

The Use of Predictive Models in Dynamic Treatment Planning

Saemundur O. Haraldsson^{*†}, Ragnheidur D. Brynjolfsdottir^{*},
John R. Woodward[‡], Kristin Siggeirsdottir^{*†} and Vilmundur Gudnason^{*†}

^{*}Janus Rehabilitation, Reykjavik, Iceland

[†]The Icelandic Heart Association, Kopavogur, Iceland

[‡]Computing Science and Mathematics, University of Stirling, UK

Abstract—With the expanding load on healthcare and consequent strain on budget, the demand for tools to increase efficiency in treatments is rising. The use of prediction models throughout the treatment to identify risk factors might be a solution. In this paper we present a novel implementation of a prediction tool and the first use of a dynamic predictor in vocational rehabilitation practice. The tool is periodically updated and improved with Genetic Improvement of software. The predictor has been in use for 10 months and is evaluated on predictions made during that time by comparing them with actual treatment outcome. The results show that the predictions have been consistently accurate throughout the patients' treatment. After approximately 3 week learning phase, the predictor classified patients with 100% accuracy and precision on previously unseen data. The predictor is currently being successfully used in a complex live system where specialists have used it to make informed decisions.

Keywords-Prediction Models, Healthcare, Dynamic Planing, Machine Learning, Vocational Rehabilitation, Genetic Improvement of Software

I. INTRODUCTION

Computational Intelligence (CI) for eHealth [1] includes Machine Learning for pattern recognition in patient data [2], [3], Image Processing algorithms for diagnosis [4] and monitoring [5], and epidemiology analysis [6], [7]. The use of predictive models has been of particular interest to the healthcare industry [8]. Specifically since sufficient quantity of documentation has been stored digitally to form such quantities to be considered “Big Data” [9], [10]. Traditionally the use has been limited to predicting outcomes before treatment begins, to help with treatment selection or determine the risks versus benefits [11]–[13]. Prior to this work, predictive models have not been used to guide treatment when it is already in progress. Progress in CI and automatic algorithm design [14] has created the potential for using predictive models continually, not only at the beginning of a treatment or to select one, but to make dynamic decisions throughout. The use of predictive models in healthcare serves mainly two goals:

- Increasing the likelihood of successful treatment
- Reducing overall cost of treatments

As individual goals, the former is arguably more important since it affects health and quality of life. The latter has multifaceted effects on society. However, they are closely

intertwined since a successful treatment reduces the risk of relapse. This results in lower future financial burden on healthcare and means being able to effectively prioritize treatments without the loss of service [10]. Being able to adjust treatment that has already begun would contribute to achieving both goals. It would help with guidance towards a successful outcome by providing objective insight into the patients needs and situation, without the risk of “Diagnostic overshadowing” [15]. This is specifically important when the patients suffer from multiple difficulties, both mental, and physical.

This paper presents a novel implementation of a CI predictive tool. The tool dynamically maintains a predictive model and specialises it, in situ, to a single facility with the most up to date information. The implementation is in use by Janus Rehabilitation (JR), Reykjavik. It has been in constant testing in a busy and complicated treatment facility since June 2016. The predictor is currently an add-on feature on Janus Manager (JM), a bespoke software for a vocational rehabilitation centre, developed and maintained by JR [16], [17].

The remainder of this paper is structured as follows: Section II gives a brief overview of related work, Section III describes the implementation of the prediction method, Section IV details how the method has been used in JR and the data it is used on, and finally, Sections V and VI discuss the evaluation of the use case in a practical setting.

II. BACKGROUND AND RELATED WORK

The term eHealth covers a vast literature [1] where Google Scholar search returns over 70K results. It has various sub-fields of healthcare and health related research which all combine technology and information with the goal of assisting or treating people. The focus of this paper is the use of CI in healthcare as increasingly more hospitals and institutions convert from paper administration systems to digitally stored records. Searching through electronic medical records is less time consuming than doing it manually [9], [10]. In addition, the search does not only involve looking up specific details of a single patient, but searching for general patterns [18], [19]. Although humans can identify patterns in data, the vast amount of heterogeneous and often unstructured data involved in medical records [20] will make

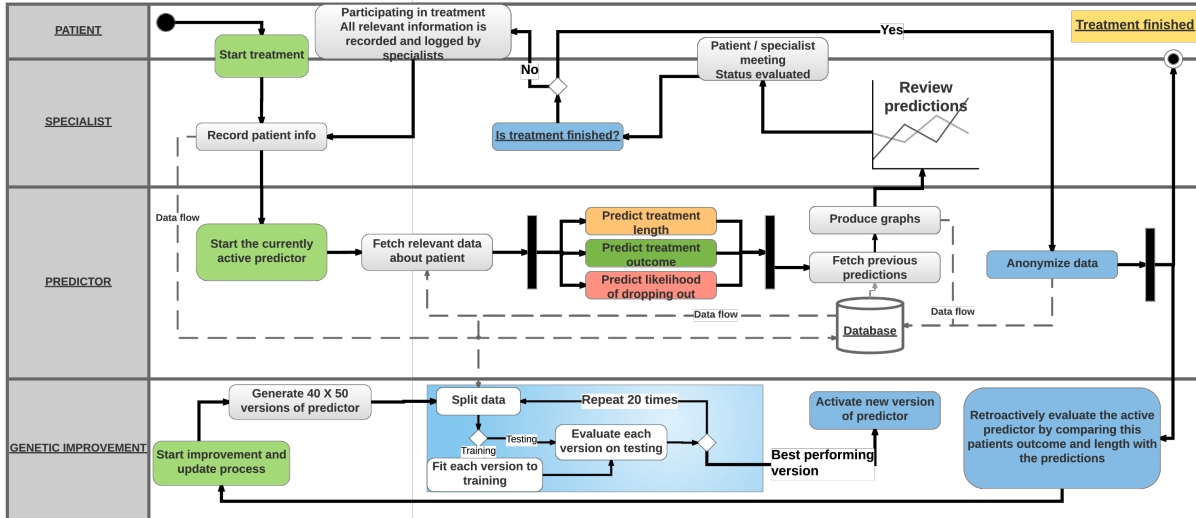


Figure 1. A flow chart showing the prediction and update process while a single patient receives treatment. The patient attends their treatment schedule and provides data. The specialist records the information, reviews predictions, and plans the treatment jointly with the patient. The predictor processes the data, makes predictions and visualises them. Lastly, the Genetic Improvement updates the predictor when the patient finishes.

this more difficult. For that we need predictive models and Machine Learning algorithms [21] to identify these patterns and help us draw inferences. Predictive models have been prevalent in healthcare research for some time and particularly after advances in genome mapping [8], [22], [23]. Examples of successful use of models include predicting outcomes of vocational rehabilitation in patients with brain tumours [11], planning home care rehabilitation [12], and predicting depression treatment outcome [13].

Rehabilitation plays a large part in healthcare and is often a complicated process. Many factors affect both the outcome and its length [24] but current advances in CI could be applied to related problems. Furthermore, there are no examples of predictive models that dynamically adapt themselves during the rehabilitation process. This paper seeks to bridge that gap by discussing a successful implementation of such a model in practice.

The underlying predictive models in our software for classification are based on Random Forest classifiers [25], an ensemble method of tree predictors which has also been used to diagnose chronic kidney disease [4]. Other models include: Bayesian Interpolation [26], Support Vector Machines regression [27], and Neural Networks [28].

The model is periodically updated with new information and improved with Genetic Improvement (GI) of software [29] procedure. GI is an emerging field from Search Based Software Engineering [30] which uses computational search to improve existing software. Such as fixing bugs [16], [17], [31], [32], and reducing execution time [33]. Typically, GI uses Genetic Programming [34] as the search method but other search methods can be used.

III. JANUS REHABILITATION AND THE PREDICTOR

JR was established in 2000 and is one of the largest vocational rehabilitation centres in Iceland [24], [35]. It employs around 40 specialists that work in multiple interdisciplinary teams. Each treatment plan is individually tailored by the specialists in cooperation with the patient and periodically reviewed and updated. There are three main reasons to end treatment: *a)* Patient has begun work or education, *b)* JR's has exhausted its options for treatment, or *c)* Patient decides to end the treatment prematurely with or without notification. JR has no control over *c* other than trying to identify warning signs and intervene whenever possible. For *a* and *b*, each team and their patient share the responsibility for deciding when treatment has ended and planning follow-up measures. JR has a large database of earlier patients that has gradually been building up. In less than a year nearly 400 new instances from 73 patients have been added to the database which already contained over 4300 instances at the start. Each event in a patient's prediction history counts as one instance. Each instance currently has 180 features of 4 data types as listed in Table I. The features describe each patient's circumstances: physical, psychological, personal, and sociological. The data processing is generic enough to allow JR to add features whenever they decide to collect new information about patients.

Since June 2016 JR has used the predictor and successfully confirmed it as a viable tool. It differs from the traditionally off-line predictors by updating its rules on-line whenever new data is recorded. JR has developed a predictive model which the rehabilitation specialists use to inform decisions at every stage of the rehabilitation process.

Figure 1 shows a flow chart of a consultation between

a specialist and patient while the predictor operates and is updated in the background. When the patient enters the treatment, the specialist records the information to the database which initiates the predictor. Three predictions are made and stored, based on the entered data:

- Likelihood of successful rehabilitation
- Drop out probability
- Treatment length, in months

The first two are made with classification, while the last is made with regression. The specialist can then review the patient’s status with this new perspective and plan the next steps. Additionally, the predictor lists the ten most influential features for each prediction to help identify risk factors that affect the outcome and length. These risk factors vary between patients, because each has their unique circumstances.

This cycle is repeated, every time new information is recorded and as long as the treatment lasts. When patients finish treatment, all the collected data regarding them is anonymized and added to the database. The current version of the predictor is then evaluated by comparing all its previously stored predictions about them with the actual outcome.

The bottom layer of Figure 1 is the GI procedure which updates the predictor with the new data. Its implementation has been described previously [16], [17], [32], [33]. In short, it evolves a population of edits that represent small changes to a program. An edit can change the source code by either: Deleting, replacing, copying, or swapping code segments. In this work the targeted software is a Python script that pre-processes the data, and selects a prediction algorithm from scikit learn [36] and tunes its parameters. The objective of the improvement process is to minimise the mean squared error of regression models and maximise accuracy of classification models. The GI uses Monte Carlo cross-validation [37], with 20 repetitions, to evaluate fitness. The dataset is randomly divided into training and testing sets of equal sizes. The fitness of each edit list is then the average performance over 20 splits.

When the GI has finished, the best performing variation, out of 2000 tested, replaces the current predictor instance which fits three models, one for each of the three predicted variables. They are then used for all predictions until the next person leaves treatment and the updating process starts again.

IV. EVALUATION OF THE PREDICTOR

JM has been in use since March 2016 and the predictor was added in June 2016. For JR, the most important evaluation of the predictor is how its specialists experience it in practice.

However, we also verify the predictor objectively by evaluating its performance on those patients that have completed their treatment after it was implemented. The procedure involves iterating over 73 versions of the predictor from

Table I
DATA TYPES OF FEATURES IN THE SET, NUMBER AND EXAMPLES.

Data type	Amount	Examples
Float	120	Age, Length of unemployment, Quality of Life measurement, Current treatment duration
Integer	18	Number of children, Number of medical diagnoses
Boolean	37	Bullied, Dyslexic, Been JR patient before
Categorical	5	Education, Income, Gender, Housing, Relationship status

June 2016 until March 2017 and compare each version’s predictions with actual outcome. For the classification problems, we measure accuracy ($\frac{c}{n}$) and precision ($\frac{p}{n}$), where n is the number of predictions, c is the number of correctly labelled predictions, and p is the number of correctly labelled predictions of the positive class. Accuracy is the proportion of correct labels while precision is the proportion of correct positive labels. For the regression problems, predictions (denoted with \hat{Y}), and true labels (Y), we measure *mean squared error* (MSE) (1)

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (1)$$

and *median absolute deviation* (MAD) (2)

$$MAD = med(|Y_i - med(Y)|) \quad \forall i \in [1, 2, \dots, n] \quad (2)$$

The MAD is a robust variation measurement while the MSE is a well know measure of spread. Additionally, to evaluate how the GI is affecting performance of the predictor we compare these values before and after each improvement process.

V. RESULTS FOR THE PREDICTOR

A. The Practical use of the predictor

JR’s specialists have expressed that being able to identify important factors of the patient’s current status is particularly helpful, along with the graph of previous predictions. It has been used as a visual aid by demonstrating an increased likelihood of a positive outcome, and also to encourage the patient when they cannot perceive progression themselves. Some specialists have also used it to expedite appointments with their patient when the predictor shows increased drop out probability. However, to be able to verify if the use of the predictor has decreased the number of drop outs or shortened rehabilitation length we need to collect data over a longer period. There are a number of seasonal variables that might have confounding effects, such as seasonal affective disorder.

B. Results for classification problems

The predictor did well on two classification tasks; if treatment will be successful, and if a patient will drop out. The

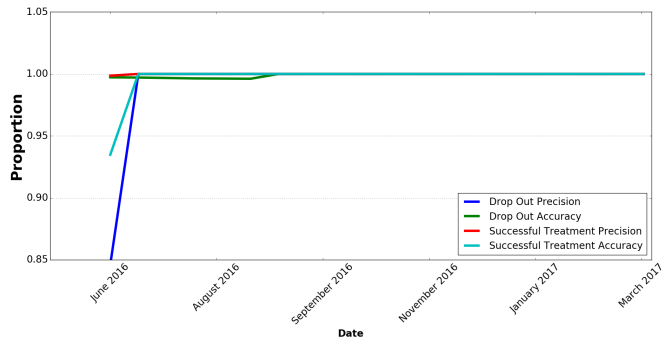


Figure 2. Precision and accuracy of predictions for dropping out and successful treatment over the trial period.

predictions for dropping out had over 99% accuracy from the start, and after August 2016, the accuracy was 100%. Its precision started at 85% but increased to 100% within 2 weeks (see Figure 2). Similarly, predicting a successful treatment was also at 100% accuracy and precision in week two after the release of the first version of the predictor.

C. Results for regression problems

The regression models were able to predict treatment length within three months from the actual duration. A three month difference is an acceptable estimation because in practice this is a lead-in time. Two predictor versions out of total 73 had an error of up to five months. Those versions were updated within a week of being activated and had no measurable effect on therapy length of prevailing patients. Figures 3 and 4 show the performance, as measured post hoc, of number of different models for every two weeks over the period, June 2016–March 2017. The boxes in the figures are the first and third quartiles, the blue line are the best performing models, and the green line is the performance of the models that were in use each time. Note that the performance of the model in use was always better than the mean and median performing model. Furthermore, the variation in the performance of the models gets increasingly larger when adding more data to the training set. It is possibly linked to the variation of treatment length as seen in Figure 5.

D. Genetic Improvement

The GI was used to improve the selection and tuning of both classification and regression models. The GI could not improve beyond maximum regarding the classification accuracy as seen in Figure 2 and the variation of the accuracy between different versions of the predictor was less than 1×10^{-5} . However, the regression models were quite different as mentioned in Section V-C. In Figure 3 and 4 we can see performance spread of the top performing version of each generation from the GI. The green lines indicate the performance of the best version on the test set after correct

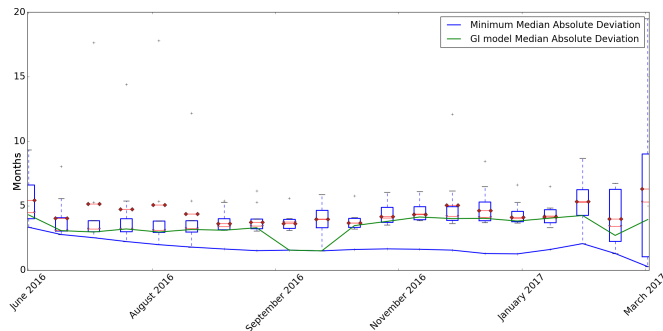


Figure 3. Distribution of post hoc evaluation of MAD for every two weeks of updated models for treatment length. Both mean (diamond) and median (triangle) are marked with each box.

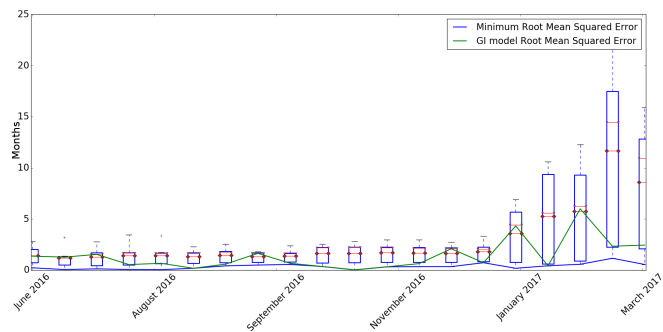


Figure 4. Distribution of post hoc evaluation of MSE for every two weeks of updated models for treatment length. Here, converted to Root Mean Squared Error for scaling on y-axis. Both mean (diamond) and median (triangle) are marked with each box.

results were known. The GI managed to keep predictions mostly within three months from the actual length, causing the overall performance of the predictor to be more than adequate.

VI. CONCLUSION

The performance of the predictor, presented here, was outstanding, with almost 100% accuracy and precision in classification predictions over a 10 month period. Additionally, it has consistently predicted treatment length within satisfactory margin for a vocational rehabilitation.

The predictor was developed to meet the demand for an objective view of each patient’s status while receiving treatment. To the authors’ best knowledge, JR is the first facility to use a predictor, designed for that purpose in practice. JR’s specialists have integrated the use of the predictor into their daily routine to get a clear view of the progress of each patient, at every stage of their treatment. With the increasing number of patients in treatment, the predictor helps by identifying possible risk factors. It assists the specialist to know where and when to intervene, possibly shortening the treatment time. They have used the predictor in various ways, discussing progress or the lack there of

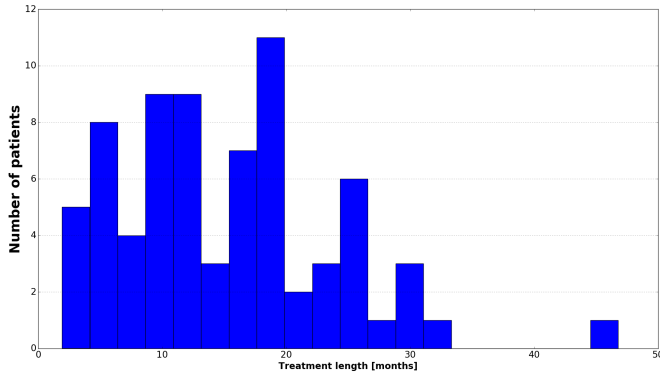


Figure 5. The distribution of treatment length for the 73 patients that finished treatment during the ten month period.

with the patient, to encourage or recognise what might be interfering with the treatment. Some specialist have used the predictor at the start of the rehabilitation to focus efforts on specific areas in the patient’s circumstances.

The predictor’s first 10 months in use have been evaluated by comparing the initial predictions for patients that were receiving treatment with actual outcomes and treatment lengths after they finished. The results for classifying a patient’s treatment as drop out, unsuccessful or successful were more than satisfactory according to experts in rehabilitation, with near 100% accuracy.

The graph in Figure 5 shows us that the treatment time can vary from 2 months to 47 and that is a possible culprit for the large variation in treatment time predictions for JR’s patients. However, the GI was consistently able to find versions of the predictor that had decent performance. Overall the predictions for treatment length were within 2 to 3 months from what actually occurred. The combination of GI and prediction models has proven to be beneficial for the vocational rehabilitation treatment.

The predictor is still in use and continually evolving with the expanding dataset, providing dynamic predictions for the specialists. With current progress in software and hardware development it is well worth exploring automatic adjustments of predictive models. Automatic algorithm design [38] and GI are two of many methodologies to make portable CI tools for healthcare, rather than depending on predefined models that might work well for the general population but not in specific treatments or facilities. In other words, GI can adapt the predictor to the specific data and patients at a given facility. Therefore the predictor was able to perform so well for JR, it was specialised to their database, which contains a narrow population. The predictor is a valuable asset for specialists, patients, and the facility as a whole. The predictor needs to be adapt to each treatment facility and database, and this can be achieved with GI. This predictor can reduce cost and identify possible risk factors, helping specialists to intervene earlier.

ACKNOWLEDGMENT

The work presented in this paper was done in collaboration with Janus Rehabilitation. The authors would like to thank all the specialists for using the predictor and for providing valuable feedback. Two of the authors are also part of the DAASE project which is funded by the EPSRC Grant EP/J017515/1

REFERENCES

- [1] C. Pagliari, et al., “What Is eHealth (4): A Scoping Exercise to Map the Field,” *Journal of Medical Internet Research*, vol. 7, no. 1, p. e9, 3 2005.
- [2] I. D. Falco, “Differential Evolution for automatic rule extraction from medical databases,” *Applied Soft Computing Journal*, vol. 13, no. 2, pp. 1265–1283, 2013.
- [3] S. Schneeweiss, “Learning from Big Health Care Data,” *New England Journal of Medicine*, vol. 370, no. 23, pp. 2161–2163, 2014.
- [4] A. Subasi, E. Alickovic, and J. Kevric, “Diagnosis of Chronic Kidney Disease by Using Random Forest,” in *CMBEBIH 2017: Proceedings of the International Conference on Medical and Biological Engineering 2017*, A. Badnjevic, Ed. Singapore: Springer Singapore, 2017, pp. 589–594.
- [5] D. Mulfari, A. Celesti, M. Fazio, M. Villari, and A. Puliafito, “Using Google Cloud Vision in assistive technology scenarios,” in *2016 IEEE Symposium on Computers and Communication (ISCC)*, vol. 2016-Augus. IEEE, 6 2016, pp. 214–219. [Online]. Available: <http://ieeexplore.ieee.org/document/7543742/>
- [6] P. Saripalli, “Analytic and learning framework for quantifying Value in Value Based Care,” in *Proceedings - IEEE Symposium on Computers and Communications*. IEEE, 2016, pp. 261–266.
- [7] K. Siggeirsdottir, et al., “Epidemiology of fractures in Iceland and secular trends in major osteoporotic fractures 1989–2008,” *Osteoporosis international*, vol. 25, no. 1, pp. 211–2019, 2014.
- [8] J. Danesh, et al., “C-Reactive Protein and Other Circulating Markers of Inflammation in the Prediction of Coronary Heart Disease,” *New England Journal of Medicine*, vol. 350, no. 14, pp. 1387–1397, 4 2004.
- [9] K. Uragaki, et al., “Sequential Pattern Mining on Electronic Medical Records with Handling Time Intervals and the Efficacy of Medicines,” in *Proceedings - IEEE Symposium on Computers and Communications*. IEEE, 2016, pp. 1–6.
- [10] E. Soares, et al., “Modular Health Kiosk for health self-assessment,” in *Proceedings - IEEE Symposium on Computers and Communications*. IEEE, 2016, pp. 278–280.
- [11] S. L. Rusbridge, N. C. Walmsley, S. B. Griffiths, P. A. Wilford, and J. H. Rees, “Predicting outcomes of vocational rehabilitation in patients with brain tumours,” *Psycho-Oncology*, vol. 22, no. 8, pp. 1907–1911, 2013.

- [12] M. Zhu, Z. Zhang, J. P. Hirdes, and P. Stolee, "Using machine learning algorithms to guide rehabilitation planning for home care clients." *BMC medical informatics and decision making*, vol. 7, no. 1, p. 41, 2007.
- [13] A. M. Chekroud, et al., "Cross-trial prediction of treatment outcome in depression: A machine learning approach," *The Lancet Psychiatry*, vol. 3, no. 3, pp. 243–250, 2016.
- [14] E. K. Burke, M. R. Hyde, G. Kendall, and J. Woodward, "Automatic heuristic generation with genetic programming," in *Proceedings of the 9th annual conference on Genetic and evolutionary computation - GECCO '07*. New York, New York, USA: ACM Press, 2007, p. 1559.
- [15] G. Shefer, C. Henderson, L. M. Howard, J. Murray, and G. Thornicroft, "Diagnostic Overshadowing and Other Challenges Involved in the Diagnostic Process of Patients with Mental Illness Who Present in Emergency Departments with Physical Symptoms - A Qualitative Study," *PLoS ONE*, vol. 9, no. 11, 2014.
- [16] S. O. Haraldsson, J. R. Woodward, A. E. I. Brownlee, and D. Cairns, "Exploring Fitness and Edit Distance of Mutated Python Programs," in *Proceedings of the 17th European Conference on Genetic Programming, EuroGP*. Amsterdam, The Netherlands: Springer Berlin Heidelberg, 2017.
- [17] S. O. Haraldsson, J. R. Woodward, A. E. Brownlee, and K. Siggeirsdottir, "Fixing Bugs in Your Sleep: How Genetic Improvement Became an Overnight Success," in *Proceedings of the 2017 Conference Companion on Genetic and Evolutionary Computation Companion*. Berlin, Germany: ACM, 2017.
- [18] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: Using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
- [19] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, p. 3, 2014.
- [20] Y. Wang, L. A. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technological Forecasting and Social Change*, no. March, 2016.
- [21] T. M. Mitchell, "The Discipline of Machine Learning." Tech. Rep., 2006.
- [22] E. A. Marques, et al., "Proximal Femur Volumetric Bone Mineral Density and Mortality: 13 Years of Follow-Up of the AGES-Reykjavik Study," *Journal of Bone and Mineral Research*, pp. n/a–n/a, 2017.
- [23] N. A. Chatterjee, et al., "Genetic Obesity and the Risk of Atrial Fibrillation Clinical Perspective," *Circulation*, vol. 135, no. 8, pp. 741–754, 2 2017.
- [24] K. Siggeirsdottir, et al., "Determinants of outcome of vocational rehabilitation," *Work*, vol. 55, no. 3, pp. 577–583, 2016.
- [25] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [26] D. J. C. MacKay, "Bayesian Interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992. [Online]. Available: <http://www.mitpressjournals.org/doi/abs/10.1162/neco.1992.4.3.415>
- [27] K. P. Bennett and C. Campbell, "Support vector machines: hype or hallelujah?" *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 2, pp. 1–13, 2000. [Online]. Available: <http://portal.acm.org/citation.cfm?doid=380995.380999>
- [28] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 1st ed. Upper Saddle River, New Jersey: Prentice Hall, 1994.
- [29] J. Petke, S. O. Haraldsson, M. Harman, W. B. Langdon, D. R. White, and J. R. Woodward, "Genetic Improvement of Software: a Comprehensive Survey," *IEEE Transactions on Evolutionary Computation*, vol. To Appear, 2017.
- [30] M. Harman and B. F. Jones, "Search-based software engineering," *Information and Software Technology*, vol. 43, no. 14, pp. 833–839, 12 2001. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0950584901001896>
- [31] C. Le Goues, T. Nguyen, S. Forrest, and W. Weimer, "GenProg: A Generic Method for Automatic Software Repair," *IEEE Transactions on Software Engineering*, vol. 38, no. 1, pp. 54–72, 2012. [Online]. Available: <http://www.cs.virginia.edu/~weimer/p/weimer-tse2012-genprog.pdf>
- [32] S. O. Haraldsson, J. R. Woodward, and A. I. E. Brownlee, "The Use of Automatic Test Data Generation for Genetic Improvement in a Live System," in *8th International Workshop on Search-Based Software Testing*. Buenos Aires: ACM, 2017.
- [33] S. O. Haraldsson, J. R. Woodward, A. E. Brownlee, A. V. Smith, and V. Gudnason, "Genetic Improvement of Runtime and its Fitness Landscape in a Bioinformatics Application," in *Proceedings of the 2017 Conference Companion on Genetic and Evolutionary Computation Companion*. Berlin, Germany: ACM, 2017.
- [34] R. Poli, W. B. Langdon, and N. F. McPhee, *A field guide to genetic programming*. (With contributions by J. R. Koza): Published via <http://lulu.com> and freely available at <http://www.gp-field-guide.org.uk>, 2008. [Online]. Available: <http://www.gp-field-guide.org.uk>
- [35] K. Siggeirsdottir, U. Alfredsdottir, G. Einarsdottir, and B. Y. Jonsson, "A new approach in vocational rehabilitation in Iceland: preliminary report." *Work*, vol. 22, no. 1, pp. 3–8, 1 2004.
- [36] F. Pedregosa, et al., "Scikit-learn: Machine Learning in {P}ython," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [37] W. Dubitzky, M. Granzow, and D. Berrar, *Fundamentals of data mining in genomics and proteomics*. Springer, 2007.
- [38] J. Woodward and J. Swan, "The automatic generation of mutation operators for genetic algorithms," in *GECCO'12, 14th annual conference on Genetic and evolutionary computation*, G. L. Pappa, J. Woodward, M. R. Hyde, and J. Swan, Eds., Philadelphia, Pennsylvania, USA, 2012, pp. 67–74. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2330796>