Early Detection of Chronic Kidney Disease using Various Machine Learning Techniques

|  |  |  |
| --- | --- | --- |
| Moh.Faturrahman | Ito Wasito | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Faculty of Computer Science, University of Indonesia, Depok Indonesia | Faculty of Computer Science University of Indonesia, Depok. Indonesia | Faculty of Computer Science University of Indonesia, Depok. Indonesia |

***Abstrak***

*Chronic Kidney Disease (CKD) could increase the risk of associated complications such as hipertention, heart disease, anemia, mortality, progression to end-stage renal disease. In addition different stages of CKD are refered to different clinical options. Therefore early detection of CKD is needed not only to support clinicians to make decisions but also to prevent and to manage the risk. Machine Learning has a potential to detect CKD in order to support clinical care of the patients. In this study, Artificial Neural Network Resilient Backpropagation (ANN-RP), Support Vector Machine (SVM) and Extreme Learning Machine (ELM) are utilized to predict the presence of CKD based on patient clinical data. The Experimental result shows that, ANN-RP achieved highest average in accuracy, sensitivity and specificity.*

***Keywords : Machine Learning, Chronic Kidney Disease, Early Detection,***

1. **INTRODUCTION**

Chronic Kidney Disease (CKD) is defined as symtomps of kidney damage for at least three months, followed by the decrease of Glomerular filtration rate (GFR ) < 60 mL/min/1.73m2[1]. The Prevalence of CKD in Indonesia is 0.2% and most of the patients are villagers,non-educated people, farmer, and private sector worker [2]. CKD increases the risk of other complications such as, hipertetion, anemia, mortality, and end-stage of CKD. In clinical practice, measurement of GFR and urinary albumin, are used to detect CKD and its stages. Other support medical tests such as Ultra Sound/CT Scan, serologic testing is not routinely required. Early Detection of CKD are required in order to avoid complications, to reduce the progression of end-stage renal disease, and to slow loss of kidney function [3].

Machine Learning Techniques are widely used in medical area in order to help clinicians making medical decisions, such as medical treatment and drug design. Most machine learning techniques including ANN,SVM and ELM are applied in this field.

Guang Jing et al, used ELM for predicting mortality rate of bladder cancer after radical cystectomy. ELM shows that the average accuracy is 80% which is better than other statistical methods [4]. Vukicevic et al, used neural network for predicting mortality rate of bladder cancer using different patient data. In this research, neural network achieved 92.5% in accuracy.

Many researchers have applied ELM in medical area [5], [6] [7], show that ELM would be one of the alternative solutions which has better result in accuracy, sensitivity and specificity. In this research aims to measure the peformance of ELM in term of predicting CKD. Further design and implementation of ELM will be explained in the next section.

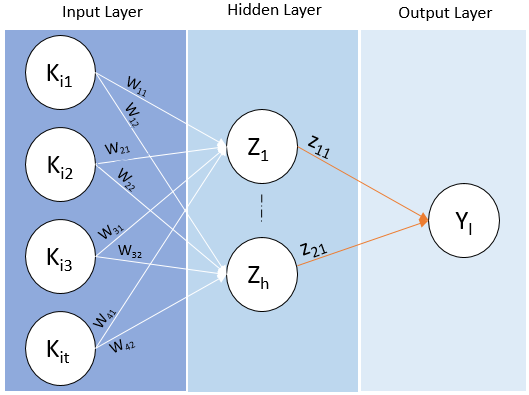
The rest of the paper organized as follows. Section II provides the design and implementation of Machine Learning Techniques including ANN, SVM and ELM. Section III describe the design experiment, results, and analysis, Section IV concludes the peformance of ANN, SVM and ELM.

1. **DESIGN AND IMPLEMENTATION**

**2.1 Extreme Learning Machine**

Extreme Learning Machine (ELM) which is a new modified of Single Feed Forward Neural Netwok (SLFN) by Guang-Bin Huang [8]. Unlike other SLFN algorithms that use *Gradient Based Learning Algorithm* to find weights in each layer, ELM uses *Moore Penrose Pseudo Invers Matrix* to find weights in once iteration. Supposed there are K training samples of data [Pi, Yi], where Ki = [Pi1,...,Pis] and Yi = [Yi1,...,Yim] . The ELM consists of single hidden layer of neural network with H hidden nodes, i represents the data in sample ith.

The result in hidden nodes are placed in a matrix , each value in matrix A is calculated by an activation function , where and are inner product of vector weight .



Gambar 1 : ELM Architecture

The weight between hidden node and output layer will be computed by

|  |  |
| --- | --- |
| Z = A†Y | (1) |

where A† is Moore Penrose Pseudo Inverse Matrix of matrik A, vector Y is class target in sample. The vector weight of Z can be used to predict the target class Y that can be expressed as

|  |  |
| --- | --- |
| Y = AZ | (2) |

In this study, the implementation of ELM can be summarized as follows

1. All weights in each layer are assigned by random number.
2. Calculate the hidden layer output matrix A.
3. Calculate Moore Penrose Pseudo Invers of Matrix A
4. Calculate ouput weigth

The different from traditional gradient-descent algorithm are extremely fast learning process, avoid local minima, can be used to train with non-differentiable activation function, not only to reach smallest minimum training error but also the smallest norm of weights [9]

* 1. **Support Vector Machine**

Support Vector Machine (SVM ) is a common kernel based algorithm in machine learning. SVM can be applied to classification and regression. The basic idea of SVM is that to find optimum separation hyperplane in order to separate data into two classes [10].

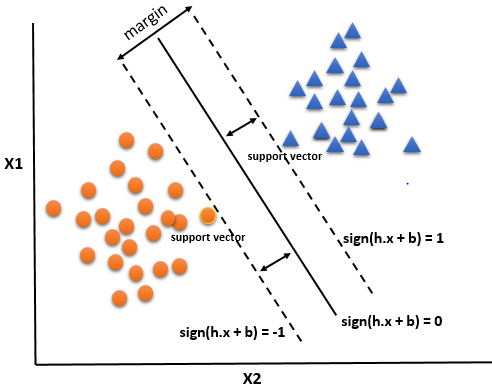
 Given training data **K**{xi,yi},i = 1,2,...,n,, xi Rd yi = {1,-1}. SVM finds hyperplane which has the largest margin that separates two classes. All training data should satisfy for yi = 1 and for yi = -1, the point which is plotted on hyperplane satisfies . Where *x* is a input vector, *h* is weight vector and b is bias. The data that have nearest distance to decision boundaries are called support vectors. In order to define separating hyperplane, the formulation can be written as

Figure : Separation Hyperplane of SVM

|  |
| --- |
|  |
|  |

and the formations above can be combined in simple form as

|  |
| --- |
|  |

Optimum separation hyperplane, can be obtained by maximizing . Because it is constraint optimization problem, it can be solved by the Lagrangian Multiplier

|  |
| --- |
|  |

After the derivative at min = 0, it can be expressed as

|  |
| --- |
| , |

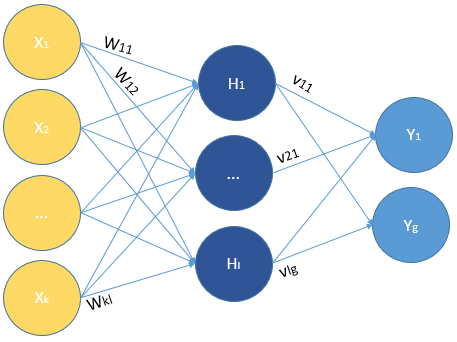
Therefore, the decision function for classification can be written as

|  |
| --- |
|  |

All the detail of SVM can be seen in [11]. In case of non-linear data, the data are mapped into high dimensional feature space using kernel function. In this study, we used several kernel functions such as, Polynomial, Radial Basis Function, and Linear.

* 1. **Artificial Neural Network**

Artificial Neural Network (ANN) algorithm, that is inspired by biological process of human brain. It represents the process of computation using interconnected neurons which send signal among them. Basically ANN which is comprised by three layers, such as Input Layer, Hidden Layer and Output Layer is called Multilayer Perceptron.



Gambar : ANN Architecture

The nodes in each layer are connected by weights. Given training data , where xi = (x1,x2,...,xk) and yi = (y1,y2,...,yg) . Input layer will receive values from x, and Hidden layer will calculate the result that can be expressed as

where wkl is weight of input layer that is connected to any hidden node, l is the number of hidden nodes, and f is the tanh activation function that can be written as

|  |
| --- |
|  |

The calculation in ouput layer uses purelin activation function which can be expressed as

Where Hl represents a hidden node in hidden layer, vi is the weight between hidden layer and output layer, b is the bias, and s represents pureline activation function that can be formulated as

the error will be calculate by computed value of e, which is

|  |
| --- |
|  |

In order to minimize the error, the computation in each layer will be repeated and all weights will be updated using backpropagation algorithm. Several optimized backpropagation algorithms such as Resilient Backpropagation, Levenberg Marquardt, Scale Conjugate Gradient are used in this study. For further details of those techniques can be seen in [12], [13]

1. **EXPERIMENT, RESULT AND DISCUSSION**

In this section, we describe the experiment scenario including result and disscussion. In the last, we compare the performance of ANN,ELM and SVM.

.

* 1. **Experiment**

The 386 clinical data are gained from UCI Machine Learning Repository which is available online in. The data consist 25 attributes including the target class. There are two types of target class, “ckd” and “notckd”. “ckd” refers to the patient who has kidney disease, and “notckd” refers to the patient who has no kidney disease. 237 sample data are diagnosed as chronic kidney disease and 149 sample data are diagnosed as healthy patient. Most of attributes have missing values . The detail of the dataset can be seen in table 1.

Table 1 : CKD dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Attribute Names** | **Data type** | ***NA*** |
| 1 | Age | Numerical | 9 |
| 2 | Blood Pressure | Numerical | 9 |
| 3 | Specific Gravity | Nominal | 45 |
| 4 | Albumin | Nominal | 44 |
| 5 | Sugar | Nominal | 46 |
| 6 | Red Blood Cell | Nominal | 143 |
| 7 | Pus Cell | Nominal | 62 |
| 8 | Pus Cell Clumps | Nominal | 4 |
| 9 | Bacteria | Nominal | 4 |
| 10 | Blood Glucose Random | Numerical | 42 |
| 11 | Blood Urea | Numerical | 18 |
| 12 | Serum Creatinine | Numerical | 16 |
| 13 | Sodium | Numerical | 81 |
| 14 | Potassium | Numerical | 82 |
| 15 | Hemoglobin | Numerical | 48 |
| 16 | Packed Cell Volume | Numerical | 66 |
| 17 | White Blood Cell Count | Numerical | 99 |
| 18 | Red Blood Cell Count | Numerical | 123 |
| 19 | Hypertension | Nominal | 2 |
| 20 | Diabetes Melitus | Nominal | 2 |
| 21 | Coronary Artery Disease | Nominal | 2 |
| 22 | Appetite | Nominal | 1 |
| 23 | Pedal Edema | Nominal | 1 |
| 24 | Anemia | Nominal | 1 |
| 25 | Class (Output) | Nominal | 0 |

In data preprocessing, KNN-Imputation algorithm has used to fill the missing values of the dataset. In this case, we set K = 3. Data Normalization [0,1] has perfomed to continues variables.

After data preprocessing, we divide the training and testing data using K-Fold Crosssvalidation, we set K = 10. Training data will be used for Training Process of ANN,SVM and ELM. The aim of Training proses is to train those algorithms to recognize ckd patterns in training data.Testing data will be used for validation process.

In training process. We tried several congfigurations in order to find best peformance of ANN,SVM and ELM. In particular for ELM, we try to add the odd number of hidden neurons between 3 – 301, we also applied those configuration for all ELM activation functions including sine, sigmoid, hard limit, radial basis and triangular radial basis. For ANN we try to add the odd number of hidden neurons between 49 – 101. We also applied different optimized backpropagation algorithm including *Levenberg-Marquardt, Resilient Backpropagation* and *Scaled Conjungat Gradient*. Especially for SVM we train the model using various kernel function including *Polynomial, Radial Basis Function, Gaussian* and *Linear.*

In this study, Accuracy, Sensitivity and Specificity are used to measure the peformance of each model of ANN,SVM and ELM. Accuracy, Sensitivity and Specificity can be expressed as

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |
|  | (5) |

TP adalah True Positif yang merujuk pada data pasien penyakit Ginjal Akut yang diprediksi secara benar. TN adalah True Negatif yang merujuk pada data pasien yang tidak mengidap Ginjal Akut yang diprediksi secara benar. FP adalah False Positive yang merujuk pada pasien yang tidak mengidap Ginjal Akut, tetapi diprediksi mengidap Ginjal Akut. FN adalah False Negative yang merujuk pada data pasien Ginjal Akut tetapi diprediksi tidak mengidap penyakit Ginjal Akut.

Seluruh Eksperimen pada penelitian ini menggunakan satu perangkat Laptop dengan Processcor Core i7 5th, RAM 4.GB, Sistem Operasi Windows 10, 64 Bit. *Source Code* program ditulis menggunakan perangkat lunak Matlab 2015. Untuk *Source Code* ELM menggunakan sumber dari[], sedangkan SVM dan ANN merupakan *build in function* dari Matlab 2015.

* 1. **Hasil**

Pada eksperimen ELM dengan variasi *Hidden Neuron* 3 - 301 terdapat pola yang sama, dimana pada setiap fungsi aktivasi memiliki peforma rata-rata *Accuracy* mencapai 0,95 – 0,97, rata-rata *Sensitivity* 0,97 – 0,99, rata-rata *Specificity* 0,93 – 0,95, jika jumlah *Hidden Neuron* berada pada rentang 49-101. Untuk peforma terbaik dari ELM dapat dilihat pada tabel 2. Pada Tabel 2 terlihat bahwa perbedaan fungsi aktivasi tidak mempengaruhi peforma ELM secara signifikan. Jumlah *Hidden Neuron* yang dipilih merupakan jumlah dengan nilai akurasi tertinggi, dalam hal ini ELM *Radial Basis Function* dengan 89 *Hidden Neuron* memiliki nilai Accuracy, Sensitivity, Specificity terbaik dari fungsi aktivasi lainnya.

Eksperimen pada SVM dengan variasi fungsi kernel dapat dilihat pada tabel 3. SVM dengan Fungsi Kernel Polynomial dan Linear memiliki peforma yang lebih baik , dibandingkan Kernel Gaussian dan Radial Basis. Sedangkan pada pada ANN, setiap variasi fungsi pelatihan menghasilkan peforma yang hampir sama, namun fungsi pelatihan dengan *Resilient Backpropation*, memberikan peforma yang lebih baik dibandingkan dengan dua variasi lainnya. Adapun peforma ANN dapat dilihat pada tabel 4.

Tabel : Performance of ELM with different activation functions and hidden neurons

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Performance | | Activation Functions and Hidden Neurons | | | | | |
| Sine (83) | | Sigmoid (75) | Hard Limit (99) | Triangular Radial Basis (83) | Radial Basis  (89) |
| Accuracy | Mean | | 0,974 | 0,973 | 0,959 | 0,972 | **0,978** |
| SD | | 0,022 | 0,023 | 0,031 | 0,023 | 0,021 |
| Sensitivity | Mean | | 0,992 | 0,989 | 0,982 | 0,992 | **0,994** |
| SD | | 0,017 | 0,020 | 0,027 | 0,017 | 0,016 |
| Specificity | Mean | | 0,952 | 0,954 | 0,931 | 0,946 | **0,956** |
| SD | | 0,049 | 0,049 | 0,063 | 0,048 | 0,043 |

Tabel : Performance of SVM with different kernel functions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Performance | | Support Vector Machine (Fungsi Kernel) | | | |
| Polynomial | Radial Basis | Gaussian | Linear |
| Accuracy | Mean | **0,988** | 0,917 | 0,917 | 0,986 |
| SD | 0,017 | 0,004 | 0,004 | 0,018 |
| Sensitivity | Mean | **0,981** | 1 | 1 | 0,984 |
| SD | 0,027 | 0 | 0 | 0,023 |
| Specificity | Mean | **0,999** | 0,786 | 0,786 | 0,989 |
| SD | 0,006 | 0,114 | 0,114 | 0,029 |

Tabel : Performance of ANN with different training function

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Peforma | | ANN Train Function | | |
| Levenberg Marquardt (85) | Resilient Backpropagation (61) | Scaled Conjungate Gradient (83) |
| Accuracy | Mean | 0,9994 | **0,9997** | 0,9994 |
| SD | 0,0036 | 0,0025 | 0,0036 |
| Sensitivity | Mean | 0,9991 | **0,9995** | 0,9991 |
| SD | 0,0058 | 0,0041 | 0,0059 |
| Specificity | Mean | 1 | **1** | 1 |
| SD | 0 | 0 | 0 |

Tabel : Peformance of ELM, ANN and SVM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Peforma | | Algoritma | | |
| ELM (Radial Basis) | ANN Resilient Backpropagation (61) | SVM-Kernel Polynomial |
| Accuracy | Mean | **0,978** | **0,9997** | **0,988** |
| SD | 0,021 | 0,0025 | 0,017 |
| Sensitivity | Mean | **0,994** | **0,9995** | **0,981** |
| SD | 0,016 | 0,0041 | 0,027 |
| Specificity | Mean | **0,956** | **1** | **0,999** |
| SD | 0,043 | 0 | 0,006 |
| Waktu Training (ms) |  | **0,0038** | **0,1691** | **0,0491** |

* 1. **Analisa**

Berdasarkan hasil evaluasi pengujian yang terdapat pada tabel 5, Model ANN-RP memiliki nilai *Accuracy* lebih baik yaitu 99.9 %, diikuti SVM-Polynomial 98.8 % dan ELM 97.8%. Nilai Sensitivitas ANN-RP dan ELM sama yaitu 99 % diikuti oleh SVM 98%, Sedangkan pada nilai Specificity, ANN-RP mencapai nilai 100 %, SVM, 99 % dan ELM 95.6%. Walaupun ELM memiliki peforma yang tidak sebaik ANN-RP dan SVM, namun perbedaan nilai *Accuracy, Sensitivity,* terhadap dua algoritma tidak signifikan, kecuali pada nilai specificity, perbedaan dengan ANN-RP mencapai 4.4%, dan 4,3 % dengan SVM. Keunggulan ELM yaitu memiliki waktu pelatihan yang sangat cepat dibandingkan dengan ANN-RP dan SVM, hal ini dikarenakan kedua algoritma tersebut memiliki banyak proses iterasi pada saat pelatihan, sedangkan ELM hanya menggunakan satu kali iterasi. Salah satu faktor lebih rendahnya peforma yang diperoleh oleh ELM adalah kedua algoritma ANN-RP dan SVM yang terdapat pada *Matlab 2015* secara default telah dilengkapi dengan optimisasi , sedangkan pada ELM yang digunakan merupakan standar algoritma tanpa optimisasi. Hasil dari penelitian ini tidak jauh berbeda dengan penelitian sebelumnya yang dilakukan oleh Chetty et al [14] yang melakukan seleksi fitur terhadap dataset penyakit ginjal yang sama dengan menggunakan *WraperSusetEval* dan *BestFirstSearch*, hasilnya terjadi peningkatan Akurasi pada algoritma yang digunakan sebesar 4 % pada Naive Bayes, 1.5 % pada SMO, dan 3.25% pada IBK. Walapun terjadi peningkatan peforma, namun penelitan tersebut tidak menjelaskan secara rinci berapa kali ekperimen tersebut dilakukan,dan tidak menghitung nilai *sensitivity* dan *specificity*. Jika dibandingkan dengan hasil peforma ELM pada penelitian ini, ELM mampu mencapai peforma akurasi rata-rata 97.8% tanpa dilakukan seleksi fitur, dari 100 eksperimen.

Pada penelitian lain yang dilakukan oleh *Z.Chen et al* [15] menggunakan model optimisasi dari Fuzzy Classifier, serta menghilangkan 4 atribut yang memiliki dari dataset, hasil pada tabel 6 menunjukan bahwa hasil yang didapatkan tidak jauh berbeda, dimana peforma terbaik dari algoritma *FuRES,*mencapai 99% untuk *Accuracy*, Sensitivity, dan Specificity, sedangkan peforma ELM, berbeda dengan selisih 2 % pada *Accuracy* dan 4% pada Specificity, adapun pada *Sensitivity* ELM mendapat nilai yang sama dengan algoritma FuRES. Faktor lain yang membuat peforma dari setiap model klasifikasi menampilkan hasil yang sangat baik adalah dataset yang digunakan memiliki kompleksitas yang rendah, sehingga nilai *Accuracy, Sensitivity,* dan *Specificity* hampir mendekati nilai 1.

Tabel : Perbandingan hasil penelitian sebelumnya

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Penelitian | Algoritma | Acc | Sen | Spec |
| Chetty et all (2015) | IBK | 1 | - | - |
| Naive Bayes | 0,99 | - | - |
| SMO | 0.98 | - | - |
| Z.Chen et al (2016) | FuRES | 0.99 | 0.99 | 0.99 |
| FOAM | 0.98 | 0.99 | 0.95 |
| PLS-DA | 0.95 | 1 | 0.89 |
| Penelitian ini | ELM | 0.97 | 0.99 | 0.95 |
| ANN-RP | 0,99 | 0.99 | 1 |
| SVM | 0.98 | 0.98 | 0.99 |

1. KESIMPULAN

Dari hasil eksperimen didapatkan beberapa fakta bahwa, peforma ELM dapat diterapkan dalam memprediksi penyakit Ginjal Akut, dengan *Accuracy* 97%, *Senstivity* 99%*,* dan *Specificity* 95%. Lebih rendahnya peforma Accuracy dan Specificity ELM dibandingkan dengan ANN-RP dan SVM dikarenakan ELM tanpa optimisasi, namun nilai *Sensitivitas* dari ELM lebih baik dari SVM dan sama dengan ANN-RP. Salah satu keunggulan yang tidak dimiliki oleh ANN-RP dan SVM adalah waktu pelatihan yang dibutuhkan oleh ELM yaitu 100 kali lebih cepat dari ANN-RP dan 10 kali lebih cepat dari SVM, dikarenakan proses pelatihan pada ELM hanya dengan 1 kali iterasi. Optimisasi pada algoritma ELM dan pra-pemrosesan data pada dataset seperti fitur seleksi dapat berpotensi meningkatkan peforma ELM.

1. DAFTAR PUSTAKA

[1] “Chronic kidney disease - ClinicalKey.” [Online]. Available: https://remote-lib.ui.ac.id:2194/#!/content/medical\_topic/21-s2.0-1014826. [Accessed: 25-Mar-2016].

[2] Ministry of Health, “Riset Kesehatan Dasar (Basic Health Research),” p. 303, 2013.

[3] J. A. Vassalotti, R. Centor, B. J. Turner, R. C. Greer, M. Choi, and T. D. Sequist, “A Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician.,” *Am. J. Med.*, vol. 129, no. 2, pp. 153–162.e7, 2015.

[4] G. Wang, K.-M. Lam, Z. Deng, and K.-S. Choi, “Prediction of mortality after radical cystectomy for bladder cancer by machine learning techniques,” *Comput. Biol. Med.*, vol. 63, pp. 124–132, 2015.

[5] Z. You, L. Li, Z. Ji, M. Li, and S. Guo, “Prediction of protein-protein interactions from amino acid sequences using extreme learning machine combined with auto covariance descriptor,” *Memetic Comput. (MC), 2013 …*, pp. 80–85, 2013.

[6] M. Fathurachman and U. Kalsum, “Heart disease diagnosis using extreme learning based neural networks,” *… Concept, Theory …*, 2014.

[7] S. Ismaeel, A. Miri, and D. Chourishi, “Using the Extreme Learning Machine ( ELM ) Technique for Heart Disease Diagnosis,” no. 1, pp. 1–3, 2015.

[8] G.-B. H. G.-B. Huang, Q.-Y. Z. Q.-Y. Zhu, and C.-K. S. C.-K. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” *2004 IEEE Int. Jt. Conf. Neural Networks (IEEE Cat. No.04CH37541)*, vol. 2, pp. 25–29, 2004.

[9] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification.,” *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 42, no. 2, pp. 513–29, 2012.

[10] “SVM - Support Vector Machines.” [Online]. Available: http://www.support-vector-machines.org/. [Accessed: 25-May-2016].

[11] C. J. C. J. C. J. C. C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.

[12] M. Riedmiller and H. Braun, “A direct adaptive method for faster backpropagation learning: The RPROP algorithm,” *IEEE Int. Conf. Neural Networks - Conf. Proc.*, vol. 1993-January, pp. 586–591, 1993.

[13] M. T. Hagan and M. B. Menhaj, “Training Feedforward Networks with the Marquardt Algorithm,” *IEEE Trans. Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[14] N. Chetty, “Role of Attributes Selection in Classification of Chronic Kidney Disease Patients,” pp. 0–5, 2015.

[15] Z. Chen, Z. Zhang, R. Zhu, Y. Xiang, and P. B. Harrington, “Diagnosis of patients with chronic kidney disease by using two fuzzy classifiers,” *Chemom. Intell. Lab. Syst.*, 2016.