

Using Bootstrap Aggregated Neural Networks for Peripheral Nerve Injury Treatment

Munish B. Shah¹, Wei Chang¹, Kathleen McGuire¹, William Koch², Yan Meng², and Xiaojun Yu¹

¹Chemistry, Chemical Biology, and Biomedical Engineering, Stevens Institute of Technology, Hoboken, NJ 07010, USA

²Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07010, USA

Abstract - *Accidents and trauma can cause severe peripheral nerve injuries and may require surgical intervention. While autografts are considered the current gold standard for complete regeneration of damaged nerves: their scarcity, potential loss of function at the donor site, and potential mismatch in axon diameter limits their use in practice and begs the need for optimal nerve guidance conduits (NGCs), which are the current viable alternative. The major challenges in current NGC research is the inability to account for variations in gap lengths, materials, and enhancement factors. Also, there is an inability to estimate the performance of NGCs, without in vitro and in vivo studies, so that it may be optimized to achieve maximum recovery for an injury. We propose a prediction model based on bootstrap aggregated neural networks in this paper that addresses these challenges and can alleviate the conventional burdens involved in the development of an NGC.*

Keywords: peripheral nerve injury; neural networks; nerve guidance conduit; data collection, bootstrapping

1 Introduction

The peripheral nervous system is made up of the nerves and ganglia outside of the central nervous system, which is the brain and spinal cord. The function of the peripheral nervous system is to bridge the central nervous system with the rest of body. Peripheral nerve injury (PNI) affects approximately 200,000 patients in the United States, with greater numbers reported globally [1]. PNI can be severe enough to result in a loss of function to certain parts of the body especially if there is a gap in the nerve. The gold standard to repair nerve injuries is through two methods: direct coaptation of the proximal and distal stump when the nerve gap is ≤ 4 mm and when the nerve gap is > 4 mm, application of an autograft, in which a patient's own nerve from a secondary location on the body is applied to the primary site of injury. Autografts are scarce, can lead to a loss of function at the secondary site, and can have a mismatch in axon diameters; all of which encompass the drawbacks associated with this approach [2]. While allografts can be applied, they too have the same limitations as autografts. A viable alternative is nerve guidance conduits (NGCs), which are tubular guidance channels that are coapted to the proximal and distal ends of an

injured nerve to guide the regenerating axon from the proximal stump to the distal stump. NGCs can enhance nerve regeneration if coupled with growth factors, cells, and proteins as well as aligned fibers [3]. Conventional methods of assessing the best method to optimizing NGCs is through in vitro and in vivo studies which expend time, capital, and resources. This is a continuous process that is repeated after the critical parameter(s) is altered until the expected results are achieved. Having an ability to bypass this process or at least alleviate the burdens faced by researchers so as to progress in optimizing NGCs and more effectively treat PNI.

A second challenge within the research and development of NGCs is the inability to account for the variety of materials used for fabrication, variety of factors that can be coupled with an NGC to promote nerve regeneration, variations in gap length, and the variety of experimental practices used by researchers to conduct in vivo studies to name a few which limit the comparison of NGCs. While these two challenges are currently hindering the pace of NGC optimization and its potential to match the ability of autografts, we address these challenges in this paper through the presentation of a prediction model based on a normalization standard noted as L/Lc.

2 Data Preparation

This section is focused on the normalization standard, the criteria of which we used to build our dataset. Additionally we detail our process of collecting our data and the format of our dataset.

2.1 Normalization Standard

A normalization standard allows one to compare products or processes that may vary in at least one aspect by adjusting a critical common parameter(s) associated with them. A normalization standard for NGCs related to PNI has been proposed by Dr. Ioannis Yannas's group of the Massachusetts Institute of Technology. This standard is denoted as L/Lc and is a calculated ratio of the gap length divided by the grafts critical axon elongation. It was developed to compare the regenerative activity of NGCs while accounting for aforementioned factors. L/Lc is based on the value of %N which reflects the successful rate of regeneration within an experimental group. For example if within a sample size of 5, 4 samples showed successful regeneration then the %N value is 80%. Having a parameter such as %N is critical in assessing how effective an NGC is and exactly how

repeatable and reliable successful regeneration is. While researchers generally report results and compare the performance of an NGC to the results of a tubular conduit and an autograft, it is also vital to report the %N to give an insight of how repeatable results are.

A plot of %N with respect to gap length was constructed using data collected from studies performed on rat sciatic nerves using silicone based NGCs. This yielded an S shaped curve which is used to approximate the value of L_c , the critical axon elongation, which is the number of samples within a group for which an axon reached at least 50% of the distance from the proximal stump to the distal stump. The L_c is simple a linear shift on the x-axis with respect to the value of %N and the gap length. Once the L_c is approximated for an NGC, it is compared against a standard NGC, defined as a tubular NGC, which is known to yield the poorest performance in nerve regeneration, using equation shown below [4,5].

$$\Delta L = \frac{L^{exp}}{L_c} - \frac{L^{exp}}{L_c} \quad (1)$$

The equation (1) was used to calculate the ΔL value, to provide a quantified measure of regenerative activity of an NGC relative to a standard NGC. This standard is also meant to be used to compare NGC effectiveness in nerve regeneration across different species, but for the purposes of the proof of concept it was validated on only mice and rat data. Since this concept has not yet been extended beyond applicability to rats and mice we focused on using it strictly for dealing with parameters and data from and for rats. Having a normalization standard is critical in the development of NGCs as it can allow researchers to gauge whether their NGC(s) is better than the standard NGCs which in turn will help them assess whether it is a move forward in the field of PNI. Additionally by being able to compare across gap lengths it can allow researchers to assess the capability of their NGC or NGCs developed by others to characterize its limitations as well as understand its advantages for nerve regeneration.

2.2 Data Collection

In the development of any prediction model, a vast and diverse dataset is key in developing an effective prediction model. A large number of input parameters requires a more diverse and larger dataset so its accuracy is high and can be applied for its intended purpose. In our particular case we had a large number of input parameters and a relatively small dataset.

We focused on collecting data related to NGCs studied on rat sciatic nerves since that is the preferred anatomical site of choice for PNI as well as the fact we used the L/L_c normalization standard to build our dataset. Due to the absence of a central forum dedicated to the exchanging of ideas and research for NGCs related to PNI, we relied on collecting data from scientific publications through an array of

databases and journals. We searched for publications on Google Scholar, EBSCOHOST, Science Direct, and the Wiley Online Library. The criteria for including a publication in our dataset was that in it: researchers detailed their protocol for animal studies, use the rat sciatic nerve as their anatomical site, provided values of %N, reported their recovery time(s), detailed their NGC designs, and used a standard NGC (tubular NGCs) to compare their experimental NGC performance. This criteria was designed to conform to the L/L_c normalization standard and to ensure the procedures used and the form of data reporting within the papers was uniform.

Using this criteria we were able to isolate a total of 28 scientific publications and within these publications each type of NGC that served as an experimental group was designated as a 'case' within our dataset. A total of 138 cases were recorded in Microsoft Excel and each parameter involved in the development of the NGC was noted. The output parameters were the value of L_c , ΔL , %N, and L/L_c . The L_c value for each of the negative controls in from each publication were designated as 0, since the goal was to compare the performance of each experimental NGC relative to the negative control. L_c was calculated for each experimental NGC using the S-shaped curve based on the gap length and %N, which were provided in each of the 28 publications. The ΔL for each experimental case was calculated using Eq 1. In our analysis we observed a few cases in which the performance of the experimental NGCs was equal to the negative control, however there were several cases in which the ΔL was as high as 6.8. This provides an assessment that progress is being made within the field to improve NGC design, and quantified measures such as ΔL can allow a better understanding of exactly how well the progress are being made and whether design can be improved or not. Each aspect related to the design of an NGC served as an input to the prediction model and was categorized as shown in Table I. A total of 46 parameters served as inputs. The parameters we noted are not the complete set of parameters used by researchers currently and the list can certainly be expanded in the future. These parameters were selected based on the fact that the publications that met our criteria for inclusion in the building of our dataset used these parameters to achieve successful nerve regeneration with their NGCs. We categorized the parameters as shown in Table I based on what their purpose was in NGC development and how they were applied. Categorizing the parameters clarifies the aspects associated with NGC development and allows a better understanding of which are essential to keep and which are optional. Such categorization can alleviate the challenge with optimization by allowing researchers to focus on a single category.

The output of the prediction model was selected as ΔL , since it reflects the effectiveness of a conduit relative to a standard conduit. The justification for using ΔL as the output for the prediction model stems from the fact it is the only output of the ones noted in the dataset that provides a quantified measure of the effectiveness of an NGC. While L_c , L/L_c , and %N provide a measure of NGC performance, these

parameters are specific to a single NGC case and cannot be used to compare the performance of different NGCs with each other. These parameters would be applicable in the assessment of an individual NGC, however in our approach we aimed to provide researchers with feedback about the status of the design of their NGCs through the ΔL value. ΔL does not have a known maximum limit or minimum limit, but can provide a researcher the opportunity to improve their NGC design and add/remove parameters that may improve/hinder its performance without the need to undertake the burden of doing in vitro and in vivo studies to come to this conclusion.

Table I. Inputs categorized for Prediction Model

Category	Model Input Parameters
Materials Processing	Phase Separation, Hydrogels, Electrospinning
Structure	Fibers, Gel, Permeable, Impermeable, Microsphere, Porous, Lumen
Materials	Collagen, Ethyl Vinyl Acetate, Polycaprolactone, Poly Lactide, Poly Glycolic Acid, Poly Lactide Co Glycotide, Chitosan, Poly Phosphazene, Poly Pyrrole, Poly Sulfone, Silicone
Form	Hydrogel, Liquid, Gel, Matrix, Fiber-Aligned, Fiber-Random, Microsphere, Solid
Growth Factors	NGF, BDNF, CNTF, GDNF, FGF, Denatured FGF, IGF, Laminin, Fibronectin, Schwann Cells, Bone Marrow Stromal Cells, Neural Crest Stem Cells
Growth Factor Arrangements	Gradients or Anisotropic, Isotropic

3 Prediction Model

This section details the predictors we considered in the development of our prediction model as well as the design of our final product. We also note the challenges we faced in the application of our model and how we addressed these challenges.

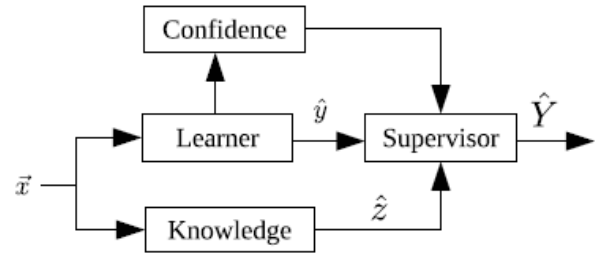
3.1 Design of Prediction Model

In our particular dataset for NGCs we constructed based on the L/Lc normalization standard there was high correlation within the data and a very limited amount of data was available. However in compensation for the lack of data we had some domain knowledge about peripheral nerve injury with which the accuracy of the predictions could be improved if used with the appropriate predictor. The predictors we considered using to incorporate prior domain are virtual examples, hints, support vector machines, and artificial neural networks. Virtual examples are new training examples developed during a learning process using prior domain knowledge [6]. There are two methods to implement virtual examples: transforming the original data or synthesizing new examples [6-8]. The former has the challenge of assessing which method of transformation is appropriate and the potential that the virtual examples may correlate to the original data [9]. Also, it has been noted that virtual examples when tested with real data can result in poor performance [8]. Hints are simply additional knowledge about the target function which is useful during learning [10]. Hints can be

applied as virtual examples. While hints have potential in making valid predictions we could not apply them for our purpose due to the drawbacks associated with virtual examples. Support vector machines is a machine learning algorithm that has shown promise in an array of biological areas, however it is computationally expensive and timely which is why it is inapplicable for our purpose [11]. While there is no shortage of predictors, our choice was the artificial neural network (ANN) as it commonly used and some 3rd party software even implement ANNs. While ANN is more complex than some predictors, it allows us the flexibility we require as well as that we have more experience using this predictor than others.

In our use of ANNs we used the technique of bootstrapping, which is a re-sampling technique used in machine learning beneficial when the amount of data available is limited as well as it can improve accuracy [12, 13]. Below is an illustration of the design of the prediction model.

Figure 1. Overview of our Prediction Model



The basic principle is that the PI indicates the level of confidence the learner has on the prediction. The domain knowledge is useful to compensate for the lack of knowledge of the learner. The learner output gets adjusted depending on how confident it is on estimating the output for the current input values. The learner component was defined as the bagged neural networks and was trained using the aforementioned dataset. In our case we used bootstrap aggregated neural networks which are developed when multiple neural networks are trained on bootstrap samples. Aggregated neural networks improve accuracy, improve generalization, and are more robust compared to just a single neural network as we assessed in [14]. Aggregated neural networks have a reduced chance of error compared to a single neural network. Based on the experiments we conducted in [14], we determined the optimal number of neural networks to be 100 when trained with particle swarm optimization, which provides greater accuracy than backpropagation for training neural networks. The confidence component computes the learner's PI for the current inputs. The knowledge component provides a separate estimation of the target function based on acquired domain knowledge. The supervisor component assesses the learner's confidence and determines if the learner's output should be influenced by domain knowledge. The supervisor component was responsible for fusing the prediction made by the knowledge base, the dataset, and the learner component, based on the confidence value, calculated through the prediction intervals (PI). PI determine how far an

estimated value is from the target. The PIs were calculated using the bootstrap aggregated neural networks, which can improve accuracy in predictions for even small datasets [15].

The idea was to keep the architecture simple and straightforward since this prediction model is intended to use as a guideline for future developments for models applied for NGC research. Using PIs allowed us to ensure we were able to provide as accurate of an output as possible given the size of the dataset and the number of inputs we had. The architecture of our prediction model is further detailed and described in [14].

3.2 Challenge of Applying the Prediction Model

A challenge we faced in development was to limit the number of combinations that could be created of the input parameters. This is based on the fact that if the prediction model devises potential combinations on its own then it can have lower accuracy, however if it is given a limited number of combinations of the inputs then the accuracy can be optimal especially within a large number of inputs and a small size dataset. Parameters within categories and parameters across categories can be coupled together to develop an NGC which increases the possible types of NGCs that can be developed. For example, researchers have coupled synthetic with natural materials to develop NGCs. Additionally such materials can be coupled with a growth factor, cells, proteins, or all three factors together. While our prediction model is capable of devising possible combinations that can be made of parameters for potential NGC designs, we limited this capability by providing it the potential combinations of the inputs. We devised combinations through a thorough literature review as well as our own knowledge and experience within the field. This approach allowed us to achieve optimal accuracy in our prediction model as well as account for all possible combinations of input parameters.

4 Experimental Results

As stated above the purpose of the prediction model is to estimate the performance of an NGC developed using a defined set of the inputs. We assessed the accuracy of the prediction model by assessing its training and prediction error, which were calculated using data for which the output was known. This type of data was designated as the control. The control in our case was the dataset we had used to train our prediction model. In order to effectively train the prediction model and assess its accuracy in making predictions, we applied a concept known as cross validation. In cross validation the dataset is split into several subsets of equal size which are used to train and test the accuracy in making predictions of a prediction model [16]. For this particular purpose we chose to apply the seven fold cross validation approach, and as such our dataset was split into seven parts of which 6 parts had 20 cases each and the 7th

had 18 cases, since the number of cases could not be evenly divided by 7.

The first six parts were combined to yield a dataset of 120 cases, which were used to train the prediction model. The reason to have a bigger size dataset for training is due to the fact that it is more crucial to train a prediction model better than it is to test its prediction error. A prediction model will be able to make more accurate predictions only if it is trained well. The 120 cases were also used to assess training error of the model. In addition to training the prediction model we quantified the training error and noted it as shown in Table II. We used the remaining 18 cases to assess the prediction error, the results of which are shown in Table III. There were three parts for which the dataset was crucial: training, then assessing training error, and assessing the prediction error.

Table II. Training Accuracy of the Prediction Model

Material	Growth Factors	Permeable	Lumen	Gap Length	Actual ΔL	Predicted ΔL	Difference in ΔL
Collagen	1 M Schwann Cells	Yes	1	18 mm	6.8	5.4	1.4
Collagen	None	Yes	1	22 mm	9.4	7.56	1.84
Ethyl Vinyl Acetate	.004% Fibroblast Growth Factor/BSA	No	1	15 mm	6.8	5.19	1.61
Silicon	30X10 ⁴ /m ² Calcium Ions	No	1	15 mm	6.8	5.055	1.745
Poly(lactide)	1670 Neural Crest Stem Cells	No	1	10 mm	4.533	3.196	1.337

Table III. Prediction Accuracy of the Prediction Model

Material	Growth Factors	Permeable	Lumen	Gap Length	Actual ΔL	Predicted ΔL	Difference in ΔL
Collagen	None	Yes	1	20 mm	10.8	7.343089	3.456911
Silicon	None	Yes	1	16 mm	5.67	3.5464	2.1236
Collagen	None	No	1	10 mm	5.4	4.910012	0.489988
Polysulfone	100 μ g/ml NGF	No	1	10 mm	3.4	2.61726	0.78274
Poly(lactide-co-glycolactide)	None	No	1	10 mm	3.020909	1.730524	1.290385

An aspect to note with Tables II and III is each data recorded is specific to a researcher's expertise and skillset as well as the protocol by which the studies were performed. As with any prediction model our work is also susceptible to error as shown by the results in Table II and Table III, however such an occurrence is expected especially in light of the fact we had a small volume of publications to work with and a high number of input parameters. While there is error present in training the error is precise and the precision in our error ensures that our model is able to make uniform predictions regardless of variation in protocols followed by researchers as well as the fact that basing our prediction model on the L/Lc normalization standard allowed us to normalize our data despite differences in the materials used,

growth factors, cells, and proteins that were coupled, and variations in gap length.

As for the errors in training and prediction, these errors can be reduced with the addition of more cases in our dataset. Our focus in developing the prediction model is emphasize the need for standardization in doing NGC research related to PNI so in future all researchers report their %N values. Having more papers published in this regard will allow us to expand our dataset and conduct future developments on our prediction model so when more cases can be made available through future publications similar to the ones used to develop the dataset for this model the error can ideally be reduced and reliability of our prediction model will increase. Additionally since we categorized the inputs involved in the design and development of an NGC we expect any alteration in any category to impact the value of ΔL to further the understand the relationship the mechanism of nerve regeneration related to an NGC and the parameters that are involved in development. When researchers are able to isolate which category or parameter impact the output the most it will allow them a better understanding of where to make modifications. By having a deeper understanding of the mechanism of nerve regeneration for an NGC researchers can be better prepared to optimize NGCs and have a better grasp on how to deal with critical PNIs.

5 Discussion

To our knowledge our work provides the only prediction model relevant to estimating NGC performance with bootstrap aggregated neural networks to calculate prediction intervals and uses particle swarm optimization for training to assess NGC performance. The work that has been done prior involved the application of a single neural network and ours provides the advantage of random sampling of data, which is a more practical approach to developing a prediction model since all types of data should be viewed in making a prediction rather than focus on a specific set. By focusing our model on being able to analyze all forms of NGC data it is better trained to understand and predict the performance of a novel NGC design. Our aim is to alleviate the conventional burdens associated with NGC development, particularly the necessity to perform in vitro and in vivo studies. By using such a process we ensured our model is able to sample all forms of data that is provided to it in regard to NGCs. This ideally mimics a researchers thought process of reviewing all literature and data available on NGCs to effectively critique their own design using critical thinking and make future developments. We designed our prediction model to mimic the thought process of our prediction model in terms of critical thinking. In our algorithm we design our model to process data similar to how. Since papers tend to report data for an NGC on a specific gap length of study and under conditions specific for experiments, researchers cannot assume the performance of that under different parameters which is why they must account for data provided and apply critical thinking before moving forward in their own designs and experiments.

A valid and applicable prediction model for NGCs has the potential to alleviate the conventional burdens associated with development and assessment of viability as well as the potential to allow researchers to develop the optimal NGCs for specific injuries, such as those specific for a certain gap length of injury. Despite the relatively small dataset size we had to work with we were able to successfully prove our concept of being able to develop a prediction model applicable to NGCs. This prediction model was developed to serve as a guideline for future work.

The L/Lc normalization standard allowed us to effectively develop a prediction model applicable for any gap length as well as standardize for materials and additional factors. We understand that the L/Lc normalization standard is not without error, the critical being the inability to account for recovery time which significantly impacts the outcome of PNI treatment by an NGC. We also realize that L/Lc was proposed solely as a normalization standard however by applying it in terms of building a dataset for use in developing a prediction model allows us the opportunity to contribute a novel approach in assessing the outcome of a potential NGC design rather than rely on the conventional methods.

The vast amount of parameters makes it difficult often for researchers and readers to gauge the complexity of designing an NGC and can further hinder its optimization, as such we aimed to clarify this aspect through the presentation of parameters we applied in our prediction model. Our aim is to allow an organized approach to understanding NGC development as evident by our categorization of the input parameters as well as the potential combinations of these parameters. Finally by being able to modify each parameter or parameters within the categories noted the ΔL can ideally be affected which will allow researchers to understand which category and which parameter(s) has the most impact on nerve regeneration.

By saving cost, time, resources and developing an optimal NGC researchers can expedite the steps from design to commercialization and bring more effective NGCs to market. We designed our model to use random sampling of the data so as to provide an unbiased analysis and output of a novel design. Most importantly researchers will be able to potentially devise effective NGCs that can match the performance of autografts if not outperform them to provide maximum nerve regeneration affected patients. While there are very few, if any, conduits that have come close to providing nerve regeneration comparable to autografts this potential is being attained more each day. Having a valid substitute to autografts will allow clinicians to rely less on autografts, thus eliminating the drawbacks that are associated with them and apply more NGCs to treat even critical PNIs thus effectively dealing with the growing number of PNI cases globally.

6 Acknowledgements

This work was generously supported by NIH-R15 NS074404. The author, Munish Bhupendra Shah, was supported by NSF-0740462.

7 References

- [1] Lehoe S, Zhang XF, Boyd D. "FDA approved guidance conduits and wraps for peripheral nerve injury: A review of materials and efficacy." *Injury*, vol 43, no. 5, pp. 553-572, 2012.
- [2] Yu X, Bellamkonda RV. "Tissue-Engineered Scaffolds Are Effective Alternatives to Autografts for Bridging Peripheral Nerve Gaps." *Tissue Engineering*, vol 9, no. 3, pp. 421-430, 2003.
- [3] Daly W, Yao L, Zeugolis D, Windebank A, Pandit A. "A biomaterials approach to peripheral nerve regeneration: bridging the peripheral nerve gap and enhancing functional recovery." *Journal of The Royal Society*, vol 9, no. 67, pp. 202-221, 2012.
- [4] Yannas I V, Hill B J. "Selection of Biomaterials for peripheral nerve regeneration using data from the nerve chamber model." *Biomaterials*, vol 25, pp. 1593-1600, 2004.
- [5] Zhang M, Yannas I, . "Peripheral Nerve Regeneration." *Adv Biochem Engin/Biotechnol*, vol. 94, pp. 67-89, 2005.
- [6] P. Niyogi, F. Girosi, and T. Poggio, "Incorporating prior information in machine learning by creating virtual examples," *Proceedings of the IEEE*, vol. 86, no. 11, 1998.
- [7] T. Yu, T. Jan, S. Simoff, and J. Debenham, "Incorporating prior domain knowledge into inductive machine learning," pp. 1-42, 2007.
- [8] D. Pomerleau, "Efficient training of artificial neural networks for autonomous navigation," *Neural Computation*, pp. 1-10, 1991.
- [9] T. Yu, T. Jan, S. Simoff, and J. Debenham, "Incorporating prior domain knowledge into inductive machine learning," pp. 1-42, 2007.
- [10] Y. S. Abu-Mostafa, "Hints," *Neural Computation*, vol. 7, no. 4, pp. 639-671, 1995.
- [11] Furey, Terrence S., et al. "Support vector machine classification and validation of cancer tissue samples using microarray expression data." *Bioinformatics*, vol 16, no. 10 PP. 906-914, 2000.
- [12] Efron, Bradley, and Robert J. Tibshirani. *An introduction to the bootstrap*. CRC press, vol. 57, 1994.
- [13] H. J. Ader, G. J. Mellenbergh, and D. J. Hand, *Advising on Research Methods: a consultant companion*. Johannes van Kessel Publ., 2008.
- [14] Koch, William, Yan Meng, Munish Shah, Wei Chang, and Xiaojun Yu. "Predicting nerve guidance conduit performance for peripheral nerve regeneration using bootstrap aggregated neural networks." In *Neural Networks (IJCNN), The 2013 IEEE International Joint Conference on*, pp. 1-7, 2013.
- [15] Breiman,Leo. "Bagging Predictors." *Machine Learning*, vol 24, pp. 123-140, 1996.
- [16] Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." In *IJCAI*, vol. 14, no. 2, pp. 1137-1145, 1995.