**Chronic Kidney Disease Classification using Patient Clinical Characteristic Data**

Chronic Kidney Disease (CKD) is a major public health problem with silent disease progression [1]. Gradual decline in kidney function in CKD patients can lead to End Stage Kidney Disease (ESKD) requiring kidney transplants [1]. Early intervention in CKD patients can lead to an improved quality of life by slowing down the progression of CKD and healthcare expenses [1]. Predictive models can aid in identifying patients requiring intensive management of CKD (e.g. through higher doses of blood pressure medication) as well as insights on how to manage and screen outpatients with CKD [3].

Vironix (vironix.ai) specializes in developing consumer validated, cloud-enabled application- programming interface software that – when invoked under smartphones, tablets, computers, monitoring devices, and a variety of institutional HR/Payroll software – delivers on-demand detection and triage of health deterioration. A major challenge in developing these algorithms is access to relevant data. Electronic medical records are proprietary, often inaccessible, and generally contain hoards of incomplete, dirty, and irrelevant data. Public databases that hold genetic, blood, and biomarker type data sources include some explanatory information, but as this data is generally not available to consumers at-home, it can't be used to guide software-based diagnostics in a reasonable consumer setting.

Use of available and synthetically generated patient data offers a chance to develop machine learning models to predict the progression of chronic diseases and obtain insights into their progression [2]. As such, we propose a project to identify and leverage clinical characteristic data in the literature to classify CKD patients. The project will make use of existing patient health data and synthetically generated patient health data to evaluate and develop these models. A particular focus will be on developing models with clinically usable strategies for risk stratification of CKD patient outcome.

**Specific questions/aims:**

1. Identify from scientific literature and/or public databases – a representative set of data that captures the clinical characteristics of patients presenting with CKD. A publicly available dataset to get started is found in [4].
2. Identify a methodology for utilizing that data to predict severe/non-severe presentations of CKD. Vironix has some approaches for an alternate disease, but we would be looking to explore others.
3. Develop and validate a prediction model and show reasonable accuracy, sensitivity, and specificity in CKD.
4. Describe and detail the most important clinical features relevant to predicting CKD patients.
5. Generate synthetic data using the original data and evaluate performance gains obtained when using synthetic data and real data for training and testing the prediction models. One possible way of getting started is by using medGAN [5]. Code for medGAN is publicly available at [6].

**References**

[1] Bai, Q., Su, C., Tang, W. and Li, Y., 2022. Machine learning to predict end stage kidney disease in chronic kidney disease. Scientific reports, 12(1), pp.1-8.

[2] Alaa, A.M. and van der Schaar, M., 2019. Attentive state-space modeling of disease progression. Advances in neural information processing systems, 32.

[3] Xiao, J., Ding, R., Xu, X., Guan, H., Feng, X., Sun, T., Zhu, S. and Ye, Z., 2019. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. Journal of translational medicine, 17(1), pp.1-13.

[4] <https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease>

[5] Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W.F. and Sun, J., 2017, November. Generating multi-label discrete patient records using generative adversarial networks. In *Machine learning for healthcare conference* (pp. 286-305). PMLR.

[6] <https://github.com/mp2893/medgan>

**Additional literature and links:**

Company website: <https://vironix.ai/>

A Machine Learning Methodology for Identification and Triage of Heart Failure Exacerbations  
<https://link.springer.com/article/10.1007/s12265-021-10151-7>  
  
CovidX: Remote Screening, Surveillance, Triage, and Management of Novel Coronavirus <https://vironix.ai/wp-content/uploads/2021/02/CovidX_CoronaVirus_WhitePaper.pdf>

Machine Learning Based Triage Models for Remote Detection and Care of Influenza <https://vironix.ai/wp-content/uploads/2021/02/Vironix_MP2_Influenza_Tech_report_2020-1.pdf>

A machine learning approach to triaging patients with chronic obstructive pulmonary disease <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188532>