Original Paper

A New Tool Preliminary Assessment on Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER)

Tim Chen[1,2,3,4], Asim Muḥammad[4], Bertrand Chapron[5], C. Y. J. Chen[6,7], & Alexander Babanin[8]

[1] School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

[2] Department of Artificial Intelligence, University of Maryland, Maryland, USA

[3] Faculty of Intelligence Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam

[4] Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

[5] Laboratoire d'Océanographie Physique et Spatiale, Centre de Brest, IFREMER, Plouzané, France

[6] Industrial Technology Research Institute, Hsinchu, Taiwan

[7] BRAC Univ, Department of Computer Science and Engineering, Dhaka, Bangladesh

[8] Department of Infrastructure Engineering, University of Melbourne, Melbourne, Victoria, Australia

Abstract

Patients' comorbidities, operations and complications can be associated with reduced long-term survival probability and increased healthcare utilisation. The aim of this research was to produce an adjusted case-mix model of comorbidity risk and develop a user-friendly toolkit to encourage public adaptation and incremental development.

It has been shown in healthcare research that demographics, temporal dimensions, length-of-stay and time between admissions, can noticeably improve the statistical measures related to comorbidities. The proposed model incorporates temporal aspects, medical procedures, demographics, and admission details, as well as diagnoses.

The research resulted in the development of Temporal-Comorbidity Adjusted Risk of Emergency Readmission (T-CARER) model using routinely collected hospital data.

Keywords

multiple criteria decision making, interval-valued belief distribution, evolutionary algorithm, accuracy, efficiency

1. Introduction

There is increasing evidence that the quantification of high-risk diagnoses, operations and procedures, and monitoring changes over time, can greatly improve the quality of readmission models with adequate adjustment. There have been two streams of work on risk scoring comorbidities to estimate future resource utilisation, emergency admission and mortality.

Firstly, one stream of research looks at the odds ratio of major diagnoses groups and therefore is highly reliant on the whole population statistics. These models stem from crudely summing up the derived weights for comorbidities, which are based on the most recent admission of patients with disregard of temporal patterns. A popular example is the Charlson Comorbidity Index (CCI) (Charlson, Pompei, Ales, & MacKenzie, 1987), which relies on twenty-two comorbidity groups. One of the recent translation of the CCI is the National Health Service (NHS) England version of the CCI (NHS-CCI), that is continuously being updated (Aylin, Bottle, Jen, Middleton, & Intelligence, 2010; Bottle, Jarman, & Aylin, 2011; HSCIC, 2014; 2015; 2016).

The second stream of models uses a diagnosis classification approach based on simi larities, type, likelihood or duration of care. However, they are usually very complex and specialised to highly particular settings and populations. Also, these models use a period of care records in past, but temporal patterns are greatly ignored. One prominent method is the Elixhauser Comorbidity Index (ECI) (Elixhauser, Steiner, Harris, & Coffey, 1998; AHRQ, 2016), which relies on thirty comorbidity groups and 1-year lookback period. Unlike the CCI, the ECI is using Diagnosis-Related Groups (DRG), which was first developed by Fetter et al. (Fetter, Shin, Freeman, Averill, & Thompson, 1980) and is based on ICD (International Statistical Classification of Diseases) diagnoses, procedures, age, sex, discharge status, complications and comorbidities. A recent adaptation of the ECI is the AHRQ-ECI, which is actively being maintained by the US Public Health Service (AHRQ, 2016). Another well-established method is the John Hopkin's (Weiner & Abrams, 2011) Adjusted Clinical Groups (ACGs), which is a commercial tool. The model uses a minimum of 6-month and maximum of 1-year prior care records, and it encapsulates 32 diagnoses groups, known as Aggregated Diagnosis Groups (ADGs), and their aggregations called Expanded Diagnosis Clusters (EDCs). Moreover, these indices are initially developed to adjust for particular risks, like mortality risk and care utilisation, but they are commonly used in a variety of risk adjustment problems in critical care health services research.

In the machine learning pipeline that was developed in the prior stage of our research (Mesgarpour, Chaussalet, & Chahed, 2017), comorbidity index was an extremely significant factor and has a high potential for further improvement. Presently, comorbidity risk indices have four major weakness areas: robustness, temporal adjustment, population stratification, and the inclusion of associated factors to comorbidities and complications.

In this research, we are going to improve on these four major areas. Firstly, to make the risk score relevant to different environments, an approach must be used to model complex correlation between variables and states. Secondly, to better distinguish the short- and long-term conditions (i.e. prior

43

admission, length-of-stay, and delta-time between admissions), the temporal dimension may be included in form of life-table or a polynomial weight function. Thirdly, population stratification is a major factor in the prevalence of medical conditions, and therefore must be adjusted. Fourthly, major correlated factors to diagnoses may be included directly or indirectly (latent) to improve the risk estimates, including secondary diagnoses, operations, procedures and complications.

The first main outcome of this research was the development of the Temporal Comorbidity Adjusted Risk of Emergency Readmission (T-CARER), to address the four mentioned issues. The secondary outcome was to release a generic, open source and easy-to-use environment to model the comorbidity risk. It consists of a user-friendly IPython Notebook, which calls procedures in MySQL and Python, in addition to third-party libraries. The T-CARER Toolkit and documentation are available online (Apache 2016). Furthermore, comorbidity risk models are constrained by the population and sam ple characteristics, data quality (e.g. missing diagnoses or delayed death registration) and modelling approach. There is a wide range of literature that focuses on modification and benchmarking comorbidity indices, using different datasets, cohorts, complexity, length-of-stay and claims. The prediction targets vary, it includes in-hospital and 1-year mortality, and in some cases 7-day and 30-day readmission (Austin, Stanbrook, Anderson, Newman, & Gershon, 2012; Holman, Preen, Baynham, Finn, & Semmens, 2005; Gagne, Glynn, Avorn, Levin, & Schneeweiss, 2011; Mehta, Dimou, Adhikari, Tamirisa, Sieloff, Williams, Kuo, & Riall, 2016; Sharabiani, Aylin, & Bottle, 2012; Januel Luthi, Quan, Borst, Taff'e, Ghali & Burnand, 2011). Moreover, there have been many attempts at scoring sur gical outcome and complications that are affected by comorbidity (Mehta, Dimou, Adhikari, Tamirisa, Sieloff, Williams, Kuo, & Riall, 2016; Armitage & Meulen, 2010; DFI, 2013). However, they are mainly based on non-administrative clinical variables or are specialized to very specific outcomes and populations.

2. Methodology

2.1 Data

In this study, a bespoke extract of the HES inpatient data was used, which contains records from April 1995 to April 2010. Two main samples were randomly selected from this database, which includes 20% of total unique patients from 1999-2004 and 2004-2009 periods. Then, each main sample was divided into two equal half, to be used for training and testing.

Each time-frame was divided into one year of trigger-event, one year of prediction period, and three years of prior-history. The population includes all alive patients greater than one-year-old that have an admission within the trigger year. The prediction target variables are 30- and 365-day hospital emergency admission to the inpatient.

2.2 Features

After the data extraction step, several stages of data pre-processing and feature selection were carried out using the framework introduced by Mesgarpour et al. (2016). Firstly, a set of data pre-processing

44

steps are applied, then the feature selection steps are carried out. Also, before carrying out the feature selection steps, features are aggregated and split into temporal events, to capture the events through time.

2.2.1 Pre-Processing

The pre-processing stage implements data selection, removals of invalids and imputations of observations. Also, the feature re-categorisation was applied in this stage, to reduce sparsity and to better capture non-linear relationships.

In re-categorisation step, a clinical grouper, known as the Clinical Classifications Software (CCS), was used to categorise the diagnoses, to better capture comorbidities' patterns and cross-correlations. The CCS categorises the ICD-10 (10th revision of the ICD) diagnoses and operations into a number of categories that are clinically meaningful (HSCIC, 2014; Elixhauser & Steiner, 2006; AHRQ, 2016). Furthermore, operations and procedures were categorised using the major categories of the OPCS-4, but alternative coding categorisation may be used, like ICD-10-PCS. The OPCS-4 is an alphanumeric nomenclature (similar to ICD-10-PCS), and is used by the NHS England and has an implicit categorisation for operations based on clinical categories rather than cost or risks.

2.2.2 Life-Table and Aggregation

Healthcare administrative data are severely unbalanced regarding the amount of longitudinal (panel) data per patient and their distributions over the years. Statistical methods are not equipped to handle these type of unbalances directly. Therefore, the survival analysis's life-table approach was used to keep track of temporal events (Singer & Willett, 2003).

Based on previous studies and the initial statistical analyses, four levels of temporal features were generated: 0-30, 30-90, 90-365 and 365-730 days. These four levels capture part of the temporal aspect of comorbidities, in addition to the delta-time between admissions (gapDays) and the length-of-stay (epidur) features that include temporal metadata. Furthermore, in the modelling stage, we applied several techniques to capture the complex temporal patterns of patients' Comorbidities. The temporal features were summarized in each temporal level based on several aggregation functions, including prevalence, count and average. This stage increased the number of features by more than fifty folds.

2.2.3 Feature Selection

After feature generation, a feature pool was produced based on the developed features. Thereafter, the feature selection step has been carried out. Firstly, the features were filtered out based on their linear cross-correlation, as well as frequency and sparseness (percentage of distinct, and the ratio of the most common value to the second most common).

Thereafter, the continuous features have been transformed using two feature transformations methods: scale-to-mean and Yeo-Johnson (Yeo & Johnson, 2000; Breiman, 2001; Hawkins & Blakeslee, 2007). Both methods can be used to transform the data, to improve normality. Although, feature transformations would not guarantee better convergence or very stable variance for any dataset, they have been applied to avoid inputting skewed features into models. Moreover, a downside of

45

transformations is that they make model interpretation harder, and can negatively impact the relationship between correlated features in the model. Therefore, the highly correlated features were removed after transformations.

2.3 Modelling Approaches

The aim of this research is to model emergency readmission using a minimal number of generic features that can be used for short and long-term predictions with high correlation to the comorbidity risk. In this study, there is no condition on the trigger event admission and a minimal number of features are used. This makes it different from general readmission models, like the ERMER (Mesgarpour, Chaussalet, & Chahed, 2017), that use a wide range of features and may enforce the emergency admission condition for both trigger-event and future-event.

2.3.1 Logistic Regression

The first algorithm is a logistic regression with L1 regularisation (1.0), using liblinear optimisation algorithm (Fan, Chang, Hsieh, Wang, & Lin, 2008) with the maximum of hundred iterations and a warm-up period. The random forest method is an ensemble decision tree, which was first introduced by Breiman et al. (2001). It is based on the CART algorithm (Breiman, Friedman, Stone, & Olshen, 1984) and the bagging ensemble method (Breiman, 1996). However, the Breiman random forest is sensitive to highly correlated features, and the scale or categories of features (Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Tolosi & Lengauer, 2011).

2.3.2 Random Forest

We used a random forest method using the Breiman algorithm (Breiman, 2001), with gradient boosted regression trees, Gini index criterion and 1000 trees in the forest.

3. Deep Neural Network and Results

We implemented a Deep Neural Network (DNN) based on the Wide and Deep Neural Network (WDNN) algorithm, which was introduced by Cheng et al. (2016). DNNs are a class of Artificial Neural Networks (ANN) with multiple hidden layers, which allow modelling more complex non-linear problems (Bengio et al., 2009; Schmidhuber, 2015). DNN act like ANN, but with better ability to model complex non-linear models with a more effective representation of features in each layer. The WDNN is a DNN which combines benefits of memorization and generalization. The WDNN consists of two parts: the wide model and the deep model. The wide part of the model consists of a wide linear model for highly sparse features (random features that are rarely active) and is good at memorizing specific cases. The wide part may also include groups of crossed features. Inside of a group of crossed features, each level of one feature occurs in combination with each level of other features. The GLM (Eq. 1) and the cross-product transformation (Eq. 2) for the wide part are defined as the following:

$$y = \mathrm{w}^T x + b \tag{1}$$

$$\phi_k(x) = \prod_{i=1}^{d} x_i^{c_{ki}} \quad ; \quad c_{ki} \in \{0,1\} \tag{2}$$

, where y is the prediction, x is a vector of features of d features, w is model parameters and b is the bias.

On the other hand, the deep part of the model composed of hidden layers of feed forward neural network with an embedding layer and several hidden layers for any other variable. The deep part can be particularly good in the generalization of cross-correlations. Each hidden layer performs the following operation (Eq. 3).

$$a^{(l+1)} = f(\mathrm{w}^{(l)} a^{(l)} + b^{(l)}) \tag{3}$$

, where W(l), a(l) and b(l) represent weights, actuation's and bias for layer l, respectively. Finally, the WDNN for the logistic regression problem (Y) can be formulated as the following (Eq. 4):

$$p(Y = 1 \mid x) = \sigma(w_{wide}^{T}[x, \phi(x)] + w_{deep}^{T} a^{(l_f)} + b) \tag{4}$$

where $\sigma(.)$ is the sigmoid function, $\Phi(x)$ is the cross-product transformations of x features and w. are the weights. In our study, the WDNN model applies Adadelta optimiser (Duchi, Hazan, & Singer, 2011) for the gradients of the deep part, and implements the Rectified Linear Unit (ReLU) activation function to each layer of the ANN (LeCun, Bengio, & Hinton, 2015).

In overall, the WDNN and the random forest models provide a better fit for the 30- and 365-day emergency readmission problems. For the 365-day, the WDNN produces a marginally better Receiver-Operating Characteristic (ROC) compared to the random forest, and significantly better ROC compared to the logistic regression (Figure 1). Also, the WDNN models have very strong precision (positive predictive value), accuracy, and micro-average f1-score. On the other hand, the random forest models have very high sensitivity (true positive rate) and f1-score. Because the classes are highly unbalanced, the precision-recall curve (Figure 2) was used, to compare the area under the curve, average precision and average recall. The plot demonstrates that the area under the curves are significantly lower for 30-day models compared to 365-day models. Also, sample-2 models have a bigger area under the curve across the models.
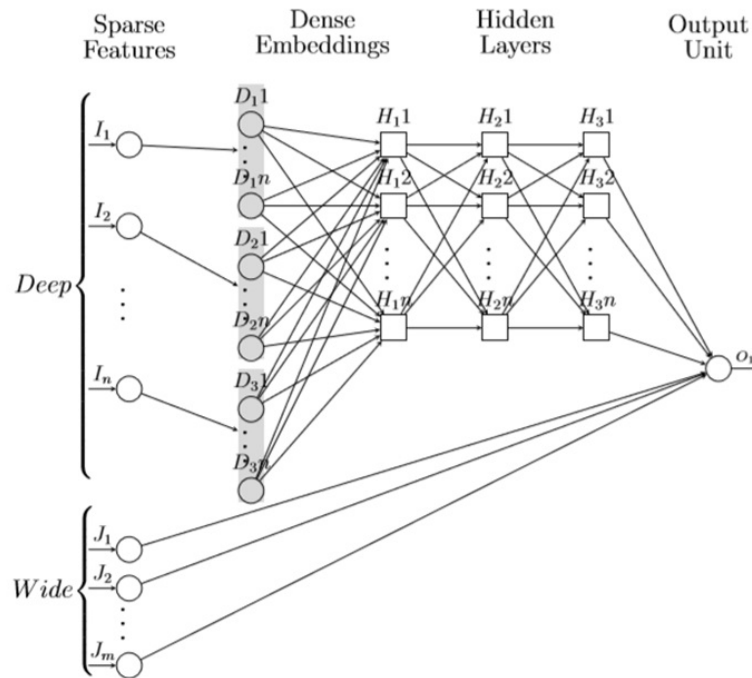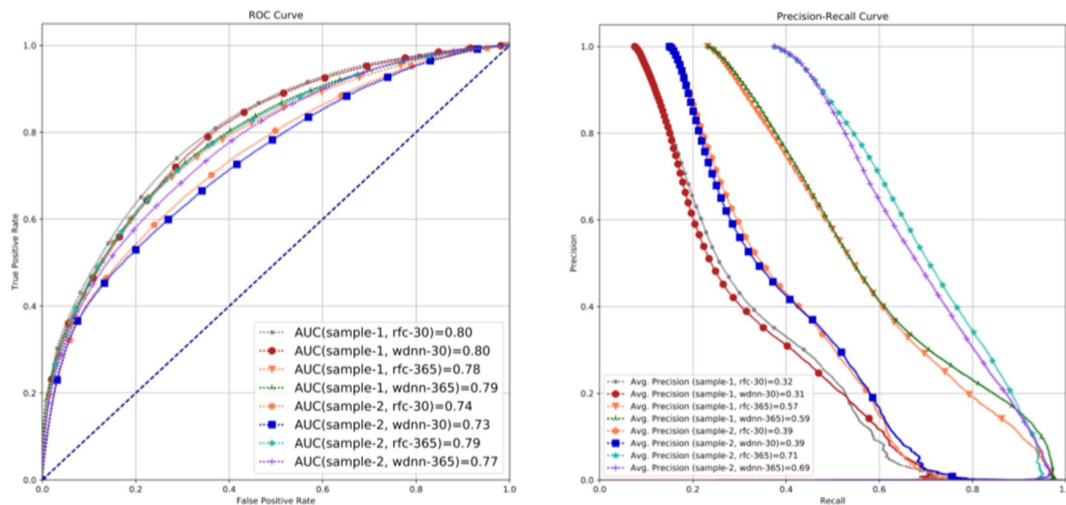
Figure 1. The Abstract Graph of the Wide and Deep Neural Network (WDNN)

(a) Receiver Operating Characteristic (ROC).   (b) Precision-recall curve.

Figure 2. Performance Comparison

4. Discussion and Conclusions

We compared the performance of the T-CARER against commonly used comorbidity index models using different samples and population cohorts across a ten year period. Our analyses of the T-CARER and the NHS-CCI for different diagnoses categories demonstrated that our model performed best in the majority of comorbidity groups, and in overall T-CARER models show better results against previous

48

surveys of CCIs and ECIs.

An advanced research will be sought to identify an approach to score commodities by the inclusion of diverse categories of diagnoses, operations and complexities. The T-CARER performs consistently across tests and validations, and it outperformed against Charl-son and Elixhauser indices which are widely used for prediction of comorbidity risks.

Funding Acknowledgement

References

AHRQ. (2016). *AHRQ—HCUP—clinical classifications software (ccs) for ICD-10-CM* (accessed 1 May 2018). Retrieved from https://www.hcup-us.ahrq.gov

AHRQ. (2016). *AHRQ—HCUP—elixhauser comorbidity software for ICD-10-CM* (accessed 1 May 2018). Retrieved from https://www.hcup-us.ahrq.gov

Apache. (2016). *Apache license, version 2.0* (accessed 1 May 2018). Retrieved from https://www.apache.org

Armitage, J., & Van der Meulen, J. (2010). Identifying co-morbidity in surgical patients using administrative data with the royal college of surgeons charlson score. *British Journal of Surgery*, *97*(5), 772-781. https://doi.org/10.1002/bjs.6930

Austin, P. C., Stanbrook, M. B., Anderson, M. B., Newman, A., & Gershon, A. S. (2012). Comparative ability of comorbidity classification methods for administrative data to predict outcomes in patients with chronic obstructive pulmonary disease. *Annals of epidemiology*, *22*(12), 881-887. https://doi.org/10.1016/j.annepidem.2012.09.011

Aylin, P., Bottle, A., Jen, M. H., Middleton, S., & Intelligence, F. (2010). HSMR mortality indicators (accessed 1 May 2018). Retrieved from https://www1.imperial.ac.uk

Bottle, A., Jarman, B., & Aylin, P. (2011). Strengths and weaknesses of hospital standardized mortality ratios. *BMJ, 342*, c7116. https://doi.org/10.1136/bmj.c7116

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, L., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*, *40*(5), 373-383. https://doi.org/10.1016/0021-9681(87)90171-8

Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, T., … Ispir, M. (2016). Wide &

deep learning for recom mender systems in: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. *ACM*, 7-10.

DFI. (2013). *My hospital guide* (accessed 1 May 2018). Retrieved from http://myhospitalguide.drfosterintelligence.co.uk

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*, 2121-2159.

Elixhauser, A., & Steiner, C. (2006). *Readmissions to us hospitals by diagnosis, 2010: Sta tistical brief #153*.

Elixhauser, A., Steiner, A., Harris, D. R., & Coffey, D. R. (1998). Comorbidity measures for use with administrative data. *Medical care*, *36*(1), 8-27. https://doi.org/10.1097/00005650-199801000-00004

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). Liblinear: A library for large linear classification. *Journal of machine learning research*, *9*, 1871-1874.

Fetter, R. B., Shin, Y., Freeman, J. L., Averill, R. F., & Thompson, J. D. (1980). Case mix definition by diagnosis-related groups. *Medical care*, *18*(2), 53.

Gagne, J. J., Glynn, R. J., Avorn, J., Levin, R., & Schneeweiss, S. (2011). A combined comorbidity score predicted mortality in elderly patients better than existing scores. *Journal of clinical epidemiology*, *64*(7), 749-759. https://doi.org/10.1016/j.jclinepi.2010.10.004

Hawkins, J., & Blakeslee, S. (2007). *On intelligence*.

Holman, C. J., Preen, D. B., Baynham, N. J., Finn, J. C., & Semmens, J. B. (2005). A multipurpose comorbidity scoring system performed better than the charlson index. *Journal of clinical epidemiology*, *58*(10), 1006-1014. https://doi.org/10.1016/j.jclinepi.2005.01.020

HSCIC. (2014). *National clinical coding standards OPCS-4—Accurate data for quality information* (accessed 1 May 2018). Retrieved from http://systems.hscic.gov.uk

HSCIC. (2014). *Summary hospital-level mortality indicator* (accessed 1 May 2018). Retrieved from http://www.hscic.gov.uk

HSCIC. (2015). Summary *hospital-level mortality indicator (SHMI)—Methodology development log* (accessed 1 May 2018). Retrieved from http://www.hscic.gov.uk

HSCIC. (2016). *Summary hospital-level mortality indicator* (accessed 1 May 2018). http://www.hscic.gov.uk

Januel, J.-M., Luthi, J.-C., Quan, H., Borst, F., Taff'e, P., Ghali, W. A., & Burnand, B. (2011). Improved accuracy of co-morbidity coding over time after the introduction of icd-10 administrative data. *BMC health services research*, *11*(1), 194. https://doi.org/10.1186/1472-6963-11-194

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436-444. https://doi.org/10.1038/nature14539

Mehta, H. B., Dimou, F., Adhikari, D., Tamirisa, N. P., Sieloff, E., Williams, T. P., … Riall, T. S. (2016). Comparison of comorbidity scores in predicting sur gical outcomes. *Medical care*, *54*(2), 180-187.

50

https://doi.org/10.1097/MLR.0000000000000465

Mesgarpour, M., Chaussalet, T., & Chahed, S. (2016). *Risk modelling framework for emer gency hospital readmission, using hospital episode statistics inpatient data, in: 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 219-224). https://doi.org/10.1109/CBMS.2016.21

Mesgarpour, M., Chaussalet, T., & Chahed, S. (2017). Ensemble of risk models for emergency readmissions (ermer). *International Journal of Medical Informatics*, *103*, 65-77. https://doi.org/10.1016/j.ijmedinf.2017.04.010

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

Scikit Learn. (2016). *Scikit-learn—machine learning in python* (accessed 1 May 2018). Retrieved from http://scikit-learn.org

Sharabiani, M. T., Aylin, P., & Bottle, P. (2012). Systematic review of comorbidity indices for administrative data. *Medical care*, *50*(12), 1109-1118. https://doi.org/10.1097/MLR.0b013e31825f64d0

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, *8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Tolo¸si, L., & Lengauer, T. (2011). Classification with correlated features: Unreliability of feature ranking and solutions. *Bioinformatics*, *27*(14), 1986-1994. https://doi.org/10.1093/bioinformatics/btr300

Weiner, J., & Abrams, C. (2011). *The johns hopkins ACG system—technical reference guide—version 10* (accessed 1 May 2018). Retrieved from http://acg.jhsph.org

Y. Bengio et al. (2009). Learning deep architectures for ai. *Foundations and trends R in Machine Learning*, *2*(1), 1-127.

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 954-959. https://doi.org/10.1093/biomet/87.4.954