 20, n. 1, p. 48–57, Jan 2012. ISSN 1534-4320. Citations on pages 58, 59, 60, 63, 65, 66, 70, 71, 75, 116, 125, 126, 132, and 140.

DNIT. **Anuário Estatístico das Rodovias Federais**. 2010. Available: <<http://www.dnit.gov.br/rodovias/operacoes-rodoviararias/estatisticas-de-acidentes/anuario-2010.pdf>>. Accessed: 9 feb. 2015. Citation on page 31.

DUKELOW, S. P. The potential power of robotics for upper extremity stroke rehabilitation. **Int J Stroke**, v. 12, n. 1, p. 7–8, 01 2017. Citations on pages 116, 132, and 139.

DUKELOW, S. P.; HERTER, T. M.; BAGG, S. D.; SCOTT, S. H. The independence of deficits in position sense and visually guided reaching following stroke. **J Neuroeng Rehabil**, v. 9, p. 72, Oct 2012. Citation on page 116.

DURET, C.; PILA, O.; GROSMARE, A. G.; KOEPPPEL, T. Can robot-based measurements improve prediction of motor performance after robot-assisted upper-limb rehabilitation in patients with moderate-to-severe sub-acute stroke? **Restor Neurol Neurosci**, v. 37, n. 2, p. 119–129, 2019. Citation on page 143.

EDWARDS, D. J.; CORTES, M.; RYKMAN-PELTZ, A.; CHANG, J.; ELDER, J.; THICKBROOM, G.; MARIMAN, J. J.; GERBER, L. M.; OROMENDIA, C.; KREBS, H. I.; FREGNI, F.; VOLPE, B. T.; PASCUAL-LEONE, A. Clinical improvement with intensive robot-assisted arm training in chronic stroke is unchanged by supplementary tDCS. **Restor. Neurol. Neurosci.**, v. 37, n. 2, p. 167–180, 2019. Citations on pages [115](#), [116](#), [117](#), [118](#), [125](#), [128](#), [129](#), and [133](#).

FAYYAD, U.; GRINSTEIN, G. G.; WIERSE, A. **Information Visualization in Data Mining and Knowledge Discovery**. San Francisco: Morgan Kaufmann, 2002. Citation on page [36](#).

FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, v. 17, p. 37–54, 1996. Citations on pages [35](#), [36](#), and [95](#).

FITZGIBBON, A.; PILU, M.; FISHER, R. B. Direct least square fitting of ellipses. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 21, n. 5, p. 476–480, 1999. Citation on page [84](#).

FLASH, T.; HOGAN, N. The coordination of arm movements: an experimentally confirmed mathematical model. **J Neurosci**, v. 5, n. 7, p. 1688–1703, Jul 1985. Citations on pages [82](#), [83](#), and [116](#).

FORMICA, D.; CHARLES, S.; ZOLLO, L.; GUGLIELMELLI, E.; HOGAN, N.; KREBS, H. The passive stiffness of the wrist and forearm. **Journal of neurophysiology**, v. 108, p. 1158–66, 05 2012. Citation on page [84](#).

GIACOBBE, V.; KREBS, H. I.; VOLPE, B. T.; PASCUAL-LEONE, A.; RYKMAN, A.; ZEIRATI, G.; FREGNI, F.; DIPIETRO, L.; THICKBROOM, G. W.; EDWARDS, D. J. Transcranial direct current stimulation (tDCS) and robotic practice in chronic stroke: the dimension of timing. **NeuroRehabilitation**, v. 33, n. 1, p. 49–56, 2013. Citations on pages [115](#), [117](#), and [128](#).

GLADSTONE, D. J.; DANELLIS, C. J.; BLACK, S. E. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. **Neurorehabil Neural Repair**, v. 16, n. 3, p. 232–240, Sep 2002. Citations on pages [62](#) and [132](#).

GOLDSCHMIDT, R.; PASSOS, E. **Data Mining: Um Guia Prático**. Rio de Janeiro: Campus, 2005. ISBN 8535218777. Citation on page [38](#).

GOMI, H.; KAWATO, M. Human arm stiffness and equilibrium-point trajectory during multi-joint movement. **Biol Cybern**, v. 76, n. 3, p. 163–171, Mar 1997. Citations on pages [56](#) and [58](#).

GOODWIN, L. D.; LEECH, N. L. **Understanding Correlation: Factors That Affect the Size of r**. [S.l.]: The Journal of Experimental Education, 2006. Citation on page [140](#).

GREENE, C. S.; TAN, J.; UNG, M.; MOORE, J. H.; CHENG, C. Big data bioinformatics. **Journal of Cellular Physiology**, v. 229, n. 12, p. 1896–1900, 2014. Citation on page [66](#).

GROSS, R.; SHI, J. **The CMU Motion of Body (MoBo) Database**. Pittsburgh, PA, 2001. Citation on page [66](#).

HALL, M. A. Correlation-based feature selection for discrete and numeric class machine learning. In: **Proceedings of the Seventeenth International Conference on Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000. (ICML '00), p. 359–366. ISBN 1-55860-707-2. Citation on page [42](#).

HARCUM, S.; CONROY, S. S.; BOOS, A.; ERMER, E.; XU, H.; ZHAN, M.; CHEN, H.; WHITALL, J.; DIMYAN, M. A.; WITTENBERG, G. F. Methods for an Investigation of Neurophysiological and Kinematic Predictors of Response to Upper Extremity Repetitive Task Practice in Chronic Stroke. **Arch Rehabil Res Clin Transl**, v. 1, n. 3-4, Dec 2019. Citation on page [131](#).

HOEFFDING, W. Probability inequalities for sums of bounded random variables. **Journal of the American Statistical Association**, American Statistical Association, v. 58, n. 301, p. pp. 13–30, 1963. ISSN 01621459. Citation on page [47](#).

HOGAN, N. Adaptive control of mechanical impedance by coactivation of antagonist muscles. **IEEE Transactions on Automatic Control**, v. 29, n. 8, p. 681–690, 1984. Citation on page [82](#).

N. Hogan, H.I. Krebs, A. Sharon and J. Charnnarong. **Interactive robotic therapist**. 1995. 5,466,213. Citations on pages [55](#), [58](#), and [95](#).

HOGAN, N.; STERNAD, D. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. **J Mot Behav**, v. 41, n. 6, p. 529–534, Nov 2009. Citation on page [83](#).

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence**. Cambridge, MA, USA: MIT Press, 1992. ISBN 0262082136. Citation on page [42](#).

HOLZINGER, A.; DEHMER, M.; JURISICA, I. Knowledge Discovery and interactive Data Mining in Bioinformatics–State-of-the-Art, future challenges and research directions. **BMC Bioinformatics**, v. 15 Suppl 6, p. I1, 2014. Citations on pages [36](#) and [39](#).

HORAK, F. B. Assumptions underlying motor control for neurologic rehabilitation. In: **Contemporary Management of Motor Problems**. [S.l.]: Chapter 4, 1990. p. 11–27. Citation on page [31](#).

HSIEH, Y. W.; LIN, K. C.; WU, C. Y.; LIEN, H. Y.; CHEN, J. L.; CHEN, C. C.; CHANG, W. H. Predicting clinically significant changes in motor and functional outcomes after robot-assisted stroke rehabilitation. **Arch Phys Med Rehabil**, v. 95, n. 2, p. 316–321, Feb 2014. Citation on page [143](#).

HSIEH, Y. W.; LIN, K. C.; WU, C. Y.; SHIH, T. Y.; LI, M. W.; CHEN, C. L. Comparison of proximal versus distal upper-limb robotic rehabilitation on motor performance after stroke: a cluster controlled trial. **Sci Rep**, v. 8, n. 1, p. 2091, 02 2018. Citations on pages [115](#) and [128](#).

HWANG, B.-W.; KIM, S.; LEE, S.-W. A full-body gesture database for automatic gesture recognition. In: **7th International Conference on Automatic Face and Gesture Recognition (FGR06)**. [S.l.: s.n.], 2006. p. 243–248. ISSN null. Citation on page [66](#).

ITF. Road safety annual report 2016. OECD Publishing, 2016. Available: <http://www.oecd-ilibrary.org/transport/road-safety-annual-report-2016_irtad-2016-en>. Accessed: 3 oct. 2017. Citation on page 31.

JAIN, A. K.; CHANDRASEKARAN, B. Handbook of statistics. In: _____. Amsterdam: North-Holland, 1982. v. 2, chap. Dimensionality and sample size considerations in pattern recognition in practice, p. 835–855. Citation on page 43.

JAUCH, E. C.; SAVER, J. L.; ADAMS, H. P.; BRUNO, A.; CONNORS, J. J.; DEMAER-SCHALK, B. M.; KHATRI, P.; MCMULLAN, P. W.; QURESHI, A. I.; ROSENFELD, K.; SCOTT, P. A.; SUMMERS, D. R.; WANG, D. Z.; WINTERMARK, M.; YONAS, H. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. **Stroke**, v. 44, n. 3, p. 870–947, Mar 2013. Citation on page 131.

JUNG, J.-Y.; GLASGOW, J. I.; SCOTT, S. H. Feature selection and classification for assessment of chronic stroke impairment. In: **2008 8th IEEE International Conference on BioInformatics and BioEngineering**. [S.l.: s.n.], 2008. p. 1–5. Citations on pages 58, 59, 60, 61, 63, 64, and 108.

KIM, B.; WINSTEIN, C. Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review. **Neurorehabil Neural Repair**, v. 31, n. 1, p. 3–24, 01 2017. Citations on pages 128, 129, and 143.

KRABBEN, T.; PRANGE, G. B.; MOLIER, B. I.; STIENEN, A. H.; JANNINK, M. J.; BUURKE, J. H.; RIETMAN, J. S. Influence of gravity compensation training on synergistic movement patterns of the upper extremity after stroke, a pilot study. **J Neuroeng Rehabil**, v. 9, p. 44, Jul 2012. Citation on page 132.

KRAKOVSKA, O.; CHRISTIE, G.; SIXSMITH, A.; ESTER, M.; MORENO, S. Performance comparison of linear and non-linear feature selection methods for the analysis of large survey datasets. **PLoS One**, v. 14, n. 3, p. e0213584, 2019. Citation on page 140.

KRAUSE, B.; KADOSH, R. C. Not all brains are created equal: the relevance of individual differences in responsiveness to transcranial electrical stimulation. **Front Syst Neurosci**, v. 8, p. 25, 2014. Citation on page 129.

KREBS, H. I. **Robot-aided neuro-rehabilitation and functional imaging**. Phd Thesis (PhD Thesis) — Massachusetts Institute of Technology, 1997. Citations on pages 21, 88, and 89.

KREBS, H. I.; AISEN, M. L.; VOLPE, B. T.; HOGAN, N. Quantization of continuous arm movements in humans with brain injury. **Proceedings of the National Academy of Sciences**, National Academy of Sciences, v. 96, n. 8, p. 4645–4649, Apr 1999. ISSN 0027-8424. Available: <<https://www.pnas.org/content/96/8/4645>>. Citations on pages 73, 75, 88, 116, and 117.

KREBS, H. I.; DIPIETRO, L.; LEVY-TZEDEK, S.; FASOLI, S.; RYKMAN-BERLAND, A.; ZIPSE, J.; FAWCETT, J.; STEIN, J.; POIZNER, H.; LO, A.; VOLPE, B.; HOGAN, N. A paradigm shift for rehabilitation robotics. **Engineering in Medicine and Biology Magazine, IEEE**, v. 27, n. 4, p. 61–70, July 2008. ISSN 0739-5175. Citations on pages 31 and 43.

KREBS, H. I.; HOGAN, N.; AISEN, M. L.; VOLPE, B. T. Robot-aided neurorehabilitation. **IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society**, v. 6, n. 1, p. 75–87, Mar. 1998. ISSN 1063-6528. Available: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2692541/>>. Citations on pages 55, 57, 58, 65, 67, 95, and 133.

KREBS, H. I.; KRAMS, M.; AGRAFIOTIS, D. K.; DIBERNARDO, A.; CHAVEZ, J. C.; LITTMAN, G. S.; YANG, E.; BYTTEBIER, G.; DIPIETRO, L.; RYKMAN, A.; MCARTHUR, K.; HAJJAR, K.; LEES, K. R.; VOLPE, B. T. Robotic measurement of arm movements after stroke establishes biomarkers of motor recovery. **Stroke**, v. 45, n. 1, p. 200–204, Jan 2013. Citations on pages 82, 116, 132, 135, 139, 140, 141, and 148.

KREBS, H. I.; PALAZZOLO, J. J.; DIPIETRO, L.; FERRARO, M.; KROL, J.; RANNEKLEIV, K.; VOLPE, B. T.; HOGAN, N. Rehabilitation robotics: Performance-based progressive robot-assisted therapy. **Autonomous Robots**, Kluwer Academic Publishers, v. 15, n. 1, p. 7–20, 2003. ISSN 0929-5593. Citations on pages 31, 32, 59, 61, and 153.

KREBS, H. I.; VOLPE, B. T.; FERRARO, M.; FASOLI, S.; PALAZZOLO, J.; ROHRER, B.; EDELSTEIN, L.; HOGAN, N. Robot-aided neurorehabilitation: from evidence-based to science-based rehabilitation. **Top Stroke Rehabil**, v. 8, n. 4, p. 54–70, 2002. Citations on pages 65, 116, 125, and 131.

KREBS, H. I.; VOLPE, B. T.; WILLIAMS, D.; CELESTINO, J.; CHARLES, S. K.; LYNCH, D.; HOGAN, N. Robot-aided neurorehabilitation: a robot for wrist rehabilitation. **IEEE Trans Neural Syst Rehabil Eng**, v. 15, n. 3, p. 327–335, Sep 2007. Citation on page 133.

KUCEYESKI, A.; MARUTA, J.; RELKIN, N.; RAJ, A. The Network Modification (NeMo) Tool: elucidating the effect of white matter integrity changes on cortical and subcortical structural connectivity. **Brain Connect**, v. 3, n. 5, p. 451–463, 2013. Citation on page 143.

KURGAN, L. A.; MUSILEK, P. A survey of knowledge discovery and data mining process models. **Knowl. Eng. Rev.**, Cambridge University Press, New York, NY, USA, v. 21, n. 1, p. 1–24, Mar. 2006. ISSN 0269-8889. Available: <<http://dx.doi.org/10.1017/S0269888906000737>>. Citations on pages 36, 38, 39, and 65.

KWAKKEL, G.; WEGEN, E. V.; BURRIDGE, J. H.; WINSTEIN, C. J.; DOKKUM, L. van; MURPHY, M. A.; LEVIN, M. F.; KRAKAUER, J. W. Standardized measurement of quality of upper limb movement after stroke: Consensus-based core recommendations from the Second Stroke Recovery and Rehabilitation Roundtable. **Int J Stroke**, v. 14, n. 8, p. 783–791, 10 2019. Citation on page 132.

LAI, S. M.; STUDENSKI, S.; DUNCAN, P. W.; PERERA, S. Persisting consequences of stroke measured by the Stroke Impact Scale. **Stroke**, v. 33, n. 7, p. 1840–1844, Jul 2002. Citations on pages 115 and 131.

LAWRENCE, E. S.; COSHALL, C.; DUNDAS, R.; STEWART, J.; RUDD, A. G.; HOWARD, R.; WOLFE, C. D. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. **Stroke**, v. 32, n. 6, p. 1279–1284, Jun 2001. Citations on pages 115 and 131.

LEE, D.; ROH, H.; PARK, J.; LEE, S.; HAN, S. Drinking behavior training for stroke patients using action observation and practice of upper limb function. **J Phys Ther Sci**, v. 25, n. 5, p. 611–614, May 2013. Citation on page 131.

LEFEBVRE, S.; LIEW, S. L. Anatomical Parameters of tDCS to Modulate the Motor System after Stroke: A Review. **Front Neurol**, v. 8, p. 29, 2017. Citations on pages 115 and 128.

LINCOLN, N.; CROW, J.; JACKSON, J.; WATERS, G.; ADAMS, S.; HODGSON, P. The unreliability of sensory assessments. **Clinical Rehabilitation**, v. 5, n. 4, p. 273–282, 1991. Available: <<https://doi.org/10.1177/026921559100500403>>. Citation on page 65.

LINDE, R. van der; LAMMERTSE, P. Hapticmaster - a generic force controlled robot for human interaction. **Industrial Robot: An International Journal**, v. 30, n. 6, p. 515–524, 2003. Citations on pages 55 and 58.

LO, A. C.; GUARINO, P. D.; RICHARDS, L. G.; HASELKORN, J. K.; WITTENBERG, G. F.; FEDERMAN, D. G.; RINGER, R. J.; WAGNER, T. H.; KREBS, H. I.; VOLPE, B. T.; BEVER, C. T.; BRAVATA, D. M.; DUNCAN, P. W.; CORN, B. H.; MAFFUCCI, A. D.; NADEAU, S. E.; CONROY, S. S.; POWELL, J. M.; HUANG, G. D.; PEDUZZI, P. Robot-assisted therapy for long-term upper-limb impairment after stroke. **N Engl J Med**, v. 362, n. 19, p. 1772–1783, May 2010. Citations on pages 115 and 117.

LOHSE, K. R.; LANG, C. E.; BOYD, L. A. Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. **Stroke**, v. 45, n. 7, p. 2053–2058, Jul 2014. Citation on page 115.

LOPEZ, A. D.; MATHERS, C. D.; EZZATI, M.; JAMISON, D. T.; MURRAY, C. J. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. **Lancet**, v. 367, n. 9524, p. 1747–1757, May 2006. Citation on page 131.

LUM, P. S.; BURGAR, C. G.; SHOR, P. C.; MAJMUNDAR, M.; LOOS, M. Van der. Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke. **Arch Phys Med Rehabil**, v. 83, n. 7, p. 952–959, Jul 2002. Citations on pages 56 and 58.

LUXBURG, U. von; SCHÖLKOPF, B. Statistical learning theory: Models, concepts, and results. In: MAX-PLANCK-GESELLSCHAFT. **Handbook of the History of Logic, Vol. 10: Inductive Logic**. Amsterdam, Netherlands: Elsevier North Holland, 2011. v. 10, p. 651–706. Citations on pages 44, 48, and 50.

MANDERY, C.; TERLEMEZ, O.; DO, M.; VAHRENKAMP, N.; ASFOUR, T. Unifying representations and large-scale whole-body motion databases for studying human motion. **IEEE Transactions on Robotics**, v. 32, n. 4, p. 796–809, Aug 2016. ISSN 1941-0468. Citation on page 66.

MANOGARAN, G.; LOPEZ, D. A survey of big data architectures and machine learning algorithms in healthcare. **International Journal of Biomedical Engineering and Technology**, v. 25, p. 182, 01 2017. Citation on page 66.

MAYO, N. E.; WOOD-DAUPHINEE, S.; COTE, R.; DURCAN, L.; CARLTON, J. Activity, participation, and quality of life 6 months poststroke. **Arch Phys Med Rehabil**, v. 83, n. 8, p. 1035–1042, Aug 2002. Citation on page 141.

MAZZOLENI, S.; SALE, P.; TIBONI, M.; FRANCESCHINI, M.; CARROZZA, M. C.; POSTERARO, F. Upper limb robot-assisted therapy in chronic and subacute stroke patients: a kinematic analysis. **Am J Phys Med Rehabil**, v. 92, n. 10 Suppl 2, p. 26–37, Oct 2013. Citation on page 115.

Medical Research Council. **MRC Muscle scale**. **UK Research and Innovation**. 2020. Available: <<https://mrc.ukri.org/research/facilities-and-resources-for-researchers/mrc-scales/mrc-muscle-scale/>>. Citation on page 133.

MILLER, E. L.; MURRAY, L.; RICHARDS, L.; ZOROWITZ, R. D.; BAKAS, T.; CLARK, P.; BILLINGER, S. A. Comprehensive overview of nursing and interdisciplinary rehabilitation care of the stroke patient: A scientific statement from the american heart association. **Stroke**, v. 41, n. 10, p. 2402–2448, 2010. Available: <<http://stroke.ahajournals.org/content/41/10/2402.short>>. Citation on page 32.

MILOT, M. H.; SPENCER, S. J.; CHAN, V.; ALLINGTON, J. P.; KLEIN, J.; CHOU, C.; PEARSON-FUHRHOP, K.; BOBROW, J. E.; REINKENSMeyer, D. J.; CRAMER, S. C. Corticospinal excitability as a predictor of functional gains at the affected upper limb following robotic training in chronic stroke survivors. **Neurorehabil Neural Repair**, v. 28, n. 9, p. 819–827, 2014. Citations on pages 115 and 129.

MOORE, D. S.; NOTZ, W. I.; FLIGNER, M. **The Basic Practice of Statistics**. [S.l.]: W.H. Freeman and Co., 2013. Citation on page 135.

MORETTI, C. B.; ANDRADE, K. O.; CAURIN, G. A. P. Physiotherapy support web-based system for rehabilitation robotics: an initial architecture. In: **22nd. International Congress of Mechanical Engineering (COBEM 2013)**. [S.l.: s.n.], 2013. p. 1171–1180. Citations on pages 32 and 66.

MORETTI, C. B.; DELBEM, A. C. B.; KREBS, H. I. Human-robot interaction: Kinematic and kinetic data analysis framework. In: **2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)**. [S.l.: s.n.], 2020. p. 235–239. ISSN 2155-1782. Citations on pages 34, 65, 117, 133, and 154.

MORETTI, C. B.; JOAQUIM, R. C.; CAURIN, G. A.; KREBS, H. I.; MARTINS, J. Knowledge discovery, rehabilitation robotics, and serious games: Examining training data. In: **Biomedical Robotics and Biomechatronics (2014 5th IEEE RAS EMBS International Conference on)**. [S.l.: s.n.], 2014. p. 567–572. ISSN 2155-1774. Citations on pages 100, 153, and 154.

MORETTI, C. B.; JOAQUIM, R. C.; TERRANOVA, T. T.; BATTISTELLA, L. R.; MAZZOLENI, S.; CAURIN, G. A. P. Knowledge discovery strategy over patient performance data towards the extraction of hemiparesis-inherent features: A case study. In: **2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)**. [S.l.: s.n.], 2016. p. 717–722. ISSN 2155-1782. Citations on pages 58, 59, 60, 61, 96, 98, 100, 101, 104, 109, and 153.

MURPHY, M. A.; WILLÉN, C.; SUNNERHAGEN, K. S. Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass. **Neurorehabil Neural Repair**, v. 25, n. 1, p. 71–80, Jan 2011. Citation on page 132.

MURRE, J. M.; DROS, J. Replication and Analysis of Ebbinghaus' Forgetting Curve. **PLoS One**, v. 10, n. 7, p. e0120644, 2015. Citation on page 144.

MUTHA, P. K.; HAALAND, K. Y.; SAINBURG, R. L. The effects of brain lateralization on motor control and adaptation. **J Mot Behav**, v. 44, n. 6, p. 455–469, 2012. Citation on page 109.

OLIVEIRA, L. F.; SIMPSON, D. M.; NADAL, J. Calculation of area of stabilometric signals using principal component analysis. **Physiol Meas**, v. 17, n. 4, p. 305–312, Nov 1996. Citation on page 84.

ORFANIDIS, S. **Introduction to Signal Processing**. Prentice Hall, 1996. (Prentice Hall international editions). ISBN 9780132091725. Available: <<https://books.google.com/books?id=CQUfAQAAIAAJ>>. Citations on pages 77 and 78.

OVBIAGELE, B.; GOLDSTEIN, L. B.; HIGASHIDA, R. T.; HOWARD, V. J.; JOHNSTON, S. C.; KHAVJOU, O. A.; LACKLAND, D. T.; LICHTMAN, J. H.; MOHL, S.; SACCO, R. L.; SAVER, J. L.; TROGDON, J. G. Forecasting the future of stroke in the United States: a policy statement from the American Heart Association and American Stroke Association. **Stroke**, v. 44, n. 8, p. 2361–2375, Aug 2013. Citation on page 115.

PATEL, S.; HUGHES, R.; HESTER, T.; STEIN, J.; AKAY, M.; DY, J.; BONATO, P. Tracking motor recovery in stroke survivors undergoing rehabilitation using wearable technology. In: **2010 Annual International Conference of the IEEE Engineering in Medicine and Biology**. [S.l.: s.n.], 2010. p. 6858–6861. ISSN 1094-687X. Citations on pages 61, 62, and 108.

QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0. Citation on page 42.

RECH, K. D.; SALAZAR, A. P.; MARCHESE, R. R.; SCHIFINO, G.; CIMOLIN, V.; PAGNUSSAT, A. S. Fugl-Meyer Assessment Scores Are Related With Kinematic Measures in People with Chronic Hemiparesis after Stroke. **J Stroke Cerebrovasc Dis**, v. 29, n. 1, p. 104463, Jan 2020. Citation on page 131.

REINKENSMEYER, D. J.; EMKEN, J. L.; CRAMER, S. C. Robotics, motor learning, and neurologic recovery. **Annu Rev Biomed Eng**, v. 6, p. 497–525, 2004. Citation on page 116.

REINKENSMEYER, D. J.; KAHN, L. E.; AVERBUCH, M.; MCKENNA-COLE, A.; SCHMIT, B. D.; RYMER, W. Z. Understanding and treating arm movement impairment after chronic brain injury: progress with the ARM guide. **J Rehabil Res Dev**, v. 37, n. 6, p. 653–662, 2000. Citations on pages 56 and 58.

RICHARDSON, M. J.; FLASH, T. Comparing smooth arm movements with the two-thirds power law and the related segmented-control hypothesis. **J Neurosci**, v. 22, n. 18, p. 8201–8211, Sep 2002. Citation on page 82.

ROHRER, B.; FASOLI, S.; KREBS, H. I.; HUGHES, R.; VOLPE, B.; FRONTERA, W. R.; STEIN, J.; HOGAN, N. Movement smoothness changes during stroke recovery. **Journal of Neuroscience**, Society for Neuroscience, v. 22, n. 18, p. 8297–8304, 2002. ISSN 0270-6474. Available: <<https://www.jneurosci.org/content/22/18/8297>>. Citations on pages 116 and 140.

ROHRER, B.; FASOLI, S.; KREBS, H. I.; VOLPE, B.; FRONTERA, W. R.; STEIN, J.; HOGAN, N. Submovements grow larger, fewer, and more blended during stroke recovery. **Motor Control**, v. 8, n. 4, p. 472–483, Oct 2004. Citations on pages 91, 125, and 126.

SACHDEVA, J.; KUMAR, V.; GUPTA, I.; KHANDELWAL, N.; AHUJA, C. K. Segmentation, feature extraction, and multiclass brain tumor classification. **J Digit Imaging**, v. 26, n. 6, p. 1141–1150, Dec 2013. Citation on page 40.

SANCHEZ, R. J.; LIU, J.; RAO, S.; SHAH, P.; SMITH, R.; RAHMAN, T.; CRAMER, S. C.; BOBROW, J. E.; REINKENSMEYER, D. J. Automating arm movement training following severe stroke: functional exercises with quantitative feedback in a gravity-reduced environment. **IEEE Trans Neural Syst Rehabil Eng**, v. 14, n. 3, p. 378–389, Sep 2006. Citations on pages 56, 58, and 65.

SANTOS, W. M. **Desenvolvimento de um dispositivo robótico interativo para reabilitação de lesões da articulação do joelho**. Master's Thesis (Master's Thesis) — Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos., 2013. Citation on page 43.

SCOTT, S. H.; DUKELOW, S. P. Potential of robots as next-generation technology for clinical assessment of neurological disorders and upper-limb therapy. **J Rehabil Res Dev**, v. 48, n. 4, p. 335–353, 2011. Citation on page 116.

SEE, J.; DODAKIAN, L.; CHOU, C.; CHAN, V.; MCKENZIE, A.; REINKENSMEYER, D. J.; CRAMER, S. C. A standardized approach to the Fugl-Meyer assessment and its implications for clinical trials. **Neurorehabil Neural Repair**, v. 27, n. 8, p. 732–741, Oct 2013. Citations on pages 131 and 132.

SEMRAU, J. A.; HERTER, T. M.; SCOTT, S. H.; DUKELOW, S. P. Robotic identification of kinesthetic deficits after stroke. **Stroke**, v. 44, n. 12, p. 3414–3421, Dec 2013. Citations on pages 65, 116, 125, 131, 132, and 139.

SHADMEHR, R.; WISE, S. **A minimum-jerk trajectory. Supplementary documents for “Computational Neurobiology of Reaching and Pointing”**. 2005. Available: <<https://courses.shadmehrlab.org/Shortcourse/minimumjerk.pdf>>. Citation on page 82.

SHEARER, C. The crisp-dm model: The new blueprint for data mining. **Journal of Data Warehousing**, v. 5, n. 4, p. 13–19, 2000. Citations on pages 36 and 38.

Shirley Ryan AbilityLab. **Barthel Index. RehabMeasures Database**. 2016. Available: <<https://www.sralab.org/rehabilitation-measures/barthel-index>>. Citation on page 133.

_____. **Fugl-Meyer Assessment of Motor Recovery after Stroke. RehabMeasures Database**. 2016. Available: <<https://www.sralab.org/rehabilitation-measures/fugl-meyer-assessment-motor-recovery-after-stroke>>. Citation on page 133.

_____. **Wolf Motor Function Test. RehabMeasures Database**. 2016. Available: <<https://www.sralab.org/rehabilitation-measures/wolf-motor-function-test>>. Citation on page 133.

_____. **Manual Muscle Test. RehabMeasures Database**. 2020. Available: <<https://www.sralab.org/rehabilitation-measures/manual-muscle-test>>. Citation on page 133.

SINGH, K.; SCOTT, S. H. A motor learning strategy reflects neural circuitry for limb control. **Nat. Neurosci.**, v. 6, n. 4, p. 399–403, Apr 2003. Citations on pages 56, 57, and 58.

SIVAN, M.; O’CONNOR, R. J.; MAKOWER, S.; LEVESLEY, M.; BHAKTA, B. Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. **J Rehabil Med**, v. 43, n. 3, p. 181–189, Feb 2011. Citation on page 141.

SMITH, L. **A tutorial on Principal Component Analysis**. 2002. Available: <http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf>. Citation on page 40.

SPERRY, R. W. Lateral specialization in the surgically separated hemispheres. In: SCHMITT, F.; WORDEN, F. (Ed.). **Neurosciences Third Study Program**. Cambridge, Massachusetts: MIT Press, 1974. v. 3, p. 5–19. Citation on page 97.

STINEAR, C. M.; BYBLOW, W. D.; ACKERLEY, S. J.; SMITH, M. C.; BORGES, V. M.; BARBER, P. A. PREP2: A biomarker-based algorithm for predicting upper limb function after stroke. **Ann Clin Transl Neurol**, v. 4, n. 11, p. 811–820, 11 2017. Citation on page 149.

STRAUDI, S.; FREGNI, F.; MARTINUZZI, C.; PAVARELLI, C.; SALVIOLI, S.; BASAGLIA, N. tDCS and Robotics on Upper Limb Stroke Rehabilitation: Effect Modification by Stroke Duration and Type of Stroke. **Biomed Res Int**, v. 2016, p. 5068127, 2016. Citations on pages 115, 128, and 129.

TERRANOVA, T. T. **Avaliação da função de membros superiores: desafios na reabilitação física de pacientes pós-AVC**. Master’s Thesis (Master’s Thesis) — Faculdade de Medicina, Universidade de São Paulo., 2016. Citation on page 174.

TOZLU, C.; EDWARDS, D.; BOES, A.; LABAR, D.; TSAGARIS, K. Z.; SILVERSTEIN, J.; LANE, H. P.; SABUNCU, M. R.; LIU, C.; KUČEYESKI, A. Machine Learning Methods Predict Individual Upper-Limb Motor Impairment Following Therapy in Chronic Stroke. **Neurorehabil Neural Repair**, v. 34, n. 5, p. 428–439, 05 2020. Citations on pages 143, 145, and 149.

TRUNK, G. V. A problem of dimensionality: A simple example. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-1, n. 3, p. 306–307, July 1979. ISSN 0162-8828. Citation on page 43.

TSUJI, T.; KURITA, N.; YAMADA, M.; SAKAINO, S.; KANEKO, Y. Rehabilitation database based on haptic signal processing. In: **2013 IEEE Region 10 Humanitarian Technology Conference**. [S.l.: s.n.], 2013. p. 13–18. ISSN null. Citation on page 66.

VAPNIK, V. N. **The Nature of Statistical Learning Theory**. New York: Springer Verlag, 1995. Citations on pages 44 and 45.

WANG, J.; YU, L.; WANG, J.; GUO, L.; GU, X.; FANG, Q. Automated fugl-meyer assessment using svr model. In: **2014 IEEE International Symposium on Bioelectronics and Bioinformatics (IEEE ISBB 2014)**. [S.l.: s.n.], 2014. p. 1–4. Citation on page 148.

WANG, L.; LI, H.; WANG, Z.; MENG, F. Study on upper limb rehabilitation system based on surface EMG. **Biomed Mater Eng**, v. 26 Suppl 1, p. 795–801, 2015. Citation on page 44.

WARE, C. **Information Visualization: Perception for Design**. 2nd. ed. San Francisco: Morgan Kaufmann, 2004. Citation on page 36.

WELLING, M. **Fisher Linear Discriminant Analysis**. 2009. Available: <<https://www.ics.uci.edu/~welling/teaching/273ASpring09/Fisher-LDA.pdf>>. Citations on pages 40 and 41.

WHO. **World Health Statistics**. 2012. Available: <http://www.who.int/gho/publications/world_health_statistics/EN_WHS2012_Full.pdf>. Citation on page 31.

WOODWORTH, R. S. **The accuracy of voluntary movement ...** Phd Thesis (PhD Thesis), New York, 1899. Citations on pages 73 and 88.

YAO, Q.; TIAN, Y.; LI, P.-F.; TIAN, L.-L.; QIAN, Y.-M.; LI, J.-S. Design and development of a medical big data processing system based on hadoop. **Journal of Medical Systems**, v. 39, n. 3, p. 23, Feb 2015. ISSN 1573-689X. Available: <<https://doi.org/10.1007/s10916-015-0220-8>>. Citation on page 66.

YOSHIKAWA, T. **Foundations of Robotics: Analysis and Control**. Cambridge, MA, USA: MIT Press, 1990. ISBN 0262240289. Citation on page 84.

ZOLLO, L.; GALLOTTA, E.; GUGLIELMELLI, E.; STERZI, S. Robotic technologies and rehabilitation: new tools for upper-limb therapy and assessment in chronic stroke. **Eur J Phys Rehabil Med**, v. 47, n. 2, p. 223–236, Jun 2011. Citations on pages 132, 139, and 140.

MATHEMATICAL PERSPECTIVE ON THE SPREAD OF CLINICAL SCORES

Apart from the clinical debate on high predictability of clinical scores from chronic outpatients, it is also necessary to employ a mathematical perspective on the distribution of data points with respect to their respective clinical scores. In this paper we use the correlation between model prediction and the expected output to evaluate the goodness of fit of a model, given by R^2 , defined as follows:

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{VAR_{Model}}{VAR_{Total}} \quad (A.1)$$

where VAR_{Total} is the variance, i.e., mean squared distances between every i -th point y_i and the mean $\bar{y}$, whereas VAR_{Model} is the mean squared residual with respect to a model fitted to the dataset.

From a linear perspective, for example, we can consider both VAR_{Total} and VAR_{Model} as mean squared residuals arising, respectively, from a model with no slope (the mean) and a model with a slope adjusted to the data. Therefore, VAR_{Total} is always larger because a model non adjusted to data (no slope) yields larger residuals, whereas VAR_{Model} yields lower residuals because it is adjusted to the data. That said, the fraction tends to be smaller, as the variance of data points increases. Hence, R^2 tends to 1.

In practical terms, a shorter window of clinical scores may yield lower values of R^2 in comparison to the full range of a scale.

NONLINEAR MODELS FOR ESTIMATING THE FMA-UE

For both nonlinear models (Tables 12 and 13), an ensemble of the four models, in terms of mean, must yield the final output.

Table 12 – Weights of the ensemble of MLP neural networks that constitute the nonlinear model for the FMA-UE correlation, given S/E data.

Layer/node	Weight	Model 1	Model 2	Model 3	Model 4
L2N1	w0	2.651745	1.358790	-2.288135	-2.623445
	w1	1.705711	0.178818	-2.668738	1.231372
	w2	0.196343	1.322677	-0.126410	-0.673580
	w3	-0.025334	-0.208155	0.401209	-0.224087
	w4	-2.059048	0.651006	1.131175	1.364455
	w5	-0.169044	-1.014909	0.328886	-1.796512
	w6	-0.522972	0.544617	0.817448	-0.337860
L2N2	w0	0.209553	-3.869454	-2.525723	-1.880432
	w1	0.066768	-0.855441	0.893196	-2.088235
	w2	0.702611	-0.128405	-0.075536	0.038249
	w3	-0.884474	-0.065442	-0.597529	0.201751
	w4	0.841676	1.421780	0.883846	1.057567
	w5	-0.757341	-1.417613	-0.720293	0.442159
	w6	0.289868	0.302277	-0.532560	0.835733
L3N1	w0	0.458473	1.156568	5.558105	1.241911
	w1	-1.058998	0.206183	1.041446	1.055794
	w2	0.346208	2.015478	5.429382	1.104092

Source: Elaborated by the author.

Table 13 – Weights of the ensemble of MLP neural networks that constitute the nonlinear model for the FMA-UE correlation, given S/E and wrist data.

Layer/node	Weight	Model 1	Model 2	Model 3	Model 4
L2N1	w0	0.100405	-0.499624	0.917564	0.287330
	w1	0.412031	-0.644570	-0.910925	0.089558
	w2	-0.256844	-0.515516	0.244053	-0.012349
	w3	0.359730	0.291215	-0.495911	0.254710
	w4	0.224924	-0.271208	0.197361	-0.032417
	w5	-0.610725	0.488541	1.052910	-0.033304
	w6	-0.125555	-0.297185	-1.170956	-0.019990
	w7	1.204007	0.564216	0.451516	0.403285
	w8	-0.295050	0.090108	-0.376013	-0.124665
	w9	-0.654584	-0.471420	0.276252	-0.199220
	w10	0.464651	0.788101	0.134621	0.363913
	w11	-0.292340	-0.168995	-0.226710	-0.002602
L2N2	w0	0.675229	-1.719442	-0.220476	3.099416
	w1	0.628326	-0.590140	0.184195	-1.748653
	w2	0.450943	0.436204	0.367614	-0.897005
	w3	-0.306028	-0.832092	-0.235570	1.395623
	w4	0.905761	0.254605	0.032158	0.893904
	w5	-0.537267	-0.044102	0.039324	1.244433
	w6	0.042878	0.246911	-0.028137	-1.058052
	w7	-0.140713	-1.507550	-0.409774	0.778307
	w8	0.132989	1.317929	-0.142964	-0.593424
	w9	0.214891	1.315244	0.566939	-1.379620
	w10	-0.636054	-0.210806	-0.582138	-1.723260
	w11	-0.242302	0.231229	-0.271664	0.015366
L3N1	w0	-0.123029	-0.104083	0.193950	0.444519
	w1	-0.617778	-0.491291	-0.248219	-1.123216
	w2	0.263766	0.375458	0.826389	-0.398329

APPROVAL FROM CAPPESQ ETHICS COMMITTEE

Figure 51 shows the approval from Comissão de Ética para Análise de Projetos de Pesquisa (CAPPesq), related to the data used in Chapter 6.

For Chapters 7, 8 and 9 the study was approved¹ by the Institutional Review Board at Burke Rehabilitation Hospital (BRC 426), MIT-COUHES (Committee on the Use of Humans as Experimental Subjects, 1203004976) and Feinstein Institute for Medical Research IRB (12-102B). Participants provided written informed consent.

---

¹ ClinicalTrials.gov Identifier: NCT03562663. More details available at <<https://www.clinicaltrials.gov/ct2/show/NCT03562663>>.

Figure 51 – Approval from CAPPesq ethics committee.

Hospital das Clínicas da FMUSP
Comissão de Ética para Análise de Projetos de Pesquisa
CAPPesq

Nº Protocolo: 0916/11

Título: EFEITO DO APRENDIZADO MOTOR E TREINAMENTO ROBÓTICO DE MEMBROS SUPERIORES SOBRE A NEUROPLASTICIDADE E CAPACIDADE FUNCIONAL: ESTUDO CLÍNICO RANDOMIZADO CEGO EM SEQUÊLA DE ACIDENTE VASCULAR ENCEFÁLICO – ESTUDO NARLEI

Pesquisador Responsável: Linamara Rizzo Battistella

Pesquisador Executante: Fabio Marcon Alfieri

Co-autores: Marta Imamura; Thais Tavares Terranova; Michele Shultz; Daniel Gustavo Goroso; Felipe Fregni; Gracinda Tsukimoto; Mário Pedrazzoli Neto; Ayrton Massaro

Departamento: MEDICINA LEGAL, ÉTICA MÉDICA, MEDICINA SOCIAL E DO TRABALHO

A Comissão de Ética para Análise de Projetos de Pesquisa – CAPPesq da Diretoria Clínica do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, **APROVOU / TOMOU CIÊNCIA** na sessão datada de 01/02/2012, o protocolo acima.

A CAPPesq em obediência à Resolução CNS 196/96, solicita ao pesquisador (a) a elaboração de relatório parcial e final.

No caso de relatório parcial é necessário informar o tempo previsto para a conclusão do protocolo e breve resumo dos resultados obtidos.

CAPPesq, 06 de Fevereiro de 2012

PROF. DR. LUIZ EUGÊNIO GARCEZ LEME
Coordenador
Comissão de Ética para Análise de
Projetos de Pesquisa - CAPPesq

Rua Dr. Ovídio Pres de Campos, 225 - Prédio da Administração - 5º andar - CEP 05403-010 - São Paulo - SP.
Fone: 55 11 2661-6442 ramais 16, 17, 18 e 20 - e-mail: cappesq@hcnetsp.usp.br

Source: Terranova (2016).