Automatic Classification of Long-term Kidney Functions

Santosh Tirunagari[1, 2], Norman Poh[1], Simon de Lusignan[3] Farmer Christopher[4]
1: Department of Computing; 2: Centre for Vision, Speech, and Signal Processing;
3: Clinical Informatics Research Group, University of Surrey. 4: East Kent Hospitals University NHS Foundation Trust.

Introduction

- Screening through thousands of chronic kidney disease (CKD) patients for identifying those who are at risk of progressing into more severe stages is extremely important.
- No automated methods exist to solve this problem.
- This is the first reported attempt that compares automated methods with human annotation.

Potential advantages

- Monitor CKD patients remotely.
- Reduce the human effort (and cost) in labelling the data.
- Improve the understanding of the eGFR trends (Kidney functions).

Importance

- Chronic Kidney Disease (CKD) is a significant cause of morbidity and mortality across the developed world.
- Patients with CKD have increased risk of death from cardiovascular disease and End Stage Kidney Failure, leading to dialysis and kidney transplant.

Cost

- CKD was estimated to cost £1.45 billion in 2009-10.

Prevalence

- 1.8 million people were diagnosed with CKD in England.
- Potentially 900,000 to 1.8 million people with undiagnosed CKD [1].

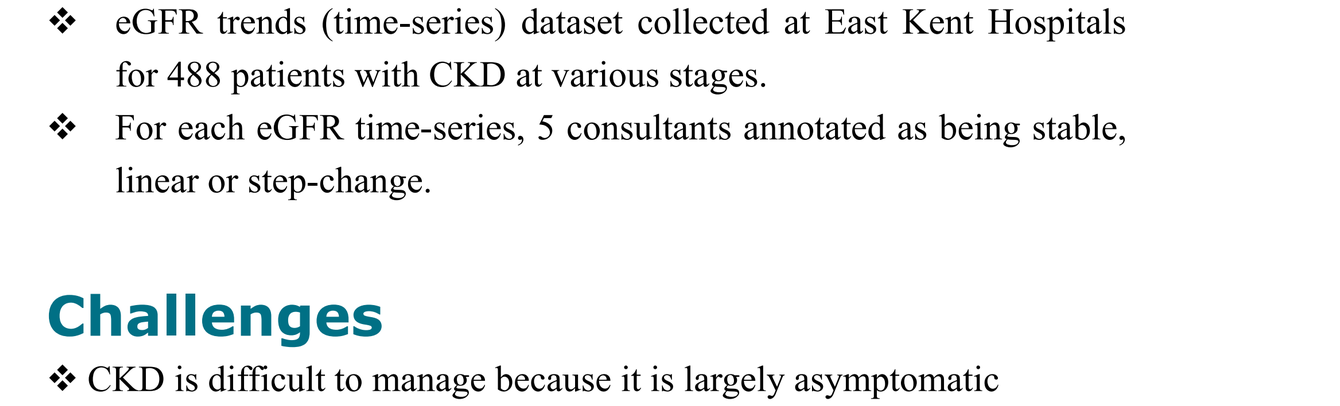
Dataset

- eGFR trends (time-series) dataset collected at East Kent Hospitals for 488 patients with CKD at various stages.
- For each eGFR time-series, 5 consultants annotated as being stable, linear or step-change.

Challenges

- CKD is difficult to manage because it is largely asymptomatic
- Its progression, often quantified via estimated Glomerular Filtration Rate or eGFR, and is difficult to interpret due:
  - the change in assay methods
  - treatment effects
  - person-dependent characteristics
  - quality of data recording
- Long-term eGFR trend not well understood.
- Unequal size of eGFR trends across the age of each patients.

System Architecture



Proposal

- Automatic classification of eGFR trends.
- N=50 and K=5 are found to be optimal via cross-validation.
- σ is found via automatic hyper parameter tuning procedure [2].

Evaluation Methodology

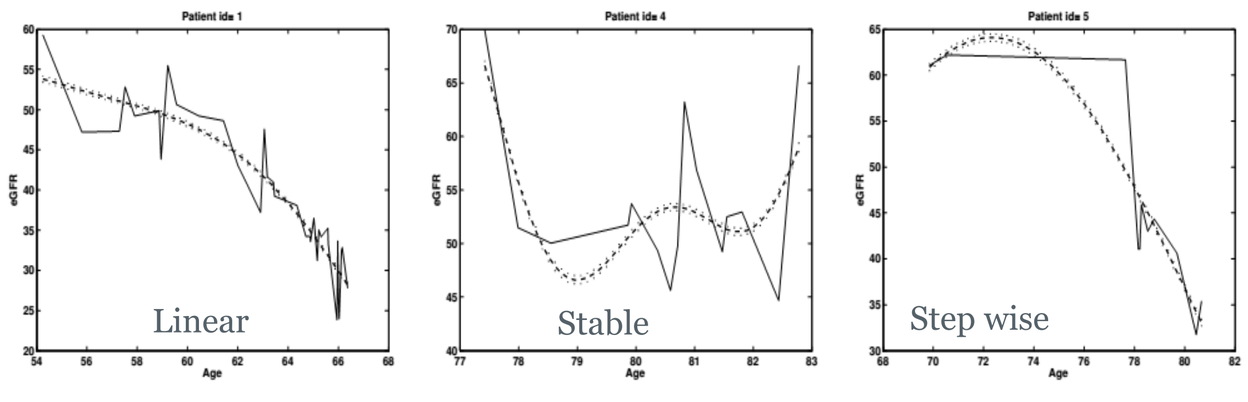
- Human annotations – 5 experts.
- Determination of the ground-truth via maximum vote.
- 5-fold cross-validation whilst maintaining the class priors.

Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Stable	Linear	Linear	Linear	Stable
Linear	Stable	Stable	Stable	Stable
Stable	Stable	Stable	Step	Step

Voting

Ground truth
Linear
Stable
Stable

Equalizing GFR trends



Dashed line: *regression fit ;* Continuous solid line: *eGFR signal*

Recognition Accuracy

Trends	Automated classifier	Human expert
Linear	38.5%	92.3%
Stable	85.7%	98.4%
Stepwise	62.9%	93.7%

Limitations

- Relatively low accuracy compared to human performance
- Better modelling of the eGFR

Future work

- Combine both human and automated annotations in order to improve the accuracy in recognition
- Analyse the impact of the GPR fitting on the classification performance of the trend

Acknowledgements

The funding for this work has been provided by the Medical Research Council (MRC) under the grant no: MR/M023281/1.

References

[1] An NHS Kidney Care report in 2012 (http://goo.gl/oLJD2)
[2] Bishop, Christopher M. *Pattern recognition and machine learning*. Vol. 4, no. 4. New York: springer, 2006.

MRC Medical Research Council
RESEARCH COUNCILS UK
British Renal Society
registered charity no. 1091024