Master of Science Topics

e-Health

Title: The Virtual Patient

Coordinators: *Prof. Adina Magda Florea (adina.florea@upb.ro)*
Dr. Alex Awada (alex.awada@upb.ro)

Tasks:

- Develop an intelligent agent and a virtual patient to provide students in the medical domain with an interactive class, to cope for their limited ability to attend live classes during the current pandemic.
- Develop a virtual on-line class.
- We want to create a framework system of interaction between the student and the patient, which is developed initially for the symptoms of pulmonary thromboembolism (PET), and then, after validation of this use case, to be extended to other symptoms and to make a complex diagnosis depending on the symptoms.
- Collaboration with prof. med dr. Alexandru Scafa Udriște (UMF and Spitalul de Urgență Floreasca).

Title: Using GANs to generate cardiology disease case descriptions

Coordinator: *Prof. Adina Magda Florea (adina.florea@upb.ro)*
Dr. Alex Awada (alex.awada@upb.ro)

Description:

General Adversarial Networks (GANs) are used for generating new, synthetic realistic and consistent data. Introduction in 2014 [1], many new GAN-based models have been proposed since to overcome some of the difficulties of the original model such as vanishing gradient, difficulty in training, poor diversity and stability. GANs have proven to play an important role in generating images and have shown good performance in generating continuous data. Considering the fact that deep neural networks require very large quantities of data for a good training, GANs are a good option to generate synthetic data in the medical domain to create new or enlarge existing data sets and guarantee anonymity (it has been said that anonymized medical data come sometimes be traced back). Since the gradient function is required to be differentiable, the traditional GAN cannot generate discrete data. For example, the diagnosis of a disease and the severity of the disease are discrete data. Different proposals have been made in this respect. In [2] the authors generated synthetic electronic health records (EHR) by using medical Generative Adversarial Networks (medGAN) based on the Sutter Palo Alto Medical Foundation and the Medical Information Mart for Intensive Care datasets. The task is to use GANs to produce synthetic data for cardiology diseases and cases.

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio. Generative Adversarial Networks, arXiv:1406.2661, 2014

[2] E. Choi, S. Biswal, B. Malin, J. Duke, W. Stewart, J. Sun. Generating Multi-label Discrete Patient Records using Generative Adversarial Networks. In Proc. of Machine Learning for Healthcare, JMLR W&C Track Volume 68, 2017.

Title: Classifying Gait Patterns from wearable sensors using Deep Learning methods

Coordinator: *Assist. Prof. Alexandru Sorici (alexandru.sorici@upb.ro)*

Description:

In the treatment and monitoring of various neurological diseases (e.g., Parkinson's Disease - PD, Multiple Sclerosis - MS, Stroke Recovery) the classification of walking / gait patterns and the quantification of balance issues are important aspects of the disease monitoring process and the adjustment of medication or physical rehabilitation treatments. In order to support such monitoring over a longer period of time, outside of clinical labs, the use of wearable sensors (e.g. smart watches; smart bracelets that can fitted on the wrist, thigh or waist; smart insoles) is promoted as a means to extend disease monitoring at home and to enable a continuous assessment of disease progression / regression.

The research topic starts investigation from existing datasets (e.g. WhuGAIT), but will also aim to build a new dataset as part of the ALAMEDA project which will take input from a smartphone, a Fitbit smartwatch, a smart bracelet (e.g. GENEActiv Original) and Loadsol smart insole devices.

Specific research directions will involve:

- Classification of balancing issues based on standardized medical tests for Stroke mobility assessment
- Classification of walking patterns for Stroke, MS and PD patients based on standardized medical tests
- Applying fine-tuning / domain adaptation and few shot learning techniques to time series classification models with variability given by:
  - intra-patient prediction performance using the same data input modalities (i.e. the same devices)
  - models trained on a subset of data inputs (e.g. using only the smart watch) and adapted to work on extended input (e.g. smart watch + smart insoles)

Keywords: Activity Recognition, Gait Recognition, Wearable Sensors, Time series Classification, Deep Learning, Fine Tuning / Domain Adaptation

Bibliography

[1] Zou Q, Wang Y, Zhao Y, Wang Q and Li Q, Deep learning-based gait recogntion using smartphones in the wild, IEEE Transactions on Information Forensics and Security, vol. 15, no. 1, pp. 3197-3212, 2020.
[2] Fawaz, H. I., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., ... & Petitjean, F. (2020). Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery, 1-27.

[3] Buffelli, D., & Vandin, F. (2020). Attention-Based Deep Learning Framework for Human Activity Recognition with User Adaptation. arXiv preprint arXiv:2006.03820.

Title: Neurological Disease Monitoring using Objective Measurements and Subjective Patient Reported Data

Coordinator: *Assist. Prof. Alexandru Sorici (alexandru.sorici@upb.ro)*

Description:

In the treatment of various neurological diseases (e.g. Parkinson's Disease - PD, Multiple Sclerosis - MS, Stroke Recovery) it is important to extend monitoring capabilities beyond the clinical visits to enable continuous assessment of disease progression / regression.
Within the context of the ALAMEDA project remote monitoring of patients is proposed, whereby **objective data** (from wearables such as smart watches, smart insoles and smart bracelets), as well as subjective **patient reported data** (e.g. questionnaires for mental state, emotional state and limitations in performing activities of daily living).
The objective of this research topic is to evaluate the **capability** of patient acquired data from the above mentioned modalities to **predict** the result of future standardized medical tests that assess disease state.
The challenge of the topic lies in analyzing multiple and variable frequency (from hourly, to monthly entries) time series containing both numeric and categorical data.

The research topic starts investigation from existing datasets, similar in design to the previously mentioned setup (e.g. mPower [1], A Long-Term, Real-Life Parkinson Monitoring Database Combining Unscripted Objective and Subjective Recordings [2]), but aims equally to establish a **new dataset** as part of the ALAMEDA project.

Specific research directions will involve:

- Evaluation of multi modal **time series classification** algorithms on the specifics of the ALAMEDA project, with the objective of predicting future medical test outcomes.
- Analyzing correlation between objective sensor measures and subjective patient reported data.
- Analyzing intra-patient self-predictive capabilities of collected data (i.e. **time series prediction problem** aiming to estimate the future values of both objective and subjective measures).

Keywords: Multivariate time series classification, Multivariate time series prediction, Data Science, Machine Learning, Wearable Sensors, Medical Questionnaires.

Bibliography

[1] Bot, Brian M., Christine Suver, Elias Chaibub Neto, Michael Kellen, Arno Klein, Christopher Bare, Megan Doerr et al. "The mPower study, Parkinson disease mobile data collected using ResearchKit." *Scientific data* 3, no. 1 (2016): 1-9.
[2] Habets, Jeroen GV, Margot Heijmans, Albert FG Leentjens, Claudia JP Simons, Yasin Temel, Mark L. Kuijf, Pieter L. Kubben, and Christian Herff. "A Long-Term, Real-Life

Parkinson Monitoring Database Combining Unscripted Objective and Subjective Recordings." *Data* 6, no. 2 (2021): 22.

Title: Domain-specific multi-word extraction and abbreviation disambiguation for e-Health

Coordinators: *S.L. Dr. Ing. Ciprian-Octavian Truică (ciprian.truica@upb.ro)*
Conf. Dr. Ing. Elena-Simona Apostol (elena.apostol@upb.ro)

Description:
Hospitals collect huge amounts of data about their patients every year, in various ways. Such massive sets of data can provide great knowledge and information which can improve the medical services, and overall the healthcare domain, such as disease prediction by analyzing the patient's symptoms or disease prevention, by facilitating the discovery of behavioral factors that can turn into risk factors for disease. Unfortunately, only a relatively small volume of e-Health data is processed and interpreted, an important factor being the difficulty in efficiently performing Big Data operations, and that often this data, even if anonymous, is hard to obtain. This topic aims to design and implement new Natural Language Processing, Machine Learning, and Deep Learning methods and models for disambiguating abbreviations and extracting domain-specific multi-words to better understand medical documents and correlate medical terms with their meaning.
Implementation Language: Python
Libraries: Scikit-learn, PyTorch, Keras with Tensorflow, etc.
Storage & Distribution Technologies: MongoDB, Spark.
No. Students: 1