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

|  |  |  |
| --- | --- | --- |
| Moh.Faturrahman | Ito Wasito | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Fakultas Ilmu Komputer, Universitas Indonesia, Depok Indonesia | Fakultas Ilmu Komputer Universitas Indonesia, Depok. Indonesia | Fakultas Ilmu Komputer Universitas Indonesia, Depok Indonesia |

***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. Pada penelitian ini didapatkan bahwa peforma ELM dalam mendeteksi penyakit ginjal kronis mencapai 97.8 % Accuracy, 99.4% Sensitivity, dan 95.6 %. Dalam hal ini ELM dapat dijadikan solusi alternatif dalam mendeteksi penyakit Ginjal Kronis.*

***Kata Kunci :* Extreme Learning Machine, Pembelajaran Mesin, Penyakit Ginjal Kronis, Deteksi Dini, *Neural Network***

1. **PENDAHULUAN**

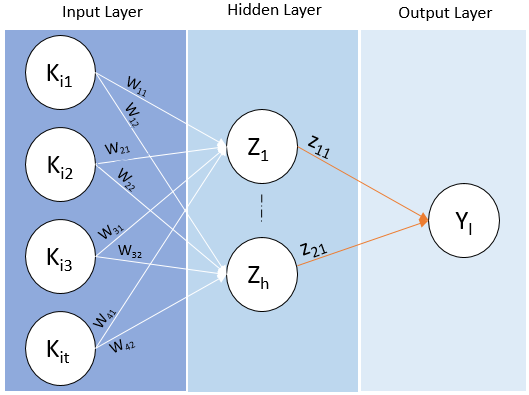
Penyakit Ginjal Kronis didefinisikan sebagai adanya tanda-tanda kerusakan pada ginjal ( abnomarlitas dalam darah, urin, atau pada tes gambar) lebih dari 3 bulan atau menurunnya *Glomerular Filtration Rate < 60 mL/min/*1.73 m2[1]⁠. Prevalensi penyakit Ginjal Kronis sebesar 0.2 % terjadi di Indonesia dan 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].

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.

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.



Gambar 2 : Arsitektur ELM

1. PENELITIAN TERKAIT

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].

1. **DESAIN DAN IMPLEMENTASI**

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.

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.
4. **EKSPERIMEN, HASIL DAN ANALISA**

Pada bagian ini akan dijelaskan model prediksi dari ELM untuk mendeteksi penyakit Ginjal Kronis, Analisa hasil eksperimen, dan perbandingan model ELM dengan ANN dan SVM.

* 1. **Eksperimen**

Dataset klinis penyakit Ginjal Kronis yang digunakan pada penelitian ini diambil dari *UCI Machine Learning Repository* yang tersedia secara online. Data yang dikumpulkan sebanyak 386 data, dengan jumlah 25 atribut 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* (NA). Detail dari dataset yang digunakan dapat dilihat pada tabel 1.

Tabel 1 : Atribut dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Nama Atribut** | **Tipe data** | ***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 |

Tahap kedua, melakukan pra-pemrosesan data, yaitu mengisi nilai missing values dengan menggunakan algoritma Imputasi KNN, dengan nilai K= 3, melakukan normalisasi [0,1] pada variable yang memiliki nilai *continous.*

Setelah pra-pemrosesan data selesai, dilakukan 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.

Proses selanjutnya yaitu melakukan pelatihan pada model ELM dengan menggunakan data latih yang telah dibagi pada proses sebelumnya. Selain melakukan pelatihan, 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*.

. Model ELM,ANN,dan SVM selanjutnya diuji dengan menggunakan data uji. Untuk mengetahui peforma dari tiap model dilakukan proses evaluasi algoritma, yang diukur dengan menghitung nilai *Accuracy*, *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.

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 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 : Peforma ELM dengan variasi jumlah hidden neuron dan fungsi aktivasi

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 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 : Peforma SVM dengan variasi Fungsi Kernel

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Peforma | | 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 : Peforma ANN dengan variasi Fungsi Pelatihan dan Hidden Neuron

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 : Perbandingan ELM, ANN dan 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 dua penelitian sebelumnya yang dilakukan oleh Chetty et al dan Z.Chen et al, penelitian pertama melakukan fitur seleksi terhadap dataset, sedangkan penelitian kedua menggunakan model optimisasi dari Fuzzy Classifier, hasil pada tabel 6 menunjukan bahwa cenderung hampir sama, disisi lain optimisasi yang dilakukan pada algoritma maupun pada dataset tidak berdampak signifikan pada peforma model klasifikasi, hal ini dapat terjadi jika dataset yang digunakan memiliki kompleksitas yang rendah.

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
2. DAFTAR PUSTAKA