

Implementasi Algoritma Genetika pada *k*-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang

Rizki Tri Prasetyo¹, Ali Akbar Rismayadi², Iedam Fardian Anshori³

¹Universitas BSI
e-mail: rizki.rte@bsi.ac.id

²Universitas BSI
e-mail: ali.aak@bsi.ac.id

³Universitas BSI
e-mail: iedam.iao@bsi.ac.id

Abstrak

Kerusakan tulang belakang dialami oleh sekitar dua pertiga orang dewasa serta termasuk ke dalam penyakit yang paling umum kedua setelah sakit kepala. Klasifikasi gangguan tulang belakang sulit dilakukan karena membutuhkan *radiologist* untuk menganalisa citra *Magnetic Resonance Imaging* (MRI). Penggunaan *Computer Aided Diagnosis* (CAD) System dapat membantu *radiologist* untuk mendeteksi kelainan pada tulang belakang dengan lebih optimal. *Dataset vertebral column* memiliki tiga kelas sebagai klasifikasi penyakit kerusakan tulang belakang yaitu, *herniated disk*, *spondylolisthesis* dan kelas normal yang diambil berdasarkan hasil ekstraksi citra MRI. Dataset akan diolah dalam lima eksperimen berdasarkan validasi dataset menggunakan *split validation* dengan pembagian *data training* dan *data testing* yang bervariasi. Pada penelitian ini diusulkan implementasi algoritma genetika pada algoritma *k*-nearest neighbours untuk meningkatkan akurasi dari klasifikasi gangguan tulang belakang. Algoritma genetika digunakan untuk fitur seleksi dan optimasi parameter algoritma *k*-nearest neighbours. Hasil penelitian menunjukkan bahwa metode yang diusulkan menghasilkan peningkatan yang signifikan dalam klasifikasi kerusakan pada tulang belakang. Metode yang diusulkan menghasilkan rata-rata akurasi sebesar 93% dari lima eksperimen. Hasil ini lebih baik dari algoritma *k*-nearest neighbours yang menghasilkan rata-rata akurasi hanya sebesar 82.54%.

Kata kunci: algoritma genetika, *k*-nearest neighbours, kerusakan tulang belakang, *vertebral*

Abstract

Spinal disorder is experienced by about two-thirds of adults and is included in the second most common disease after headaches. Classification of spinal disorders is difficult because it requires a radiologist to analyze Magnetic Resonance Imaging (MRI) images. The use of Computer Aided Diagnosis (CAD) System can help radiologists to detect abnormalities in the spine more optimally. The vertebral column dataset has three classes as a classification of spinal disorders, namely, herniated disk, spondylolisthesis and normal classes taken based on MRI Image extraction. The dataset will be processed in five experiments based on dataset validation using split validation with various training data and testing data. In this study proposed the implementation of genetic algorithms in the k-nearest neighbors algorithm to improve the accuracy of the classification of spinal disorders. Genetic algorithms are used for algorithm feature selection and parameter optimization of k-nearest neighbors. The results showed that the proposed method produced a significant increase in the classification of spinal disorder. The proposed method produces an average accuracy of 93% from five experiments. This result is better than the k-nearest neighbors algorithm which produces an average accuracy of only 82.54%.

Keywords: genetic algorithm, *k*-nearest neighbours, spinal disorder, *vertebral column*.

1. Pendahuluan

Nyeri punggung bagian bawah (*low-back pain*) adalah gangguan tulang belakang bagian bawah (pinggang) yang merupakan gangguan kesehatan yang secara signifikan mempengaruhi 50-80% dari populasi (White & Gordon, 1982). Sekitar dua pertiga orang dewasa menderita penyakit ini (Eyo & Einstein, 2001) serta termasuk ke dalam penyakit yang paling umum kedua setelah sakit kepala (Brant-Zawadzki, Dennis, Gade, & Weinstein, 2000).

Radiografi yang pada mulanya sering digunakan untuk meneliti gangguan tulang belakang saat ini menjadi kurang menarik semenjak ditemukan *Magnetic Resonance Imaging* yang mampu menggambarkan sumsum tulang belakang dan jaringan lunak lainnya dengan lebih baik (Unal & Kocer, 2013). Meskipun MRI adalah solusi yang sangat baik, menganalisa gambar MRI membutuhkan pengalaman yang cukup, kesalahan pada analisa gambar MRI dapat mengarah pada pengobatan yang salah (Videman, Nummi, Battie, & Gill, 1994). Penggunaan *Computer-Aided Diagnosis* (CAD) *system* dapat membantu *radiologist* untuk mendeteksi kelainan pada medis dilihat dengan menggunakan *pattern recognition* dan *machine learning* (Unal & Kocer, 2013).

Penelitian sebelumnya pada dataset *vertebral column* dilakukan oleh beberapa peneliti untuk dapat mengklasifikasikan gangguan pada tulang belakang menggunakan berbagai metode antara lain: *k-nearest neighbours* (Prasetyo & Riana, 2015), *Artificial Neural Network* (ANN) (Abdrabou, 2012), *Naïve Bayes* (NB) (Reddy & Kodali, 2012), *backpropagation* (Unal & Kocer, 2013), *Generalized Regression Neural Network* (GRNN) (Ansari & Sajjad, 2013) serta *embedded reject option* (Da Rocha Neto, Sousa, De A. Barreto, & Cardoso, 2011).

Menurut (Kataria & Singh, 2013), algoritma *k-nearest neighbours* merupakan algoritma klasifikasi *non-parametric* konvensional yang menghasilkan kinerja yang baik, mudah dipahami dan diimplementasikan (Wu et al., 2008) serta paling sederhana diantara algoritma *machine learning* lain (Gorunescu, 2011). *k-nearest neighbours* merepresentasikan teknik klasifikasi sebuah objek berdasarkan objek tetangga (*k*) terdekat (Harrington, 2012).

k-nearest neighbours sangat cocok digunakan pada dataset yang memiliki banyak label/kelas (Wu et al., 2008). Terdapat beberapa faktor kunci yang mempengaruhi performa dari *k-nearest neighbours*. Pertama, pemilihan parameter *k* (Wu et al., 2008). Jika *k* terlalu kecil, maka hasil dapat menjadi sensitif terhadap *noise* yang dapat mengarahkan algoritma menuju *overfitting* (Larose & Larose, 2014). Di sisi lain, jika *k* terlalu besar, maka tetangga terdekat mungkin terlalu banyak dari kelas lain (Wu et al., 2008). Pilihan paling tepat adalah menentukan *k* berdasarkan datasetnya (Gorunescu, 2011).

Faktor kunci lain yang mempengaruhi performa *k-nearest neighbours* adalah performa yang dihasilkan dapat sangat terdegradasi oleh kehadiran fitur yang bersifat *noise* atau tidak relevan (Han, Pei, & Kamber, 2011), atau jika skala fitur tidak konsisten dengan kepentingannya (Gorunescu, 2011).

Dataset *vertebral column* termasuk data medis yang melibatkan model diagnostik yang biasanya berdimensi tinggi. Data berdimensi tinggi meningkatkan kompleksitas klasifikasi dan mengurangi performa algoritma (Bharti & Singh, 2014), ini merupakan hambatan serius yang dialami algoritma *data mining* yang biasa dikenal dengan "curse of dimensionality" (Maimon & Rokach, 2005). Sehingga diperlukan reduksi dimensi data akan tetapi tetap mempertahankan informasi penting.

Ekstraksi fitur (Liu, Chai, Yu, & Tang, 2015) dan seleksi fitur (Jirapech-Umpai & Aitken, 2005) merupakan metode utama yang dapat digunakan untuk mengurangi dimensi data. Proses *data mining* membutuhkan biaya komputasi yang tinggi apabila menghadapi dataset yang besar. Mengurangi dimensi dapat mengurangi biaya komputasi (Maimon & Rokach, 2005), mengurangi waktu dan *memory* secara efektif (Shilaskar & Ghatol, 2013).

Tujuan utama seleksi fitur adalah untuk mengurangi jumlah fitur yang digunakan dalam klasifikasi serta mempertahankan akurasi klasifikasi yang dapat diterima (Raymer, Punch, Goodman, Kuhn, & Jain, 2000). Seleksi fitur dapat berdampak besar pada keefektifan algoritma klasifikasi yang diterapkan (Jain & Zongker, 1997), dalam beberapa kasus, sebagai hasil dari seleksi fitur, akurasi pada klasifikasi dapat ditingkatkan (Maimon & Rokach, 2005).

Masalah pada saat seleksi fitur adalah menentukan kandidat fitur yang akan digunakan pada saat klasifikasi (Jain & Zongker, 1997). Berbagai teknik seleksi fitur pun banyak direkomendasikan oleh peneliti lain, seperti: *relief algorithm* (Tiwari, 2014), *Particle Swarm Optimization* (PSO) (Deepa, 2013), *Genetic Algorithm* (GA) (Villar, Fernández, & Herrera, 2010) (Prasetio & Pratiwi, 2015).

Algoritma genetika sering digunakan untuk optimasi. Algoritma genetika memiliki lebih sedikit kecenderungan untuk terjebak dalam *local minima* (Gorunescu, 2011), serta termasuk salah satu metode optimasi yang mutakhir (Witten, Frank, Hall, & Pal, 2016). Algoritma genetika mampu secara efektif mengeksplorasi ruang pencarian yang besar, yang biasanya diperlukan dalam kasus pemilihan fitur (Villar et al., 2010).

Pada penelitian ini, akan diintegrasikan algoritma genetika untuk seleksi fitur dan optimasi parameter *k-nearest neighbours* untuk klasifikasi kerusakan tulang belakang. Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi dari dataset vertebral column dengan menerapkan algoritma genetika sebagai seleksi fitur serta meningkatkan kinerja algoritma klasifikasi *k-nearest neighbours* dengan mengoptimalkan parameter *k* menggunakan algoritma genetika.

2. Metode Penelitian

Penelitian ini mengusulkan metodologi berdasarkan paradigma *data mining*. Paradigma ini mengintegrasikan pencarian heuristik yang terinspirasi oleh evolusi alami yang disebut algoritma genetika dengan algoritma klasifikasi paling sederhana dan paling banyak digunakan, *k-nearest neighbours*

Algoritma Genetika

Algoritma genetika merupakan algoritma evolusioner yang paling populer yang mana algoritma ini menggunakan prinsip dasar dari seleksi alam yang diperkenalkan oleh Charles Darwin. Algoritma genetika diterapkan sebagai pendekatan untuk mengidentifikasi pencarian nilai dan solusi bagi berbagai permasalahan optimasi (Gorunescu, 2011). Terdapat beberapa kelebihan dari algoritma genetika dibandingkan algoritma optimasi tradisional lainnya.

Kelebihan algoritma genetika dua diantaranya yaitu kemampuan untuk menangani permasalahan kompleks dan paralel. Algoritma genetika dapat menganani berbagai macam optimasi tergantung pada fungsi objektifnya (*fitness*) seimbang atau tidak seimbang, linier atau tidak linier, berkesinambungan atau tak berkesinambungan, atau dengan *random noise*.

Intisari dari algoritma genetika mencakup pengkodean fungsi optimasi sebagai *array* berisi *bit-bit* atau karakter berupa *string* untuk menggambarkan kromosom, operasi manipulasi string dengan operator genetik, dan seleksi sesuai dengan *fitness*, dengan tujuan untuk menemukan solusi yang baik dan optimal terhadap masalah yang sedang dihadapi. Algoritma genetika memiliki tiga operator genetik utama yaitu: (Gorunescu, 2011)

1. *Crossover* yaitu proses menukar bagian dari solusi (kromosom) dengan bagian "*parent*" lain untuk menghasilkan jenis kromosom yang berbeda yang mungkin menjadi solusi baru untuk menyelesaikan permasalahan. Peran utamanya adalah untuk memberikan pencampuran solusi dan konvergensi dalam sub ruang (menghasilkan solusi yang baru).
2. Mutasi yaitu pergantian salah satu bagian solusi yang dipilih secara acak, yang meningkatkan keragaman dari populasi dan menghasilkan mekanisme untuk menghindari minimum lokal.
3. Seleksi *fittest* atau elitisme yaitu penggunaan solusi dengan nilai *fitness* yang tinggi untuk lulus ke generasi berikutnya, yang sering dilakukan dalam hal beberapa bentuk seleksi dari solusi terbaik

Algoritma 1. Genetic Algorithm

Begin

```

INITIALIZE population with random
candidate solutions;
EVALUATE each candidate;
REPEAT UNTIL (termination condition is
satisfied) DO
  1. SELECT parents;
  2. RECOMBINE pairs of parents;
  3. MUTATE the resulting offspring;
  4. EVALUATE new candidates;
  5. SELECT individuals for the next
generation;

```

End

***k*-nearest neighbours**

Algoritma *k*-nearest neighbours adalah suatu metode yang merupakan algoritma supervised (Wu et al., 2008). Perbedaan antara supervised learning dengan unsupervised learning adalah pada supervised learning bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada unsupervised learning, data belum memiliki pola apapun, dan tujuan unsupervised learning untuk menemukan pola dalam sebuah data (Han et al., 2011).

Tujuan dari algoritma k-NN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training samples (Larose & Larose, 2014). Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *k*-nearest neighbours.

Proses pengklasifikasian, algoritma ini tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Algoritma *k*-nearest neighbours menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari sampel uji yang baru. Pemilihan terhadap teknik perhitungan jarak merupakan hal penting lainnya. Biasanya euclidean distance (1) dan manhattan distance (2) digunakan untuk menghitung jarak pada algoritma *k*-nearest neighbours (Han et al., 2011). Berikut ini merupakan formula untuk kedua teknik perhitung jarak tersebut:

$$d(x, y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \quad (1)$$

$$d(x, y) = \sqrt{\sum_{k=1}^n |X_k - Y_k|} \quad (2)$$

Algoritma 2. *k*-nearest neighbours

Input: D , the set of training objects, the test object, \mathbf{z} , which is a vector of attribute values, and L , the set of classes used to label the objects

Output: $c_z \in L$, the class of \mathbf{z}

foreach object $\mathbf{y} \in D$ **do**

 | Compute $d(\mathbf{z}, \mathbf{y})$, the distance between \mathbf{z} and \mathbf{y} ;

end

Klasifikasi menggunakan *k*-nearest neighbours dilakukan melalui tahapan sebagai berikut:

1. Hitung jarak (kemiripan) antara semua data yang dilatih terhadap data baru.

2. Identifikasi sejumlah k objek yang paling dekat.
3. Tetapkan label berdasarkan objek yang paling dekat.

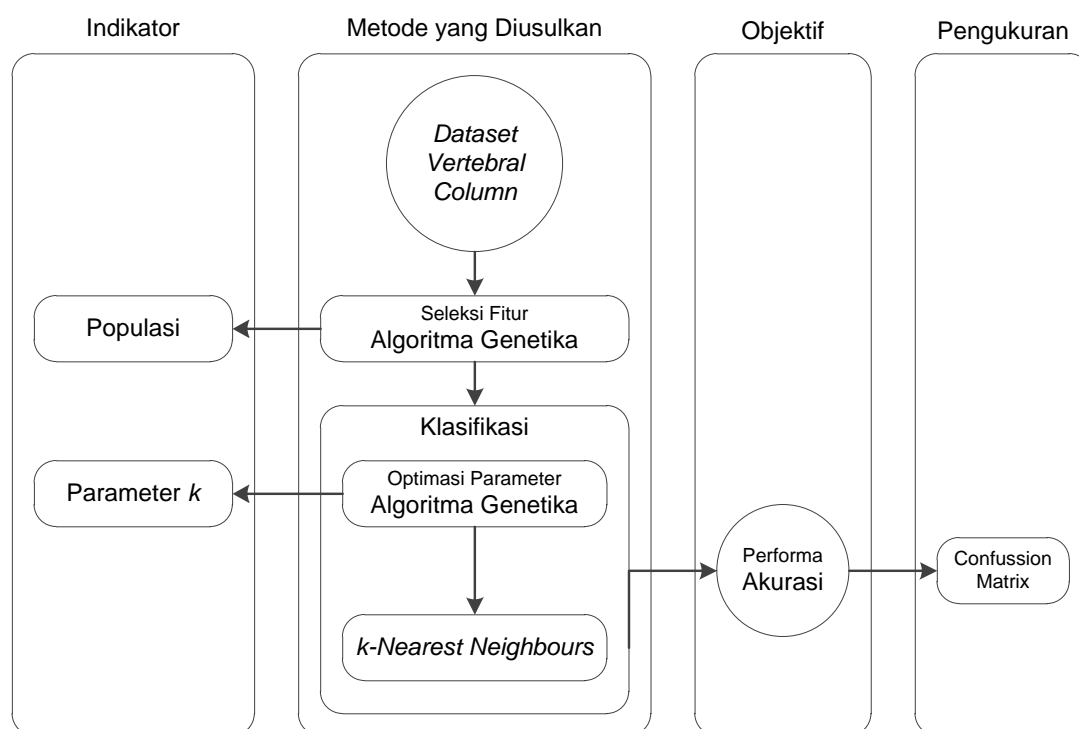
Metode yang diusulkan

Metode yang diusulkan mengintegrasikan algoritma genetika untuk seleksi fitur dan algoritma *k*-nearest neighbours yang telah dioptimasi untuk klasifikasi kerusakan tulang belakang. Penelitian dimulai pada tahapan pembagian dataset menjadi data training dan data testing menggunakan split validation dengan rasio pembagian yang bervariasi.

Algoritma genetika diimplementasikan untuk seleksi fitur. Seleksi fitur digunakan untuk menemukan fitur paling optimal yang merepresentasikan label/kelas pada dataset vertebral column. Algoritma *k*-nearest neighbours yang telah dioptimasi kemudian diterapkan untuk mengklasifikasi dataset. Setelah itu, hitung akurasi yang dihasilkan oleh algoritma *k*-nearest neighbours. Jika akurasi yang diinginkan belum tercapai, ulangi proses ini menggunakan algoritma genetika hingga mendapatkan hasil akurasi paling baik. Metode yang diusulkan dapat dilihat pada Gambar 1.

Hasil yang didapatkan pada penelitian ini kemudian dibandingkan dengan algoritma seleksi fitur lain seperti, backward elimination (Guyon & Elisseeff, 2003) (Abe, 2005a) (Derksen & Keselman, 1992), forward selection (Blanchet, Legendre, & Borcard, 2008) (Abe, 2005b) (Jain & Zongker, 1997) and greedy feature selection (Dyer, Sankaranarayanan, & Baraniuk, 2013) (Vafaie & Imam, 1994) (Farahat, Ghodsi, & Kamel, 2013). Perbandingan ini dilakukan untuk membuktikan performa algoritma genetika lebih baik dari algoritma seleksi fitur yang lain.

Hasil yang diperoleh dari metode yang diusulkan kemudian diuji dengan hasil yang diperoleh dari algoritma *k*-nearest neighbours dengan parameter standar untuk menentukan hasil kinerja metode yang diusulkan meningkatkan akurasi dataset vertebral column secara signifikan menggunakan *t*-test significance test (Prasetio & Pratiwi, 2015) (Setiyorini & Wahono, 2014) (Prasetio & Riana, 2015)



3. Hasil dan Pembahasan

Penelitian dilakukan dalam lima eksperimen berdasarkan variasi rasio *split validation*. Pada setiap eksperimen akan diterapkan algoritma *k-nearest neighbours* dan metode yang diusulkan untuk klasifikasi kerusakan tulang belakang. Hasil eksperimen dapat dilihat pada Tabel 1.

Hasil eksperimen yang tertuang pada Tabel 1 menyatakan bahwa metode yang diusulkan dapat menghasilkan rata-rata nilai akurasi yang paling baik jika dibandingkan dengan nilai akurasi yang dihasilkan oleh algoritma *k-nearest neighbours* saja.

Tabel 1. Hasil Eksperimen Klasifikasi

Rasio <i>Split validation</i>	<i>k-nearest neighbours</i>	Metode yang Diusulkan
0,9	83,87%	100%
0,8	85,48%	93,55%
0,7	84,95%	91,40%
0,6	82,26%	90,32%
0,5	76,13%	88,39%

Algoritma *k-nearest neighbours* menghasilkan nilai akurasi rata-rata sebesar 82.54% dengan hasil terbaik diperoleh dengan rasio *split validation* 0.8 sebesar 85.48%. Sementara metode yang diusulkan dapat menghasilkan nilai akurasi rata-rata sebesar 93% dengan hasil terbaik diperoleh dengan rasio *split validation* 0.9 sebesar 100%.

Berdasarkan hasil eksperimen pada Tabel 1 dapat disimpulkan bahwa metode yang diusulkan dapat meningkatkan performa kinerja klasifikasi algoritma *k-nearest neighbours*. Rata-rata peningkatan performa yang dihasilkan adalah 10.194% dengan peningkatan terbesar didapatkan pada rasio *split validation* 0.9 sebesar 16.130%. Untuk mengetahui metode yang diusulkan dapat meningkatkan performa algoritma *k-nearest neighbours* secara signifikan, maka pengujian menggunakan uji beda *t-Test Paired Two Sample for Means* (Prasetio & Riana, 2015) (Prasetio & Pratiwi, 2015) dilakukan.

Hasil pengujian *t-Test Paired Two Sample for Means* menghasilkan *P value* dari *t-Test* sebesar 0.0045. Hasil ini menunjukkan bahwa penerapan algoritma genetika untuk seleksi fitur dan optimasi parameter pada parameter *k* dapat meningkatkan performa akurasi algoritma

k-nearest neighbours secara signifikan, ditandai dengan *P Value* dari *t-Test* < 0.05 . Hasil pengujian *t-Test* dapat dilihat pada Table 2.

Tabel 2. Hasil Pengujian *t-Test*

	<i>k-nearest neighbours</i>	Metode yang Diusulkan
Mean	0.82538	0.92732
Variance	0.001434677	0.001998447
Observations	5	5
Pearson Correlation Hypothesized Mean Difference	0.551729248	0
df	4	
<i>t Stat</i>	-	
<i>P(T<=t) one-tail</i>	5.762573124	0.002249754
<i>t Critical one-tail</i>	2.131846786	
<i>P(T<=t) two-tail</i>	0.004499508	
<i>t Critical two-tail</i>	2.776445105	

Berdasarkan hasil pengujian *t-Test significance test* dapat disimpulkan bahwa algoritma genetika memiliki kinerja yang baik untuk meningkatkan performa algoritma klasifikasi melalui tahapan seleksi fitur.

Eksperimen berikutnya dilakukan pada algoritma *k-nearest neighbours* dengan parameter *k* yang telah dioptimasi, akan tetapi algoritma yang digunakan untuk seleksi fitur menggunakan *forward selection*, *backward elimination*, *greedy forward selection* dan *greedy backward selection*. Eksperimen ini dilakukan untuk mengetahui jika algoritma genetika merupakan algoritma terbaik untuk melakukan seleksi fitur. Hasil eksperimen dapat dilihat pada Tabel 3.

Hasil eksperimen yang tertuang pada Tabel 3 menyatakan bahwa algoritma genetika merupakan algoritma seleksi fitur yang menghasilkan performa klasifikasi

terbaik jika dibandingkan dengan algoritma seleksi fitur lain. Perolehan rata-rata nilai akurasi masing-masing algoritma seleksi fitur sebesar 90,04% untuk *forward selection*, 90,28% untuk *backward elimination*, 88,74% untuk *greedy forward selection* dan 90,28% untuk *greedy backward selection* sementara perolehan rata-rata nilai akurasi algoritma genetika sebesar 93%.

Berdasarkan seluruh hasil eksperimen yang telah diperoleh dapat disimpulkan bahwa algoritma genetika telah berhasil meningkatkan kinerja algoritma *k-nearest neighbours* untuk klasifikasi kerusakan tulang belakang melalui optimasi parameter *k* dan seleksi fitur. Pada penelitian yang dilakukan peneliti lain sebelumnya, telah diperoleh beberapa hasil klasifikasi kerusakan tulang belakang menggunakan berbagai algoritma klasifikasi. Perbandingan hasil penelitian ini dengan hasil penelitian lain dapat dilihat pada Tabel 4.

Tabel 4.

Perbandingan Hasil Penelitian Sebelumnya

No.	Peneliti	Metode	Akurasi
1	(Abdrabou, 2012)	Hybrid CBR dan ANN	85%
2	(Reddy & Kodali, 2012)	Naïve Bayes	83,74%
3	(Unal & Kocer, 2013)	Multi Layer Perceptron	85,48%
		Naïve Bayes	83,22%
4	(Da Rocha Neto et al., 2011)	SVM (Linear)	84,30%
		SVM (KMOD)	85,90%
5	Penelitian ini	<i>k-nearest neighbours</i> dan Algoritma Genetika	93%

4. Kesimpulan

Algoritma *k-nearest neighbours* yang sederhana dan kinerjanya yang tinggi terhadap berbagai aplikasi, karena kelebihan *k-nearest neighbours* dianggap

Tabel 3. Perbandingan Hasil Algoritma Genetika dengan Algoritma Seleksi Fitur Lain

Rasio Split Validation	Forward Selection	Backward Elimination	Greedy Forward Selection	Greedy Backward Selection	Genetic Algorithm
0,9	96,77%	96,77%	93,55%	96,77%	100%
0,8	90,32%	93,55%	90,32%	93,55%	93,55%
0,7	89,25%	88,17%	87,10%	88,17%	91,40%
0,6	88,71%	87,10%	86,29%	87,10%	90,32%
0,5	85,16%	85,81%	86,45%	85,81%	88,39%

sebanding dengan algoritma yang jauh lebih kompleks seperti *neural network* dan *support vector machine* (Wu et al., 2008).

Algoritma genetika yang dapat menangani berbagai macam optimasi tergantung pada fungsi objektifnya (fitness) apakah seimbang atau tidak seimbang, linier atau tidak linier, berkesinambungan atau tak berkesinambungan, atau dengan *random noise* (Gorunescu, 2011).

Algoritma genetika yang diterapkan untuk seleksi fitur dan optimasi parameter k pada algoritma *k-nearest neighbours* untuk meningkatkan akurasi klasifikasi kerusakan tulang belakang. Dalam lima eksperimen yang dilakukan menggunakan *split validation* dengan rasio yang bervariasi, metode yang diusulkan terbukti efektif untuk meningkatkan akurasi secara signifikan.

Perbandingan algoritma fitur seleksi diusulkan untuk membandingkan hasil akurasi dari algoritma fitur seleksi lain seperti *forward selection*, *backward elimination*, *greedy forward selection* dan *greedy backward selection*. Algoritma genetika terbukti memiliki performa yang lebih baik diantara algoritma seleksi fitur lainnya.

Pada penelitian ini, secara umum algoritma genetika diterapkan untuk seleksi fitur dan optimasi parameter algoritma *k-nearest neighbours* untuk meningkatkan performa klasifikasi kerusakan tulang belakang. Beberapa hal yang dapat dilakukan untuk menyempurnakan penelitian ini diantaranya, menggunakan algoritma optimasi parameter lain seperti *particle swarm optimization* terhadap algoritma klasifikasi lain seperti *neural network*, *support vector machine* atau *naïve bayes*.

Referensi

- Abdrabou, E. (2012). A Hybrid Intelligent Classifier for The Diagnosis of Pathology on the Vertebral Column. *Bioinformatics Using Intelligent and Machine Learning*, 297–310.
- Abe, S. (2005a). Modified Backward Feature Selection by Cross Validation. In *European Symposium on Artificial Neural Networks* (pp. 163–168). Bruges: d-side publishers.
- Abe, S. (2005b). *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*. Berlin, Heidelberg: Springer-Verlag.
- Ansari, S., & Sajjad, F. (2013). Diagnosis of Vertebral Column Disorders Using Machine Learning Classifiers. *2013 International Conference on Information Science and Applications (ICISA), Suwon*, 1–6. <https://doi.org/10.1109/ICISA.2013.6579446>
- Bharti, K. K., & Singh, P. K. (2014). A three-stage unsupervised dimension reduction method for text clustering. *Journal of Computational Science*, 5(2), 156–169. <https://doi.org/10.1016/j.jocs.2013.11.007>
- Blanchet, G., Legendre, P., & Borcard, D. (2008). Forward selection of spatial explanatory variables. *Ecology*, 89(9), 2623–2632. <https://doi.org/10.1890/07-0986.1>
- Brant-Zawadzki, M. N., Dennis, S. C., Gade, G. F., & Weinstein, M. P. (2000). Low back pain. *Radiology*, 217(2), 321–330. <https://doi.org/10.1102/1470-7330.2005.0028>
- Da Rocha Neto, A. R., Sousa, R., De A. Barreto, G., & Cardoso, J. S. (2011). Diagnostic of pathology on the vertebral column with embedded reject option. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6669 LNCS, 588–595. https://doi.org/10.1007/978-3-642-21257-4_73
- Deepa, T. (2013). An Innovative Optimization Algorithm for Feature Selection – A Comparative study, 3(1), 20–24.
- Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2), 265–282. <https://doi.org/10.1111/j.2044->

- 8317.1992.tb00992.x
- Dyer, E. L., Sankaranarayanan, A. C., & Baraniuk, R. G. (2013). Greedy Feature Selection for Subspace Clustering. *The Journal of Machine Learning Research*, 14(1). <https://doi.org/10.1109/TDEI.2009.5211872>
- Eyo, R. A. D., & Einstein, J. N. W. (2001). Low back pain. *The New England Journal of Medicine*, 344(5), 363–370. <https://doi.org/10.1016/j.ajo.2011.04.011>
- Farahat, A. K., Ghodsi, A., & Kamel, M. S. (2013). Efficient greedy feature selection for unsupervised learning. *Knowledge and Information Systems*, 35(2), 285–310. <https://doi.org/10.1007/s10115-012-0538-1>
- Gorunescu, F. (2011). Intelligent systems reference library. *Gorunescu, Ed.*
- Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research (JMLR)*, 3(3), 1157–1182. <https://doi.org/10.1016/j.aca.2011.07.027>
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Harrington, P. (2012). Machine learning in action. *Shelter Island, NY: Manning Publications Co.*
- Jain, A., & Zongker, D. (1997). Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2), 153–158. <https://doi.org/10.1109/34.574797>
- Jirapech-Umpai, T., & Aitken, S. (2005). Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes. *BMC Bioinformatics*, 6, 1–11. <https://doi.org/10.1186/1471-2105-6-148>
- Kataria, A., & Singh, M. D. (2013). A Review of Data Classification Using K-Nearest Neighbour Algorithm. *International Journal of Emerging Technology and Advanced Engineering*, 3(6), 354–360.
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.
- Liu, Z., Chai, T., Yu, W., & Tang, J. (2015). Multi-frequency signal modeling using empirical mode decomposition and PCA with application to mill load estimation. *Neurocomputing*, 169, 392–402. <https://doi.org/10.1016/j.neucom.2014.08.087>
- Maimon, O., & Rokach, L. (2005). *Data Mining and Knowledge Discovery Handbook*. Berlin, Heidelberg: Springer-Verlag.
- Prasetio, R. T., & Pratiwi, P. (2015). PENERAPAN TEKNIK BAGGING PADA ALGORITMA KLASIFIKASI UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS DATASET MEDIS. *Jurnal Informatika*, 2(2), 395–403.
- Prasetio, R. T., & Riana, D. (2015). A comparison of classification methods in vertebral column disorder with the application of genetic algorithm and bagging. In *2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)* (pp. 163–168). Bandung. <https://doi.org/10.1109/ICICI-BME.2015.7401356>
- Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A., & Jain, A. K. (2000). Dimensionality Reduction Using Genetic Algorithms. *IEEE Transactions on Control Systems Technology*, 4(2), 164–171. <https://doi.org/10.1109/TCST.2011.2171964>
- Reddy, S. K., & Kodali, S. R. (2012). Classification of Vertebral Column

- using Naïve Bayes Technique, 58(7), 38–42.
- Setiyorini, T., & Wahono, R. S. (2014). Penerapan metode bagging untuk mengurangi data noise pada neural network untuk estimasi tuat tekan beton. *Journal of Intelligent Systems*, 1(1), 36–41.
- Shilaskar, S., & Ghatol, A. (2013). Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. *Expert Systems with Applications*, 40(10), 4146–4153. <https://doi.org/10.1016/j.eswa.2013.01.032>
- Tiwari, D. (2014). Handling Class Imbalance Problem Using Feature Selection. *International Journal of Advanced Research in Computer Science & Technology*, 2(2), 516–520.
- Unal, Y., & Kocer, H. E. (2013). Diagnosis of pathology on the vertebral column with backpropagation and Naïve Bayes classifier. *Technological Advances in Electrical, Electronics and Computer Engineering (TAECE)*, 2013 International Conference On, 276–279. <https://doi.org/10.1109/TAECE.2013.6557285>
- Vafaie, H., & Imam, I. F. (1994). Feature Selection Methods: Genetic Algorithms vs Greedy-like Search. *Proceedings of the International Conference on Fuzzy and Intelligent Control Systems*, 1(March). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.160.9710&rep=rep1&type=pdf>
- Videman, T., Nummi, P., Battie, M. C., & Gill, K. (1994). Digital assessment of MRI for lumbar disc desiccation. A comparison of digital versus subjective assessments and digital intensity profiles versus discogram and macroanatomic findings. *Spine*, 19(2), 192–198.
- Villar, P., Fernández, A., & Herrera, F. (2010). A genetic algorithm for feature selection and granularity learning in fuzzy rule-based classification systems for highly imbalanced datasets. *Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-642-14055-6_78
- White, A. A. I., & Gordon, S. L. (1982). Synopsis: Workshop on Idiopathic Low-Back Pain. *Spine*, 7(2). Retrieved from https://journals.lww.com/spinejournal/Fulltext/1982/03000/Synopsis__Workshop_on_Idiopathic_Low_Back_Pain.9.aspx
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., ... Steinberg, D. (2008). *Top 10 algorithms in data mining. Knowledge and Information Systems* (Vol. 14). <https://doi.org/10.1007/s10115-007-0114-2>.