

Predicting Disease Progression in Amyotrophic Lateral Sclerosis using Machine Learning:

Learning from Longitudinal Data using Time Windows and Progression Patterns

SARA C. MADEIRA

NEUROCLINOMICS2

Unravelling Prognostic Markers in
NEUROdegenerative diseases through
CLINical and OMICS data integration

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

PTDC/EEI-SSI/1937/2014 (Jan 2016-Dec 2018)



Challenges of Biomarker discovery in Neurodegenerative diseases

Large collections of data available

BRAINnet

ADNI

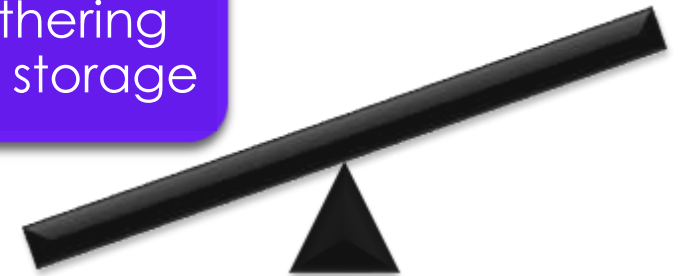
ProACT

Other
public
repositories

3

Data
gathering
and storage

Integrative
knowledge
discovery





Integrative biomarker discovery in neurodegenerative diseases

André V. Carreiro,¹ Alexandre Mendonça,² Mamede de Carvalho³
and Sara C. Madeira^{1*}

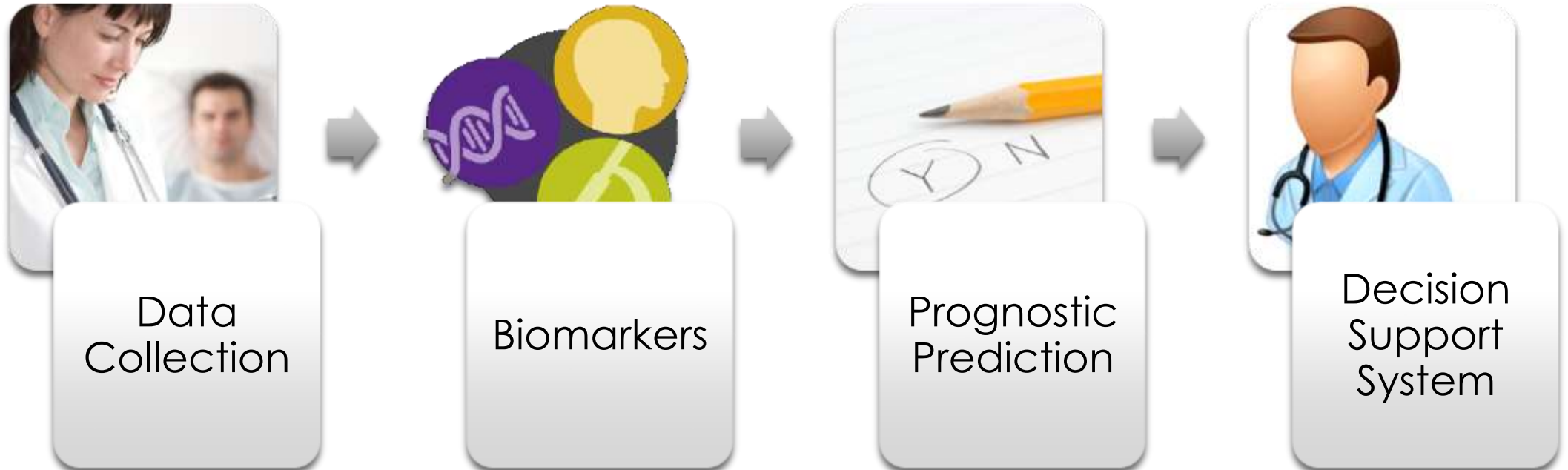
Data mining has been widely applied in biomarker discovery resulting in significant findings of different clinical and biological biomarkers. With developments in technology, from genomics to proteomics analysis, a deluge of data has become available, as well as standardized data repositories. Nonetheless, researchers are still facing important challenges in analyzing the data, especially when considering the complexity of pathways involved in biological processes and diseases. Data from single sources appear unable to explain complex processes, such as those involved in brain-related disorders, including Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis, thus raising the need for a more comprehensive perspective. A possible solution relies on data and model integration, where several data types are combined to provide complementary views. This in turn can result in the discovery of previously unknown biomarkers by unraveling otherwise hidden relationships between data from different sources, and/or validate such composite biomarkers in more powerful predictive models. © 2015 Wiley Periodicals, Inc.

How to cite this article:

WIREs Syst Biol Med 2015. doi: 10.1002/wsbm.1310

Need for Biomarkers and Predictive Models (Data Mining/Machine Learning)

5



Need for Biomarkers and Predictive Models (Data Mining/Machine Learning)

6



Disease
prognostic
markers

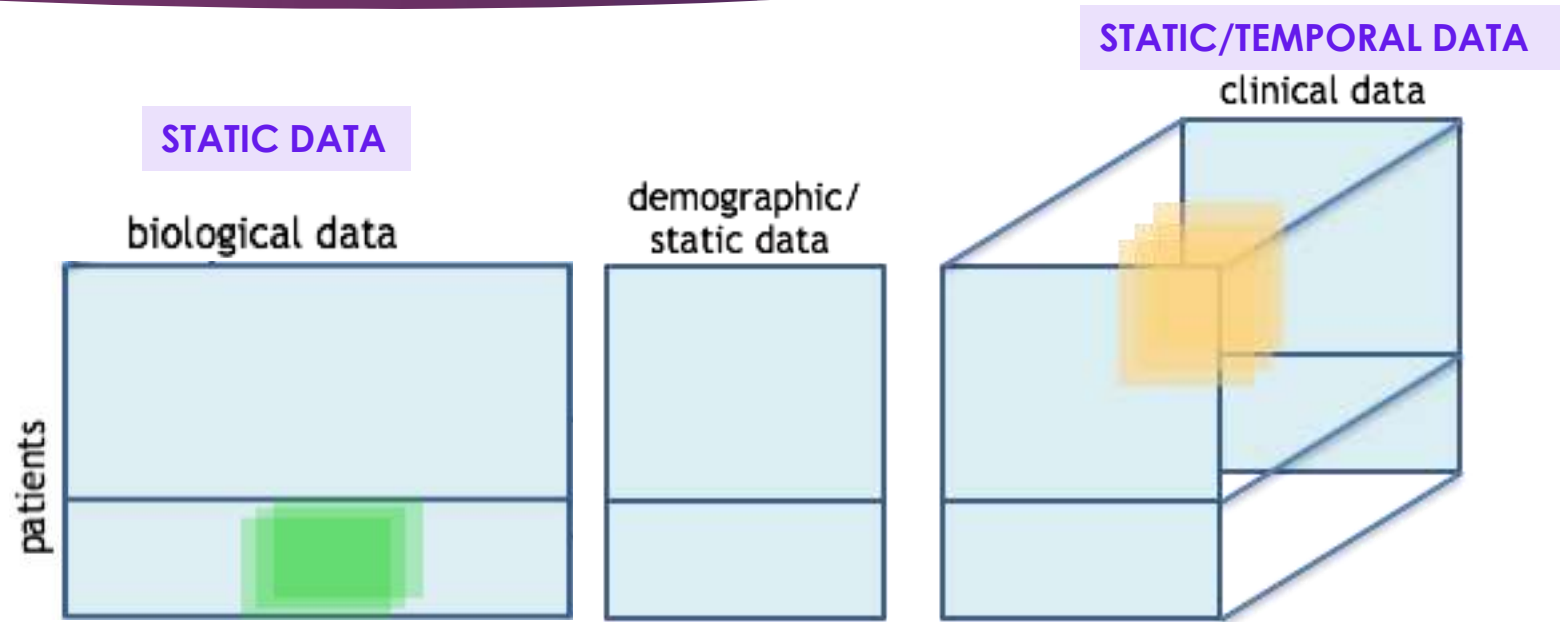
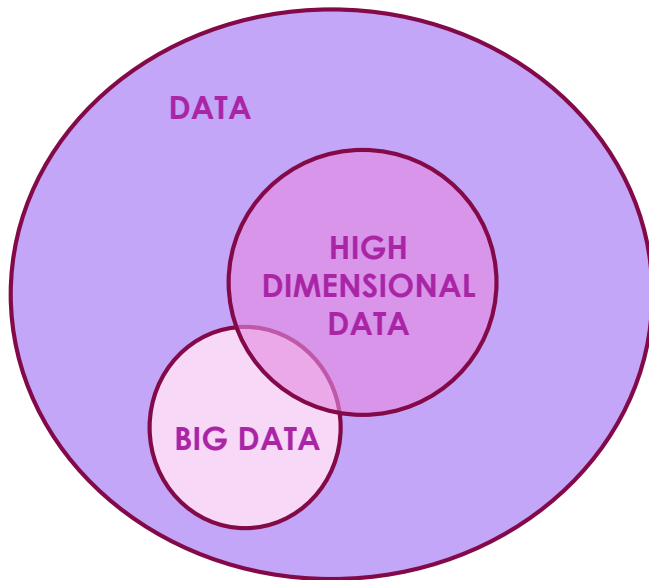


Disease
progression
rates



Patient profiles

High Dimensional and Heterogeneous Data Analysis



What type of data ****do we have**** ?

***** WHAT HAVE WE BEEN DOING ? *****

@NEUROCLINOMICS2

What do we want to do with the data ?

- Clustering
- Classification
- Pattern Discovery

OF PATIENTS

PROGNOSTIC PREDICTION IN ALS

- PATIENT SNAPSHOTS AND TIME WINDOWS
- PROGRESSION PATTERNS
- PROGRESSION GROUPS

*** WHAT HAVE WE BEEN DOING ? ***

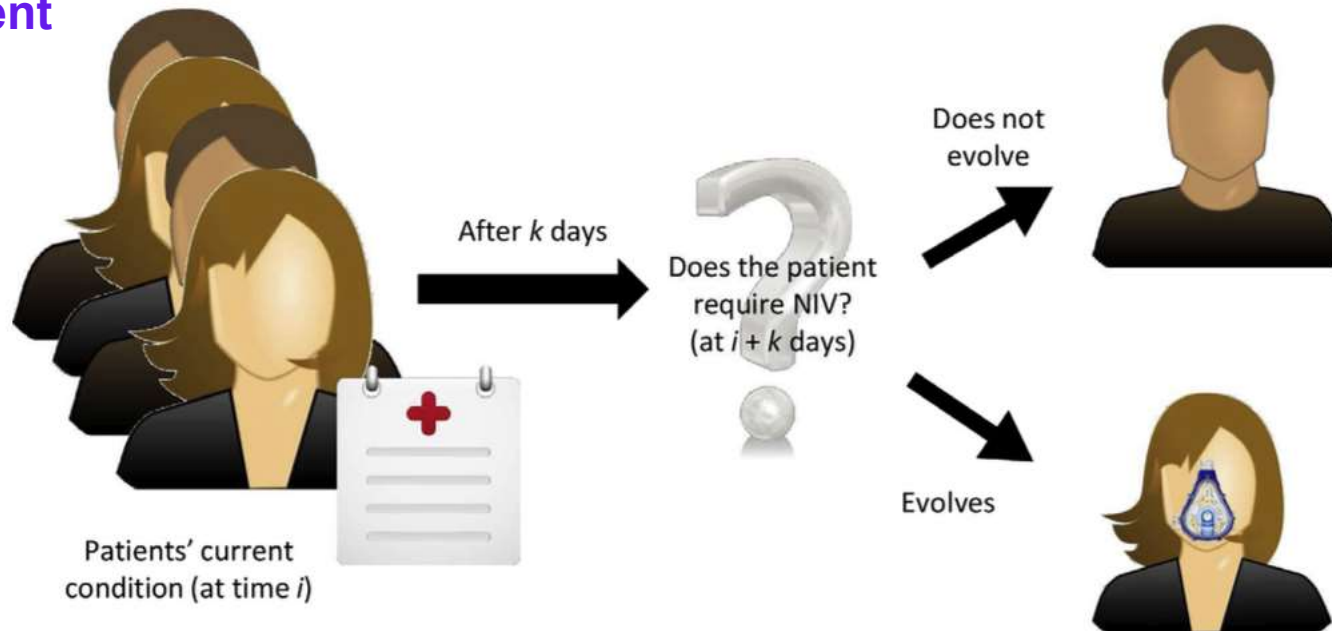
High Dimensional and Heterogeneous Data Analysis

- ▶ **NEUROCLINOMICS2 data** - 1200 national patients, with definite or probable ALS.
 - ▶ Demographic data
 - ▶ Clinical evaluation data
 - ▶ Respiratory tests data
 - ▶ Neurophysiological data
- ▶ **European project OnWebDuals** extended NEUROCLINOMICS2 data with genetic information - 500 Portuguese patients (300 controls) were sequenced to collect detailed genotype-phenotype data.
- ▶ OnWebDuals data with **1300 patients**
 - ▶ complete demographic
 - ▶ clinical (phenotype)
 - ▶ genetic (genotype) information,
 - ▶ enriched with detailed results on laboratory and respiratory function, in addition to family history, professional activity, environmental risk factors (smoking and physical activity), and survival

What data **do we have** ?

Prognostic Models using Patient Snapshots and Time Windows - One Snapshot

Using One Snapshot
(Patient's Current
Condition)



Time Window (k days)
90 days (next appointment)
180 days
360 days

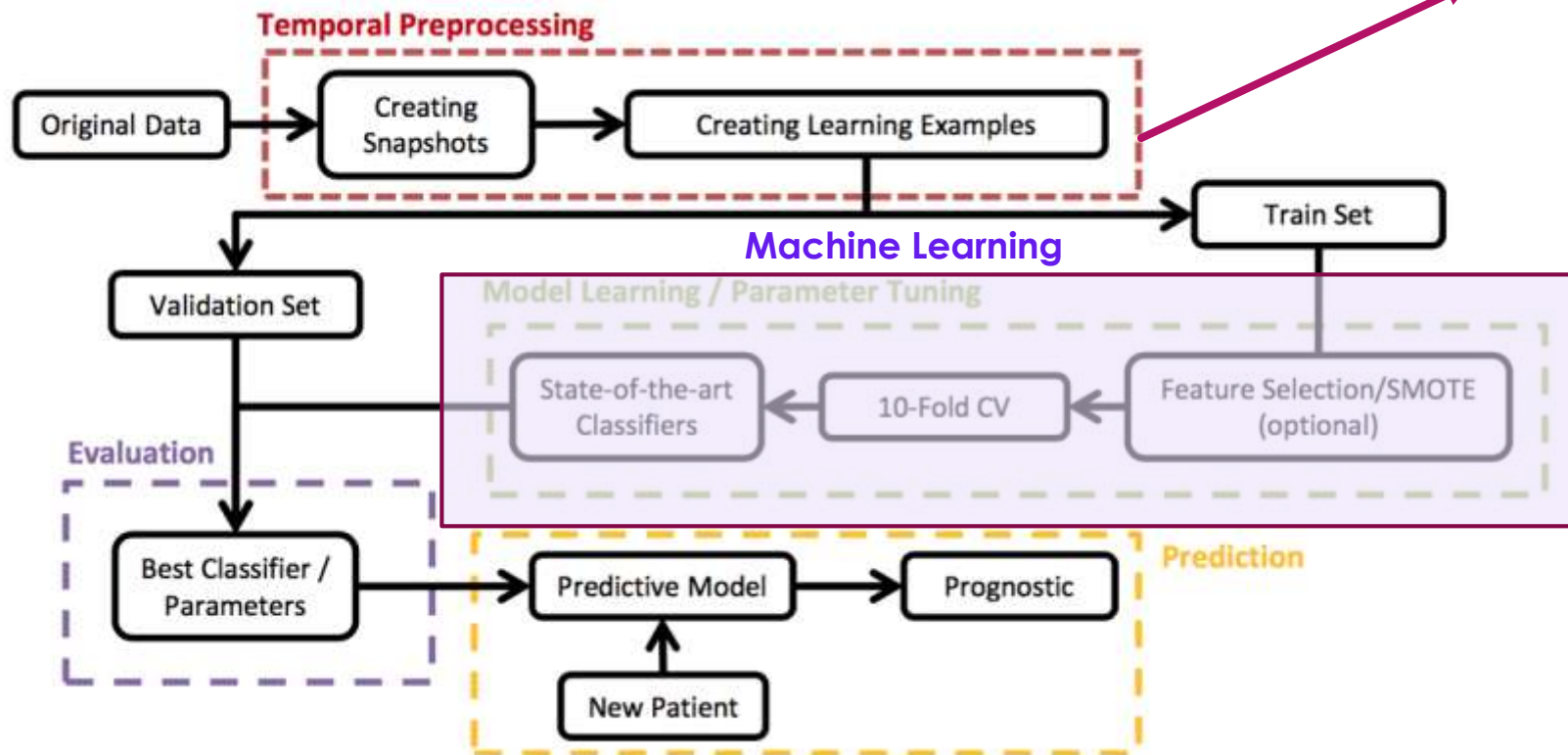
Can we predict if a patient with ALS will require NIV k days from today?

Prognostic Models using Patient Snapshots and Time Windows

11

Original Data								
ID	Date 1	TestA 1	TestB 1	...	Date N	TestA N	TestB N	NIV Date
1	01/01/2010	27	1		10/07/2011	12	0	20/04/2011
2	15/03/2010	30	1		25/06/2011	24	1	Not Applied
...								
P	21/02/2010	28	1		08/10/2011	13	0	14/08/2011

Pat ID	Snapshot	Date (j days)	Test 1	...	Test E	Evolution (E)
1	1	j = i				0
...						
P	1	j = i				1





Prognostic models based on patient snapshots and time windows: Predicting disease progression to assisted ventilation in Amyotrophic Lateral Sclerosis



André V. Carreiro^{a,*}, Pedro M.T. Amaral^a, Susana Pinto^b, Pedro Tomás^a, Mamede de Carvalho^b, Sara C. Madeira^{a,*}

^a INESC-ID Lisbon and Instituto Superior Técnico, Universidade de Lisboa, Portugal

^b Translational Clinical Physiology Unit, Institute of Molecular Medicine, Faculty of Medicine, Universidade de Lisboa, Portugal

ARTICLE INFO

Article history:

Received 19 February 2015

Revised 18 September 2015

Accepted 23 September 2015

Available online 8 October 2015

Keywords:

Prognostic model

Disease progression

Amyotrophic Lateral Sclerosis

Time windows

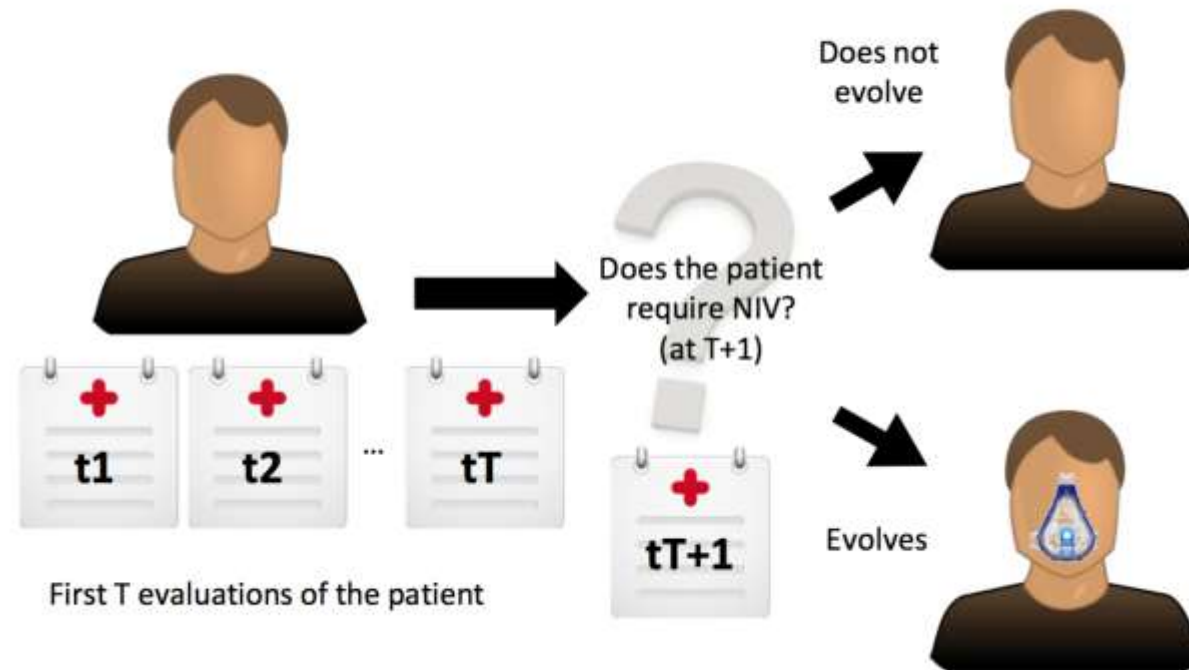
Patient snapshots

ABSTRACT

Amyotrophic Lateral Sclerosis (ALS) is a devastating disease and the most common neurodegenerative disorder of young adults. ALS patients present a rapidly progressive motor weakness. This usually leads to death in a few years by respiratory failure. The correct prediction of respiratory insufficiency is thus key for patient management. In this context, we propose an innovative approach for prognostic prediction based on patient snapshots and time windows. We first cluster temporally-related tests to obtain snapshots of the patient's condition at a given time (patient snapshots). Then we use the snapshots to predict the probability of an ALS patient to require assisted ventilation after k days from the time of clinical evaluation (time window). This probability is based on the patient's current condition, evaluated using clinical features, including functional impairment assessments and a complete set of respiratory tests. The prognostic models include three temporal windows allowing to perform short, medium and long term prognosis regarding progression to assisted ventilation. Experimental results show an area under the receiver operating characteristics curve (AUC) in the test set of approximately 79% for time windows of 90, 180 and 365 days. Creating patient snapshots using hierarchical clustering with constraints outperforms the state of the art, and the proposed prognostic model becomes the first non population-based approach for prognostic prediction in ALS. The results are promising and should enhance the current clinical practice, largely supported by non-standardized tests and clinicians' experience.

Prognostic Models using Patient Snapshots and Time Windows – Set of Snapshots (Follow-Up)

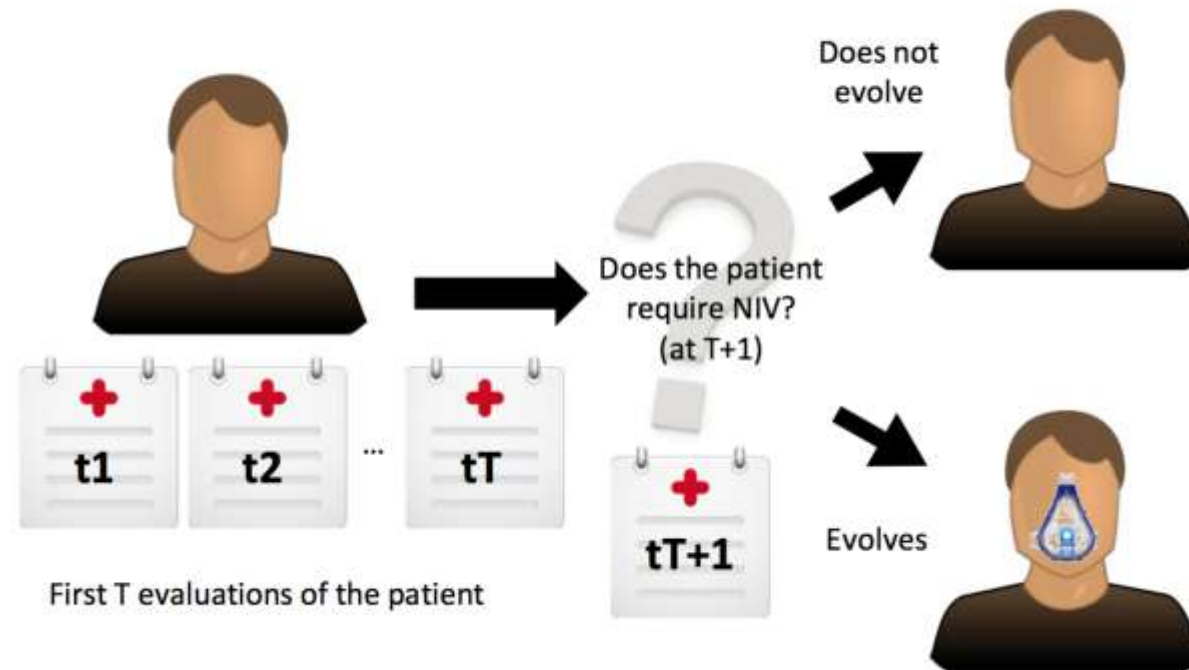
Using a Set of Snapshots
(Follow-up data)



Can we improve prediction if we use T evaluations of the patient ?

Prognostic Models using Patient Snapshots and Time Windows – Set of Snapshots (Follow-Up)

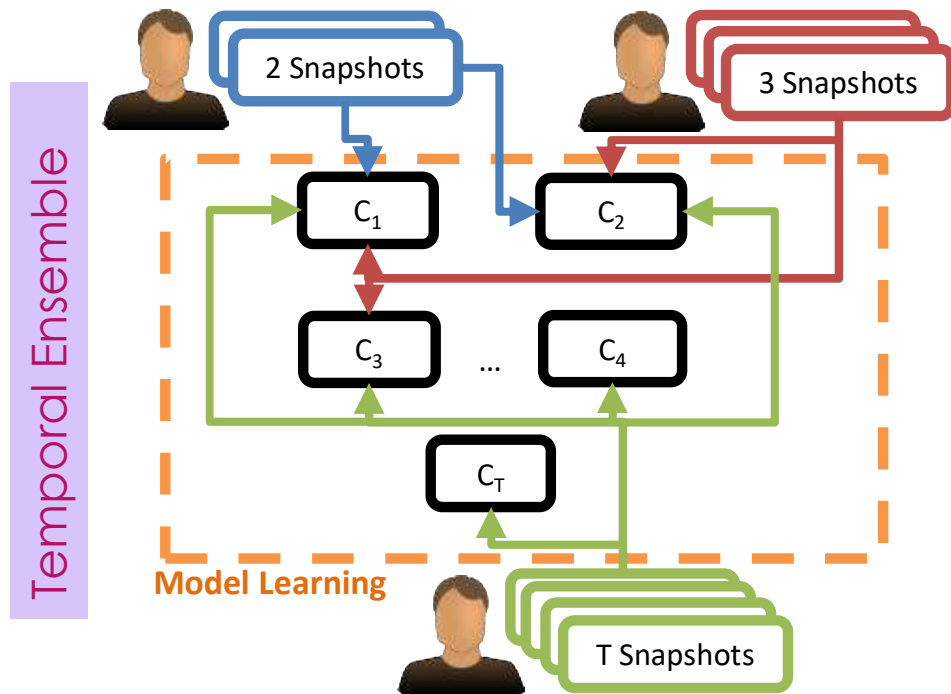
Using a Set of Snapshots (Follow-up data)



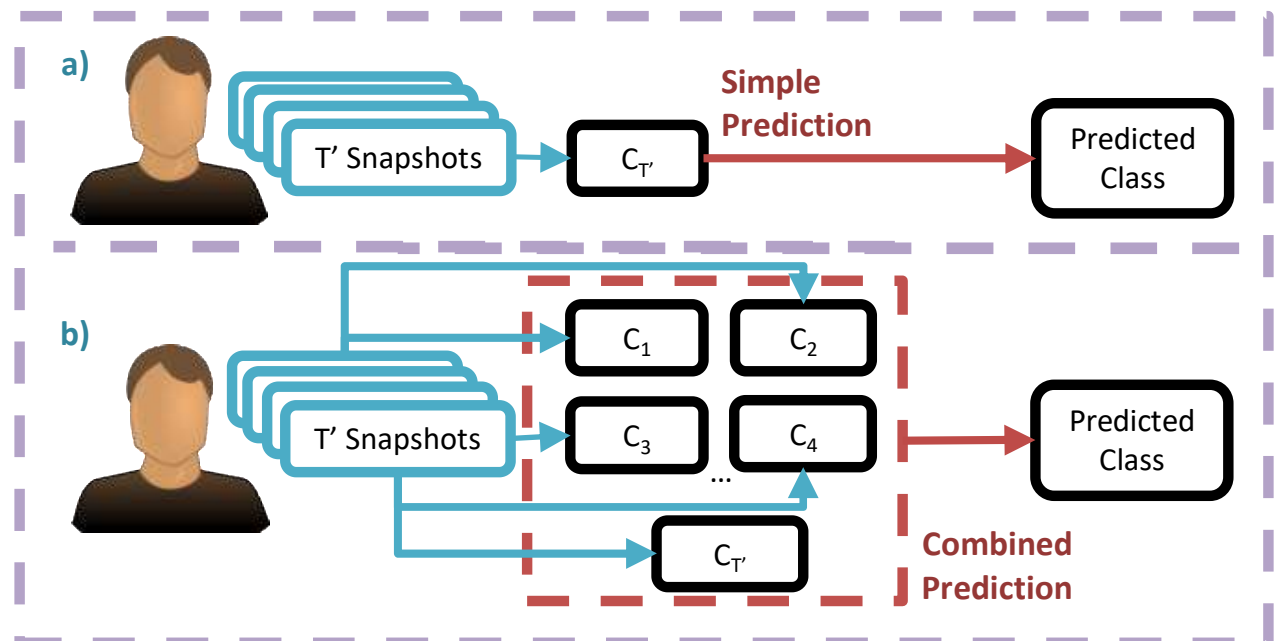
Can we improve prediction if we use T evaluations of the patient ?
Can we improve prediction if we use progression patterns ?

Prognostic Models using Patient Snapshots and Time Windows – Set of Snapshots (Follow-Up)

Can we improve prediction if we use T evaluations of the patient ?



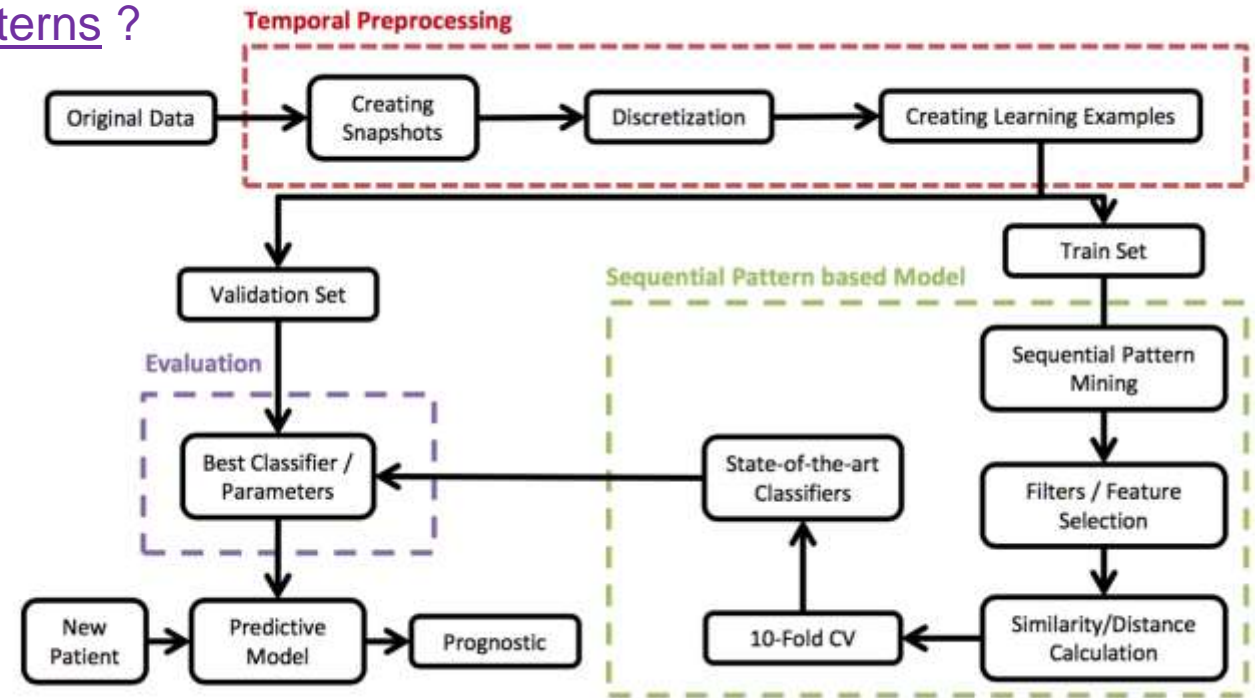
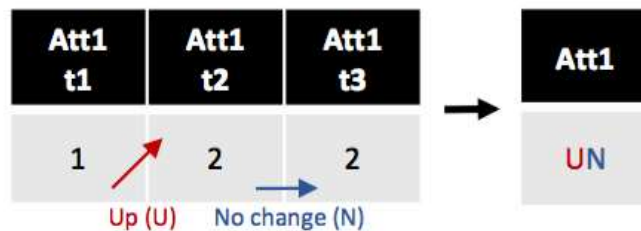
New Prediction for Variable T



Prognostic Models using Patient Snapshots and Time Windows – Set of Snapshots (Follow-Up)

Can we improve prediction if we use progression patterns ?

ID	Att1 t1	Att2 t1	...	Att1 tT	Att2 tT	Evolution
1						1
2						0
...			...			
P						1

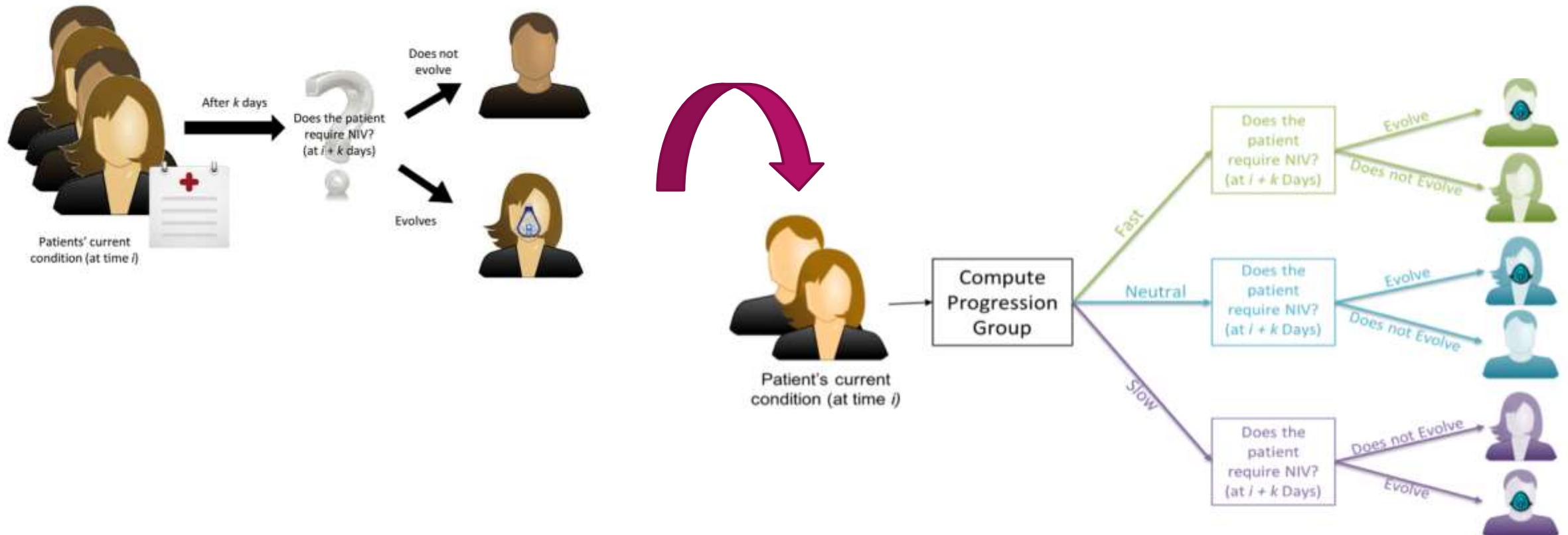


Using Frequent Sequential Patterns as

Using Temporal Dynamic Patterns as Featu

(Onset form=Spinal) (ALSFRSb=[9.6-inf]) (ALSFRSb=[9.6-inf])

Prognostic Models using Progression Groups



Can we improve prediction if we use progression groups and group-specific predictive models ?

Predicting Non-Invasive Ventilation in ALS Patients using Stratified Disease Progression Groups

Sofia Pires*

LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal
sofiaferropires@gmail.com

Marta Gromicho

Instituto de Medicina Molecular, Instituto de Fisiologia
Faculdade de Medicina, Universidade de Lisboa
Lisbon, Portugal
martalgms@gmail.com

Susana Pinto

Instituto de Medicina Molecular, Instituto de Fisiologia
Faculdade de Medicina, Universidade de Lisboa
Lisbon, Portugal
susana.c.pinto@medicina.ulisboa.pt

Mamede de Carvalho

Instituto de Medicina Molecular, Instituto de Fisiologia
Faculdade de Medicina, Universidade de Lisboa
Lisbon, Portugal
Department of Neurosciences and Mental Health
Hospital de Santa Maria CHLN
Lisbon, Portugal
mamedealves@medicina.ulisboa.pt

Sara C. Madeira*

LASIGE, Faculdade de Ciências, Universidade de Lisboa
Lisbon, Portugal
sacmadeira@ciencias.ulisboa.pt

Abstract—Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease highly known for its rapid progression, leading to death usually within a few years. Respiratory failure is the

over 10 years [3]. This aspect of the disease hinders our understanding of it, making it difficult to provide early diagnosis and develop treatments based on disease progression

Ongoing and Future Work

- WORKING ON **PROGNOSTIC MODELS**
 - Patient snapshots and time windows
 - + Progression patterns
 - + Progression groups
- STUDY PROGNOSTIC MODELS USING **OTHER PROGNOSTIC OUTCOMES** (DECLINE IN ALSFRS, ETC)
- **INTEGRATE GENOTYPE-PHENOTYPE DATA** (FROM PATIENTS AND JPND PROJECT “ONWEBDUALS – ONTOLOGY-BASED WEB DATABASE FOR UNDERSTANDING AMYOTROPHIC LATERAL SCLEROSIS”) **WITH CLINICAL DATA**
- **IDENTIFY DISEASE PROGRESSION PATTERNS** USING **LONGITUDINAL DATA** FROM PATIENTS’ FOLLOW-UP

Team (Past and Current)



Sara C. Madeira



Mamede de Carvalho + his team
(Susana Pinto, **Marta Gomiche**)



Pedro Amaral



André Carreiro



Sofia Pires

Manuel Figueiral, MSc
Joana Matos, MSc student)

<https://saracmadeira.wordpress.com>

sacmadeira@ciencias.ulisboa.pt

<http://www.lasige.di.fc.ul.pt>





Ciências
ULisboa



driven by
excellence

