Proposal Penelitian :

Deteksi Dini Penyakit Ginjal Kronis menggunakan *Extreme Learning Machine* berbasis *Neural Network*

Moh. Faturrahman

NPM : 1506706276

Fakultas Ilmu Komputer, Universitas Indonesia.

[moh.faturrahman@ui.ac.id](mailto:moh.faturrahman@ui.ac.id)

**Abstrak**

*Prevalensi penyakit Ginjal Kronis sebesar 0.2 % terjadi di Indonesia dan cenderung terjadi pada masyarakat pedesaan, tidak bersekolah, pekerja swasta, dan petani. Penyakit Ginjal Kronis dapat meningkatkan resiko terjadinya komplikasi terhadap penyakit lainnya seperti penyakit jantung, hipertensi, anemia, kematian, dan proses menuju pada stadium akhir. Deteksi dini pada penyakit ini untuk menunda proses penurunan fungsi ginjal, mencegah atau mengatur komplikasi. Penggunaan teknik Pembelajaran Mesin Extreme Learning Machine (ELM) berpotensi melakukan deteksi dini penyakit Ginjal Kronis secara cepat dan efisien untuk mendukung tujuan diatas. Keunggulan dari teknik ini adalah memiliki proses pembelajaran yang sangat cepat, tidak memiliki banyak parameter yang harus diinisialisasi, tidak terjebak pada local minima, dapat menggunakan non-differentiabale activation function. Tujuan dari penelitian ini mengukur peforma Accuracy, Sensitivity, Specificity dan menganalisa Algoritma Extreme Learning Machine dalam mendeteksi penyakit Ginjal Kronis.*

***Kata Kunci : Penyakit Ginjal Kronis, Pembelajaran Mesin, Extreme Learning Machine***

1. PENDAHULUAN

Penyakit Ginjal Kronis didefinisikan sebagai adanya tanda-tanda kerusakan pada ginjal ( abnomarlitas dalam darah, urin, atau pada tes gambar) lebih dari 3 bulan atau menurunnya *Glomerular Filtration Rate < 60 mL/min/*1.73 m2[1]⁠. Prevalensi penyakit Ginjal Kronis sebesar 0.2 % terjadi di Indonesia dan cenderung terjadi pada masyarakat pedesaan, tidak bersekolah, pekerja swasta, dan petani [2]. Penyakit Ginjal Kronis meningkatkan resiko komplikasi terhadap penyakit lainnya seperti penyakit jantung, hipertensi, anemia, kematian, dan proses menuju pada stadium akhir [3]. Deteksi Dini penyakit Ginjal dapat dilakukan pemeriksaan Laboratorium dengan estimasi nilai Glomerular Filtration Rate dan nilai Albuminnuria, serta menggunakan beberapa tes pendukung seperti Ultra Sound/CT Scan pada saluran kemih, atau tes serologic jika penderita punya riwayat diabetes atau hipertensi. Deteksi Dini bertujuan untuk menunda proses penurunan fungsi ginjal, mencegah atau mengatur komplikasi.

Pemeriksaan laboratorium memiliki kelemahan seperti biaya yang mahal dan proses yang berulang-ulang jika hasil yang didapat perlu dikomfirmasi. Penggunaan teknik Pembelajaran Mesin berpotensi melakukan deteksi dini penyakit Ginjal Kronis untuk mendukung tujuan diatas.

Pemanfaatan teknik Pembelajaran Mesin dalam bidang medis telah banyak dilakukan untuk membantu petugas medis dalam mengambil keputusan, memilih perawatan yang sesuai dengan pasien, membantu merancang pengobatan. Guang Jing et al, menggunakan ELM untuk memprediksi mortality rate dari pasien kanker kandung kemih yang telah dilakukan *Radical Cystectomy* (Operasi pengangkatan saluran kandung kemih) dengan akurasi rata-rata 80%, teknik ini menunjukan performa yang lebih baik dibandingkan dengan metode statistik [4]. Penelitian serupa juga dilakukan oleh Vukicevic et al, dengan menggunakan Neural Network dan dataset yang berbeda, hasil yang dicapai memiliki performa akuras 92.5% [5].

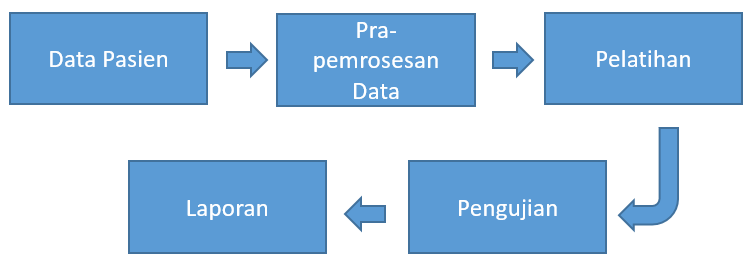
Khusus pada penyakit Ginjal Kronis, penggunaan Teknik Machine Learning Naive Bayes dalam mendeteksi penyakit Ginjal Kronis sudah dilakukan oleh Chetty et al, dengan menggunakan algoritma seleksi fitur pada pra-pemrosesan data, proses deteksi mampu menghasilkan akurasi 99%[6]. Dengan menggunakan dataset yang sama, Chen Zewei et al, menggunakan *Fuzzy Classfier* untuk mendeteksi penyakit Ginjal Akut, nilai rata-rata dari akurasi yang didapat mencapat 98% [7]. Berdasarkan beberapa hasil penelitain pemanfaatan teknik Pembelajaran Mesin pada bidang medis, menunjukan bahwa teknik Pembelajaran Mesin merupakan salah satu solusi effektif yang dapat mengatasi masalah-masalah pada bidang Medis.

Salah satu Teknik Machine Learning yang dapat digunakan pada bidang medis adalah Extreme Learning Machine [8], [9] [10], yang merupakan variansi dari Single Layer Feedforward Network (*SLFN*). Keunggulan dari teknik ini adalah memiliki proses pembelajaran yang sangat cepat, tidak memiliki banyak parameter yang harus diinisialisasi, tidak terjebak pada *local minima*, dapat menggunakan *non-differentiable activation function* [11].

Oleh karena itu tujuan dari penelitian ini mengukur peforma dari ELM dalam mendeteksi penyakit Ginjal Kronis. Adapaun penjelasan mengenai ELM akan dijelaskan pada bab selanjutnya.

Pada rancangan proposal penelitian ini ditulis melalui alur penulisan sebagai berikut, BAB II membahas skema penelitian yang akan dilakukan, mencakup tahapan-tahapan penelitian. BAB III, membahas teknik Pembelajaran Mesin yang digunakan, data klinik pasien, Perangkat Keras dan Perangkat Lunak yang digunakan. BAB IV membahas kesimpulan.

1. **METODE**



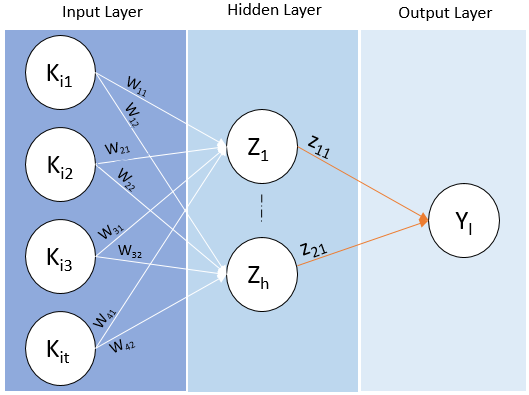
Gambar 1 : Alur Penelitian

Alur dari penelitian ini dapat dilihat pada gambar 1 diatas.

1. Tahap pertama dengan mengumpulkan Data Klinik Pasien, yang dapat diunduh pada *UCI Machine Learning Repository*
2. Tahap kedua, melakukan pra-pemrosesan data, yaitu mengisi nilai missing values dengan menggunakan nilai rata-rata pada variabel *continuous* dan nilai mode pada variabel *discrete*, melakukan normalisasi [0,1] pada variable yang memiliki nilai *continous,* melakukan pembagian data latih dan data uji menggunakan metode *K-Fold Crossvalidation* dengan nilai K= 10. Data Latih selanjutnya digunakan untuk melatih algoritma ELM dalam mengenali dan memprediksi data pasien pada penyakit Ginjal Akut. Data Uji digunakan untuk melakukan validasi dan mengukur performa dalam memprediksi penyakit Ginjal Akut pada pasien. Mekanisme pengujian yang digunakan akan dijelaskan pada bab selanjutnya.
3. Proses Latihan, dimana Algoritma ELM akan belajar mengenali pola data untuk memprediksi apakah suatu pasien mengidap penyakit Ginjal Akut atau tidak, proses pelatihan ini juga mencari konfigurasi parameter terbaik dalam memprediksi penyakit Ginjal Akut. Parameter yang akan diubah konfigurasinya adalah, fungsi aktivasi dan jumlah *Hidden Neuron* pada ELM. Hasil keluaran dari proses ini adalah sebuah model klasifikasi ELM yang memprediksi penyakit Ginjal Akut pada pasien.
4. Proses Pengujian, melakukan uji coba terhadap model ELM yang didapat pada proses latihan dengan memberikan data uji untuk diprediksi, data uji merupakan data yang tidak diketahui sebelumnya pada proses latihan. Algoritma prediksi Support Vector Machine (SVM), Backpropagation Neural Network (BPN), akan digunakan sebagai pembanding terhadap model ELM.
5. Analisa dan Laporan, Tahap analisa yaitu membandingkan hasil eksperimen pengujian model ELM,SVM dan BPN, terhadap nilai Accuracy, Sensitivity, Specifisity, melakukan analisa mendalam temuan-temuan yang terjadi saat proses Eksperimen. Menyajikan hasil Eksperimen dalam bentuk grafik agar mudah dianalisa secara visual. Laporan Analisa, adalah proses penulisan laporan akhir penelitian ditulis dalam format publikasi *paper*, yang terdiri dari Abstrak, Pendahuluan, Metode, Eksperimen, Hasil Eksperiment, Kesimpulan dan Saran.
6. EKSPERIMEN
   1. **Algoritma**

**Extreme Learning Machine**

Algoritma ELM merupakan modifikasi dari *Single Layer Feedforward Neural Network* (SLFN) [11]. Tidak seperti *Neural Network* pada umumnya yang menggunakan *Gradient Based Learning Algorithm* dalam mencari nilai bobot yang sesuai disetiap layer. ELM menggunakan *Moore Penrose Pseudo Invers Matrix* untuk mencari nilai bobot dengan sekali iterasi. Misalkan diberikan data sebanyak K sampel [Ki, Yi], dengan t jumlah atribut, dan h Hidden Node, i merupakan sampel ke i dalam data.



Gambar 2 : Arsitektur ELM

Hasil perhitungan pada *Hidden Layer* dipetakan dalam sebuah matrix Ak x h,Setiap nilai pada matrix A merupakan hasil dari fungsi aktivasi ***s( ki . wi + bi )*** dimana ki . wi adalah hasil operasi *inner product* dari vektor input ki dan vektor bobot wi . nilai bobot zi antara Hidden Layer dan Output Layer akan dihitung menggunakan

Z = A†Y

Dimana A† merupakan Moore Penrose Pseudo Inverse Matrix dari matriks A, vektor Y adalah nilai kelas dari sampel. Setelah mendapatkan nilai bobot Z maka, prediksi nilai Y dapat ditulis

Y = AZ

Secara umum, Algoritma ELM mempunyai tiga langkah :

1. Memberikan nilai bobot wi dan zi serta bias pada setiap layer, secara random.
2. Melakukan pemetaan nilai dari matrix A berdasarkan fungsi aktivasi dan menghitung nilai dari matrix A†
3. Menghitung bobot Z.
   1. **Dataset**

Dataset klinis penyakit Ginjal Kronis yang digunakan pada penelitian ini diambil dari *UCI Machine Learning Repository* yang tersedia secara online. Data yang dikumpulkan sebanyak 400 data, dengan jumlah atribut 25 termasuk atribut kelas. Terdapat 2 kelas pada dataset, “ckd” dan “notckd” (pasien dengan positif penyakit Ginjal Akut, dan pasien dengan negatif penyakit Ginjal Akut) dengan total jumlah “237” ckd dan 149 “notckd”, setiap atribut terdapat *missing values*. Detail dari dataset yang digunakan dapat dilihat pada tabel 1.

Tabel 1 : Keterangan Dataset Ginjal Akut

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input** | **Nama Atribut** | **Tipe data** | **Nilai/Satuan** | **Jumlah Missing Values** |
| 1 | Age | Numerical | Normalisai [0,1] | 9 |
| 2 | Blood Pressure | Numerical | Normalisai [0,1] | 9 |
| 3 | Specific Gravity | Nominal | (1.005,1.010,  1.015,1  .020,1.025) | 45 |
| 4 | Albumin | Nominal | 0,1,2,3,4,5 | 44 |
| 5 | Sugar | Nominal | 0,1,2,3,4,5 | 46 |
| 6 | Red Blood Cell | Nominal | Normal, Abnormal | 143 |
| 7 | Pus Cell | Nominal | Normal, Abnormal | 62 |
| 8 | Pus Cell Clumps | Nominal | Present, Not Present | 4 |
| 9 | Bacteria | Nominal | Present, Not Present | 4 |
| 10 | Blood Glucose Random | Numerical | mgs/dl | 42 |
| 11 | Blood Urea | Numerical | mgs/dl | 18 |
| 12 | Serum Creatinine | Numerical | mgs/dl | 16 |
| 13 | Sodium | Numerical | mEq/L | 81 |
| 14 | Potassium | Numerical | mEq/L | 82 |
| 15 | Hemoglobin | Numerical | Gms | 48 |
| 16 | Packed Cell Volume | Numerical | Normalisai [0,1] | 66 |
| 17 | White Blood Cell Count | Numerical | cells/cmm | 99 |
| 18 | Red Blood Cell Count | Numerical | million/cmm | 123 |
| 19 | Hypertension | Nominal | Yes, No | 2 |
| 20 | Diabetes Melitus | Nominal | Yes, No | 2 |
| 21 | Coronary Artery Disease | Nominal | Yes, No | 2 |
| 22 | Appetite | Nominal | Good, Poor | 1 |
| 23 | Pedal Edema | Nominal | Yes, No | 1 |
| 24 | Anemia | Nominal | Yes, No | 1 |
| 25 | Class (Output) | Nominal | ckd,notckd | 0 |

* 1. ***Repeated K-Fold Crossvalidation***

*Repeated K-Fold Crossvalidation* merupakan salah satu metode dalam membagi dataset menjadi data latih dan data uji, sebanyak k data. Misalkan dengan memilih nilai k = 10, maka dataset akan dibagi menjadi 10 subset data uji dan data latih yang tidak beririsan. Dengan Proporsi 75 % untuk data latih, 25 % untuk data uji. Hasil pembagian data latih dan uji diulang sebanyak R kali.

* 1. **Evaluasi**

Nilai yang akan diprediksi berdasarkan data klinis pasien adalah “ckd” dan “notckd”, “ckd” merujuk pada pasien yang terdiagnosis Ginjal Akut, sedangkan “notckd” merujuk pada pasien yang tidak didiagnosis Ginjal Akut. Tujuan model prediksi ELM adalah memperkecil kesalahan dalam memprediksi apakah pasien termasuk dalam kategori “ckd” atau “notckd”. Untuk mengetahui peforma dari model ELM dilakukan proses evaluasi Algoritma, yang diukur dengan menghitung nilai Akurasi, Sensitivity, Spesificity dengan menggunakan formula dibawah ini.

|  |  |
| --- | --- |
|  | (7) |
|  | (8) |
|  | (9) |

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.

* 1. **Perangkat Pendukung**

Seluruh Eksperimen pada penelitian ini menggunakan satu perangkat Laptop dengan Processcor Core i7 5th, RAM 4.GB, Windows Operating System 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 DAN ANALISA**
   1. **Pra-pemrosesan Data**

Dataset yang telah diperoleh dilakukan proses imputasi untuk mengisi *missing values* dari tiap variabel menggunakan Algoritma KNN, dengan konfigurasi nilai K = 3. Seluruh data dilakukan normalisasi pada skala [0 1]. Proses pembagian data latih dan uji, menggunakan metode *Repeated K-Fold Crossvalidation* yang telah dijelaskan pada bab sebelumnya.

* 1. **Hasil Eksperimen**

Eksperimen dimulai dengan mencari variasi fungsi aktivasi dan jumlah *Hidden Neuron* pada algoritma ELM,jumlah Hidden Neuron dan Fungsi Pelatihan pada ANN, dan variasi fungsi *kernel* pada SVM. Fungsi aktivasi yang digunakan pada ELM adalah *Sigmoid, Sine, Triangular Basis Function, Hard Limit, Radial Basis,* Jumlah Hidden Neuron pada ELM yang diujikan yaitu bilangan ganjil dari 3 - 301, ANN 49 -101 dengan variasi fungsi pelatihan *Levenberg-Marquard, Resilient Backpropagation,* dan *Scaled Conjungate Graduent*, sedangkan pada SVM kernel yang digunakan adalah *Polynomial, Radial Basis Function, Gaussian*, dan *Linear*.

Pada eksperimen ELM dengan variasi Hidden Neuron 3 - 301 terdapat pola yang sama, dimana pada setiap fungsi aktivasi 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 2 Pengaruh Hidden Neuron dan Fungsi Aktivasi pada ELM

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Peforma | | Fungsi Aktivasi (Hidden Neuron) | | | | |
| 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 3 : Peforma perbedaan fungsi kernel pada SVM

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Peforma | | Support Vector Machine | | | |
| 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 4 : Peforma ANN dengan variasi Fungsi Pelatihan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 5 : Perbandingan ELM, ANN dan SVM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Peforma | | Algoritma | | |
| ELM (Radial Basis) | ANN Resilient Backpropagation | 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 |

1. **KESIMPULAN**

Penyakit Ginjal Kronis mengakibatkan pasien memiliki resiko tinggi terkena komplikasi terhadap penyakit lain. Dengan melakukan deteksi dini, resiko terjadinya komplikasi dapat diperkecil, dengan memberikan perawatan untuk menghambat perkembangan penyakit Ginjal. Metode Pembelajaran ELM akan digunakan untuk mendeteksi penyakit Ginjal Akut secara dini berdasarkan data klinis pasien, untuk mendukung tenaga medis dalam mengambil keputusan atau tindakan terdahap pasien. Model ELM diharapkan dapat mengatasi kelemahan-kelemahan pada pemeriksaan penyakit Ginjal Akut seperti biaya yang mahal, proses yang berulang-ulang, serta efisien.

**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] A. M. Vukicevic, G. R. Jovicic, M. M. Stojadinovic, R. I. Prelevic, and N. D. Filipovic, “Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced bladder cancer outcome,” *Expert Syst. Appl.*, vol. 41, no. 18, pp. 8092–8100, 2014.

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

[7] 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.

[8] 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.

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

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

[11] 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.