

Feature Selection Menggunakan Binary Wheel Optimizaton Algorithm (BWOA) pada Klasifikasi Penyakit Diabetes

Lastri Widya Astuti¹⁾, Imelda Saluza²⁾, Evi Yulianti³⁾, Dhamayanti⁴⁾

¹⁾ Program Studi Teknik Informatika, Universitas Indo Global Mandiri

²⁾ Program Studi Manajemen Informatika, Universitas Indo Global Mandiri

^{3,4)} Program Studi Sistem Informasi, Universitas Indo Global Mandiri

Jl. Jend Sudirman No 629 Km 4 Palembang

Email : lastriwidya@uigm.ac.id¹⁾, imeldasaluza@uigm.ac.id²⁾, eviyulianti@uigm.ac.id³⁾, dhamayanti@uigm.ac.id⁴⁾

ABSTRACT

Diabetes Mellitus (DM) is a chronic disease characterized by blood glucose (blood sugar) levels exceeding normal, i.e. blood sugar levels being equal to or more than 200 mg/dl, and fasting blood sugar levels being above or equal to 126 mg/dl. The increase in the number of people with diabetes is due to delays in detection. Utilization of machine learning in helping to establish a fast and accurate diagnosis is one of the efforts made in the health sector. One of the important steps to produce high classification accuracy is through the selection of relevant features. The problem in feature selection is dimensionality reduction, where initially all attributes are required to obtain maximum accuracy while not all features are used in the classification process. This study uses the Binary wheel Optimization Algorithm (BWOA) as a feature selection method to increase accuracy in the classification of diabetes mellitus. The use of metaheuristic algorithms is an alternative to increase computational efficiency and avoid local minimums. The BWOA algorithm reduces the 8 attributes in the dataset to the 3 best attributes that are able to represent the original dataset. The results showed that from the six classification methods tested, namely: K-NN, Naïve Bayes, Random Forest, Logistics Regression, Decision Tree, Neural Network. then the three logistic regression methods, naive Bayes and neural network are in good classification criteria based on Area Under Curve (AUC) while the calculation of the accuracy value shows an average of above 70%.

Keywords : Feature Selection, Classification, Diabetes Mellitus, Accuracy, Area Under Curve (AUC)

ABSTRAK

Diabetes Melitus (DM) merupakan suatu penyakit menahun yang ditandai dengan kadar glukosa darah (gula darah) melebihi normal yaitu kadar gula darah sewaktu sama atau lebih dari 200 mg/dl, dan kadar gula darah puasa di atas atau sama dengan 126 mg/dl. Peningkatan jumlah penderita diabetes disebabkan keterlambatan dalam mendeteksi. Pemanfaatan machine learning dalam membantu penegakkan diagnosis yang cepat akurat merupakan salah satu upaya yang dilakukan pada bidang kesehatan. Salah satu langkah penting untuk menghasilkan akurasi klasifikasi yang tinggi, melalui pemilihan fitur relevan. Masalah dalam seleksi fitur adalah pengurangan dimensi, dimana awalnya semua atribut diperlukan untuk memperoleh akurasi yang maksimal sementara tidak semua fitur digunakan dalam proses klasifikasi. Penelitian ini menggunakan Binary wheel Optimization Algorithm (BWOA) sebagai metode seleksi fitur untuk meningkatkan akurasi dalam klasifikasi penyakit diabetes mellitus. Penggunaan algoritma metaheuristic menjadi alternatif untuk meningkatkan efisiensi komputasi dan menghindari lokal minimum. Algoritma BWOA mereduksi dari 8 atribut yang ada pada dataset menjadi 3 atribut terbaik yang mampu mewakili dataset asli. Hasil penelitian menunjukkan dari enam metode klasifikasi yang diujicoba, yaitu: K-NN, Naïve Bayes, Random Forest, Logistik Regresion, Decision Tree, Neural Network. maka tiga metode logistic regression, naive bayes dan neural network berada pada kriteria good classification berdasarkan Area Under Curve (AUC) sedangkan perhitungan nilai akurasi menunjukkan rata-rata diatas 70%..

Kata Kunci : Feature Selection, Classification, Diabetes Mellitus, Accuracy, Area Under Curve (AUC)



Article History

Received : 10/01/2022

Revised : 17/01/2022

Accepted : 02/03/2022

Online : 30/03/2022



This is an open access article under the
CC BY-SA 4.0 License

1. Pendahuluan

Penyakit tidak menular (PTM) merupakan penyakit yang tidak disebabkan oleh infeksi mikroorganisme seperti protozoa, bakteri, jamur, maupun virus. Penyakit jenis ini bertanggungjawab terhadap sedikitnya 70% kematian di dunia. Meskipun tidak dapat ditularkan dari orang ke orang maupun dari binatang ke orang, lemahnya pengendalian faktor risiko dapat berpengaruh terhadap peningkatan kasus setiap tahun. Hal ini sejalan dengan hasil Riset Kesehatan Dasar (Riskesdas) tahun 2007, 2013, dan 2018 yang menunjukkan kecenderungan peningkatan prevalensi PTM seperti diabetes, hipertensi, stroke, dan penyakit sendi/rematik/encok (profil kemenkes 2019). Diperkirakan pada tahun 2030 prevalensi DM di Indonesia akan mencapai 21,3 juta orang.

Diabetes Melitus (DM) merupakan suatu penyakit menahun yang ditandai dengan kadar glukosa darah (gula darah) melebihi normal yaitu kadar gula darah sewaktu sama atau lebih dari 200 mg/dl, dan kadar gula darah puasa di atas atau sama dengan 126 mg/dl (Azrimaidaliza, 2011). Peningkatan kadar gula darah ini akan memicu produksi hormon insulin oleh kelenjar pankreas. Keadaan hiperglikemia kronis dari diabetes berhubungan dengan kerusakan jangka panjang, gangguan fungsi dan kegagalan berbagai organ (Yusnanda, Rochadi and Maas, 2019). Klasifikasi etiologi kelainan glikemia atau Diabetes Mellitus sebagai berikut : *Tipe 1*, ditandai dengan kegagalan produksi insulin yang parsial atau total oleh sel-sel B pankreas. Faktor penyebab masih belum dimengerti dengan jelas tetapi beberapa virus tertentu, penyakit autoimun dan faktor-faktor genetik mungkin turut berperan; *Tipe 2*, ditandai dengan resistensi insulin ketika hormon insulin diproduksi dengan jumlah yang tidak memadai atau dengan bentuk yang tidak efektif. Ada korelasi genetik yang kuat pada tipe diabetes ini dan proses terjadinya berkaitan erat dengan obesitas. Anak dengan diabetes tipe 2 dilaporkan memiliki riwayat penyakit kardiovaskular dalam keluarga dan atau sindrom metabolik; *Tipe 3* merupakan Tipe spesifik lainnya dari penyakit diabetes melitus, berupa defek genetik pada fungsi sel - B, defek genetik pada kerja insulin, penyakit pada kelenjar eksokrin pankreas, endokrinopati, ditimbulkan oleh obat-obatan atau zat kimia, infeksi, bentuk immune-mediated diabetes yang langka. Kadang-kadang sindrom genetik lain yang disertai diabetes; *Tipe 4* Diabetes gestasional: bentuk diabetes yang terjadi selama kehamilan. Kebanyakan, tapi tidak semuanya, akan sembuh setelah melahirkan (Istianah, Septiani and Dewi, 2020).

Peningkatan jumlah penderita diabetes disebabkan keterlambatan dalam mendeteksi penyakit DM. berdasarkan *the National Diabetes Data Group of the USA* (NDDG) dan komite pakar pada WHO menghasilkan keseragaman hingga taraf tertentu dalam penetapan diagnose awal, salah satu diantaranya Tes Toleransi Glukosa Oral (TTGO) dengan 75 gram glukosa digunakan untuk membedakan antara DM dan

bukan DM (Fitriyanti, Febriawati and Yanti, 2019). Pemanfaatan *machine learning* dalam membantu penegakkan diagnosis yang cepat akurat merupakan salah satu upaya yang dilakukan pada bidang kesehatan. Beberapa penelitian terkait penyelesaian atau deteksi penyakit diabetes mellitus melalui metode klasifikasi, diantaranya K-NN(Yunita, 2016)(Argina, 2020), *Naive Bayes* (Ridwan, 2020), C45 (Moh. Jasri, 2017)(Ente *et al.*, 2020), *Decision Tree* (Najib, Nurcahyono and Setiawan, 2019), *SVM*(Mucholladin, Abdurrachman Bachtiar and Furqon, 2021) dan *Neural Network* (Mardiana, Ditama and Tuslaela, 2020)(Nugraha, Shidiq and Rahayu, 2019). Hasil klasifikasi pada penelitian terdahulu menunjukkan variansi keragaman akurasi yang dihasilkan. Salah satu langkah penting untuk menghasilkan akurasi klasifikasi yang tinggi, melalui pemilihan fitur relevan yang digunakan dalam penelitian. Masalah dalam seleksi fitur adalah pengurangan dimensi, dimana awalnya semua atribut diperlukan untuk memperoleh akurasi yang maksimal sementara tidak semua fitur digunakan dalam proses klasifikasi(Astuti *et al.*, 2020). *Feature selection* pada penerapannya dibagi menjadi dua kelompok *ranking selection* dan *subset selection*. *Ranking selection* melakukan perankingan antara satu *feature* dengan *feature* lainnya, dimana *feature* dengan ranking terbaik akan digunakan dalam proses klasifikasi. Beberapa metode yang digunakan dalam *feature ranking selection*, diantaranya: *regression*, *correlation* dan *mutual information*. *Subset selection* merupakan metode seleksi fitur dengan mencari *feature* yang optimal. *Subset selection* dibagi menjadi tiga metode, yaitu: *wrapper selection*, *filter selection* dan *embedded selection*

Penggunaan metode seleksi fitur diharapkan dapat meningkatkan akurasi dalam penelitian terkait penyakit diabetes mellitus. Penelitian ini menggunakan *Binary wheal Optimization Algorithm* (BWOA) sebagai metode seleksi fitur yang diterapkan pada beberapa metoda algoritma klasifikasi(Sathya and Manju Priya, 2020)(Sayed, Darwish and Hassanien, 2020)(Hussien *et al.*, 2020).

2. Pembahasan

Seleksi fitur memiliki dua tujuan utama, yaitu memaksimalkan akurasi klasifikasi dan meminimalkan jumlah fitur. BWOA digunakan untuk secara adaptif mencari kombinasi fitur terbaik, dengan mempertimbangkan dua tujuan tersebut. Seleksi fitur pada metode BWOA dititik beratkan pada penggunaan biner, di mana ruang pencarian dimodelkan sebagai kisi - kisi boolean n-dimensi, pemilihan fitur yang sesuai atau digunakan ditandai dengan binary "1" sementara binary "0" menunjukkan fitur yang tidak sesuai atau yang tidak digunakan(Eid, 2018). Algoritma BWOA dideskripsikan pada potongan pseudocode algoritma dibawah ini.

Algoritma : Binary Whale Optimization Algorithm

Input:

Number Of Whale n

Number Of Iteration for Optimization Max_Iter

Output:

Optimal Whale Binary Position X^*

Initialize the n whales population positions at random $\square [0,1]$.

Initialize a, A and C .

Calculate the fitness of each search agent.

X^* = the best search agent.

$t=1$

while $t \leq Max_Iter$ **do**

for each search agent **do**

 Update the position of the current search agent to a binary position by equation 6

end for

 Update a, A and C

 Evaluate the fitness of each search agent.

 Update X^* if there is a better solution

$t=t+1$

end while

return X^*

Fungsi fitness yang diadopsi dalam BWOA untuk mengevaluasi setiap posisi individu paus dideskripsikan yaitu:

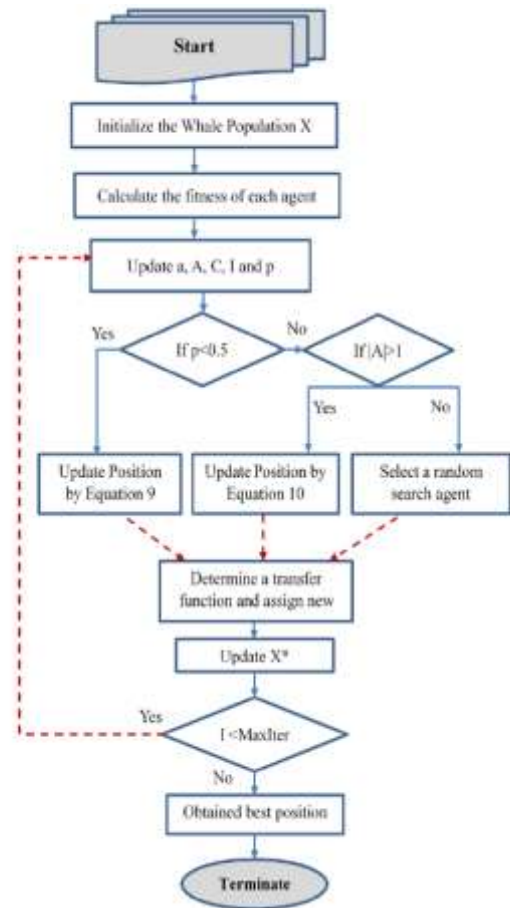
$$Fitness = \alpha E_R + (1 - \alpha) \frac{|S^*|}{|S|} \dots\dots\dots (1)$$

dimana E_R adalah tingkat kesalahan klasifikasi fitur yang dipilih, S^* adalah jumlah fitur yang dipilih dan S adalah jumlah total fitur. α dan $(1-\alpha)$ mewakili kepentingan relatif dari akurasi klasifikasi dan jumlah fitur yang dipilih, $\alpha \square (0; 1]$ (Tawhid and Ibrahim, 2020).

Flowchart dari Algorithma BWOA ditampilkan pada Gambar 1.

Dataset yang digunakan dalam penelitian ini bersumber dari Pima Indians Diabetes Database (www.kaggle.com/uciml/pima-indians-diabetes-database/diabetes.csv). Dataset terdiri dari delapan atribut yang digunakan untuk mendeskripsikan gejala yang digunakan sebagai penanda untuk penegakkan dignosa penyakit diabetes. Gejala klasik dari penderita diabetes meliputi polifagi, poliuri, dan polidipsi atau melalui pemeriksaan klinis laboratorium menunjukkan hasil pemeriksaan glukosa darah sewaktu ≥ 200 mg/dl (11,1 mmol/l) dan FPG (kadar glukosa puasa) ≥ 126 mg/dl (7,0 mmol/l) atau glukosa plasma 2jam setelah makan (2jam pp) > 200 mg/dl (11,1 mmol/l). Pada table 1 menjelaskan atribut yang digunakan pada dataset pima Indians diabetes

Pemilihan fitur merupakan masalah kombinatorial dengan ruang pencarian yang besar. Ukuran ruang pencarian tumbuh secara eksponensial seiring dengan jumlah total fitur, sehingga pencarian optimal menyeluruh untuk subset fitur dalam ruang dimensi tinggi tidak praktis.



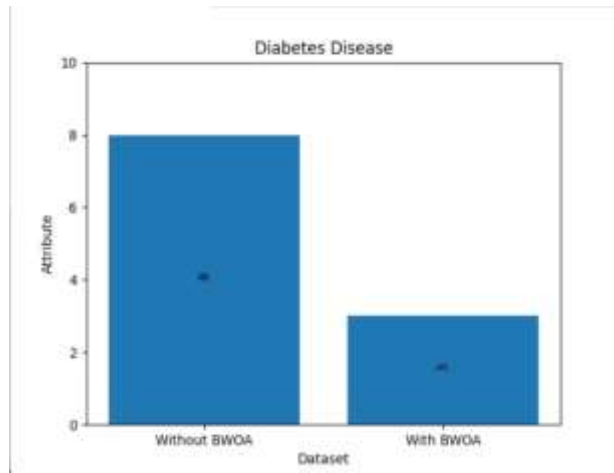
Gambar 1. Flowchart Algorithma BWOA

Pemilihan fitur dapat diformulasikan sebagai masalah optimasi kombinatorial dengan memilih subset fitur terbaik yang digunakan untuk mewakili dataset asli (Eid, 2018) (Kahya, Altamir and Algamal, 2021).

Tabel 1. Tabel Atribut Penyakit Diabetes Mellitus

| No | Kolom |
|----|--|
| 1 | Number or Times Pregnant (NTP) |
| 2 | Plasma Glucose Concentration (PGC) |
| 3 | Diastolic Blood Pressure (mmHg)(DBP) |
| 4 | Triceps Skin – Fold Thickness (mm)(TSFT) |
| 5 | 2-h Serum Insulin (mm U/ml)(H2SI) |
| 6 | Body Mass Index (Kg/m2) (BMI) |
| 7 | Diabetes Pedigree Function (DPf) |
| 8 | Age |

Penggunaan algorithma metaheuristic menjadi alternatif untuk meningkatkan efisiensi komputasi dan menghindari lokal minimum. Algoritma BWOA mereduksi dari 8 atribut yang ada pada dataset menjadi 3 atribut terbaik yang mampu mewakili dataset asli.



Gambar 2. Hasil Feature Selection BWOA

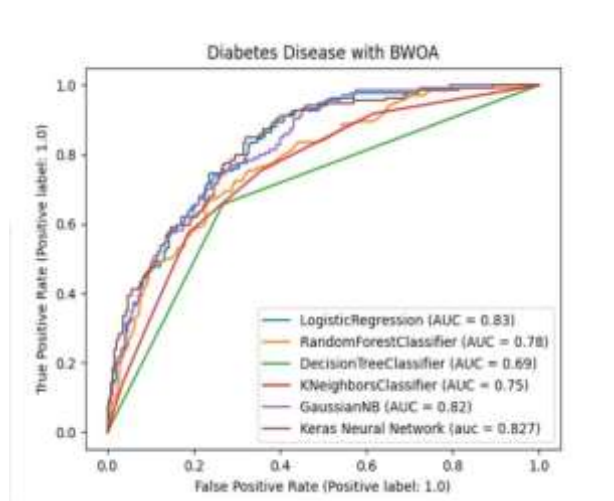
Atribut terpilih digunakan dalam proses klasifikasi dengan menggunakan beberapa metode klasifikasi. Pada penelitian metode klasifikasi yang digunakan yaitu: K-NN, Naïve Bayes, Random Forest, Logistik Regresion, Decision Tree, Neural Network. Hasil penambahan algoritma BWOA untuk seleksi fitur mampu meningkatkan persentase akurasi dan Area Under Curve (AUC). Nilai akurasi didapat melalui perhitungan menggunakan confusion matrix sementara nilai AUC didapat dari Receiver Operating Characteristic Curve (ROC)(Suwarno, 2016). Perhitungan akurasi menggunakan confusion matrik (Astuti, 2019) didapat dari persamaan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(2)$$

Sementara persamaan untuk perhitungan AUC yaitu:

$$AUC = \frac{1}{2} \sum_{i=1}^n (X_{i+1} - X_i) (Y_{i+1} - Y_i) \dots\dots(3)$$

Hasil Area Under Cover (AUC) penelitian ini, disajikan pada gambar 3



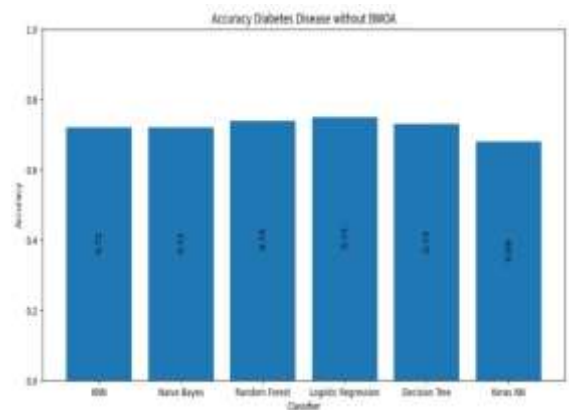
Gambar 3. Grafik AUC menggunakan BWOA

Berdasarkan tabel 2 kriteria AUC dibawah ini maka nilai AUC untuk Logistic Regression, Naïve Bayes Dan Neural Network berada pada range good classification, sementara Random Forest dan K-NN berada pada range fair classification sedangkan Decision Tree berada pada range poor classification

Tabel 2. Kriteria AUC

| Nilai AUC | Interprestasi |
|-------------|--------------------------|
| 0.90 – 1.00 | Excellent Classification |
| 0.80 – 0.90 | Good Classification |
| 0.70 – 0.80 | Fair Classification |
| 0.60 – 0.70 | Poor Classification |
| 0.50 – 0.60 | Failure |

Sedangkan hasil perhitungan akurasi menggunakan confusion matrik ditunjukkan pada gambar 4 dan gambar Hasil akurasi dari metode klasifikasi yang tidak menggunakan algoritma BWOA dalam pemilihan fitur menunjukkan untuk Metode K-NN didapat akurasi 72%, metode Naïve bayes memperoleh akurasi 72%, metode Random forest memperoleh akurasi sebesar 74%, metode Logistic regression hasil akurasi sebesar 75%, metode Decision tree mempeoleh hasil akurasi sebesar 73% dan terakhir metode Neural network memperoleh hasil akurasi sebesar 68%



Gambar 4. Hasil Akurasi Without BWOA

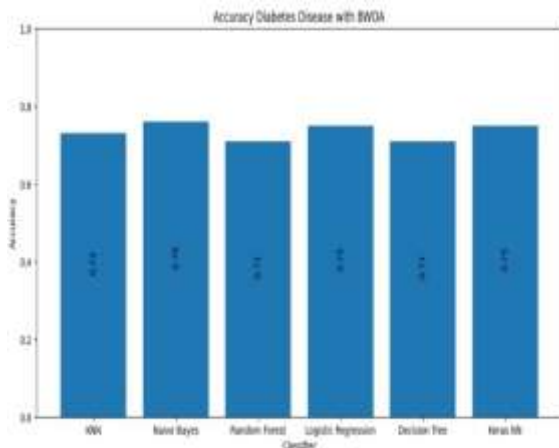
Pada percobaan perhitungan akurasi dengan menambahkan algoritma BWOA untuk pemilihan fitur maka didapatkan adanya peningkatan hasil akurasi pada beberapa metode klasifikasi. Hasil akurasi metode K-NN sebesar 73%, hasil akurasi metode naïve bayes sebesar 76%, hasil akurasi metode random forest sebesar 71%, hasil akurasi metode logistic regression sebesar 75%, hasil akurasi metode decision tree sebesar 71% dan hasil akurasi metode neurak network sebesar 75%.

Perbandingan antara penggunaan algoritma BWOA untuk seleksi fitur pada proses klasifikasi disajikan pada table:

Tabel 3. Perbandingan Akurasi

| Akurasi | Without BWOA | With BWOA |
|---------------------|--------------|-----------|
| K-NN | 72% | 73% |
| Naïve Bayes | 72% | 76% |
| Random Forest | 74% | 71% |
| Logistic Regression | 75% | 75% |
| Decission Tree | 73% | 71% |
| Neural Network | 68% | 75% |

Hasil akurasi yang diterapkan pada metoda klasifikasi menunjukkan adanya peningkatan nilai akurasi pada metode K-NN, Naïve Bayes dan Neural Network, sementara tiga metode lainnya menunjukkan hasil akurasi tetap sebelum dan sesudah menggunakan Algoritma BWOA. Pemilihan fitur juga tidak hanya berpengaruh pada peningkatan hasil akurasi, namun juga mempengaruhi waktu komputasi.



Gambar 5. Hasil Akurasi With BWOA

Peningkatan akurasi berada pada range 2% -7 %, dengan peningkatan akurasi terbaik pada *Neural Network*.

3. Kesimpulan

Hasil penelitian menunjukkan Algoritma BWOA mampu mereduksi attribute dari jumlah awal attribute sebanyak 8 hingga menjadi 3 melalui proses pemilihan fitur. Fitur terpilih digunakan pada proses klasifikasi dengan menggunakan metode yang telah ditetapkan dalam penelitian, hasil klasifikasi untuk *Logistic Regression, Naïve Bayes Dan Neural Network* berada pada range *good classification*, sementara *Random Forest* dan *K-NN* berada pada range *fair classification* sedangkan *Decision Tree* berada pada range *poor classification* berdasarkan *Area Under Curve* dan berdasarkan perhitungan akurasi berdasarkan metode *confusion matrik rata – rata nilai akurasi diatas 70%*.

Daftar Pustaka

- Argina, A. M. (2020) ‘Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes’, *Indonesian Journal of Data and Science*, 1(2), pp. 29–33. doi: 10.33096/ijodas.v1i2.11.
- Astuti, L. W. (2019) ‘Ekstrasi Fitur Citra MRI Otak Menggunakan Data Wavelet Transform (DWT) untuk Klasifikasi Penyakit Tumor Otak’, *Jurnal Ilmiah Informatika Global*, 10(2), pp. 80–86. doi: 10.36982/jig.v10i2.854.
- Astuti, L. W. et al. (2020) ‘Optimalisasi Klasifikasi Kanker Payudara Menggunakan Forward Selection pada Naive Bayes’, *Jurnal Informatika Global*, 11(2), pp. 63–67.
- Azrimaidaliza (2011) ‘Asupan Zat Gizi dan Penyakit Diabetes Mellitus’, *Jurnal Kesehatan Masyarakat*, 6(1), pp. 36–41.
- Eid, H. F. (2018) ‘Binary whale optimisation: an effective swarm algorithm for feature selection’, *International Journal of Metaheuristics*, 7(1), p. 67. doi: 10.1504/ijmheur.2018.10012912.
- Ente, D. R. et al. (2020) ‘Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus Di Rumah Sakit Unhas Menggunakan Algoritma C4.5’, *Indonesian Journal of Statistics and Its Applications*, 4(1), pp. 80–88. doi: 10.29244/ijsa.v4i1.330.
- Fitriyanti, M. E., Febriawati, H. and Yanti, L. (2019) ‘Pengalaman Penderita Diabetes Mellitus dalam Pencegahan Ulkus Diabetik’, *Jurnal Keperawatan Muhammadiyah Bengkulu*, 7(2), pp. 99–105. doi: 10.36085/jkmu.v7i2.481.
- Hussien, A. G. et al. (2020) ‘Binary whale optimization algorithm for dimensionality reduction’, *Mathematics*, 8(10), pp. 1–24. doi: 10.3390/math8101821.
- Istianah, I., Septiani and Dewi, G. K. (2020) ‘Mengidentifikasi Faktor Gizi pada Pasien Diabetes Mellitus Tipe 2 di Kota Depok Tahun 2019’, *Jurnal Kesehatan Indonesia (The Indonesian Journal of Health)*, X(2), pp. 72–78.
- Kahya, M. A., Altamir, S. A. and Algamal, Z. Y. (2021) ‘Improving whale optimization algorithm for feature selection with a time-varying transfer function’, *Numerical Algebra, Control and Optimization*, 11(1), pp. 87–98. doi: 10.3934/naco.2020017.
- Mardiana, T., Ditama, E. M. and Tuslaela, T. (2020) ‘an Expert System for Detection of Diabetes Mellitus With Forward Chaining Method’, *Jurnal Riset Informatika*, 2(2), pp. 69–76. doi: 10.34288/jri.v2i2.121.
- Suwarno, (2016) ‘Jurnal MIPA’, 39(2), pp. 98–106.
- Moh. Jasri (2017) ‘Klasifikasi Penyakit Diabetes Mellitus Tipe 2 Dengan Metode Algoritma C4.5’, 5.
- Mucholladin, A. W., Abdurrachman Bachtiar, F. and Furqon, M. T. (2021) ‘Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(2), pp. 622–633. Available at: <http://j-ptiik.ub.ac.id>.

- Najib, A., Nurcahyono, D. and Setiawan, R. P. P. (2019) 'Klasifikasi Diagnosa Penyakit Diabetes Mellitus (Dm) Menggunakan Algoritma C4.4', *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11(2), p. 47. doi: 10.46964/justti.v11i2.153.
- Nugraha, F. S., Shidiq, M. J. and Rahayu, S. (2019) 'Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara', *Jurnal Pilar Nusa Mandiri*, 15(2), pp. 149–156. doi: 10.33480/pilar.v15i2.601.
- Ridwan, A. (2020) 'Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus', *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 4(1), pp. 15–21. doi: 10.47970/siskom-kb.v4i1.169.
- Sathya, M. and Manju Priya, S. (2020) 'Modified whale optimization algorithm for feature selection in micro array cancer dataset', *International Journal of Scientific and Technology Research*, 9(3), pp. 549–556.
- Sayed, G. I., Darwish, A. and Hassanien, A. E. (2020) 'Binary Whale Optimization Algorithm and Binary Moth Flame Optimization with Clustering Algorithms for Clinical Breast Cancer Diagnoses', *Journal of Classification*, 37(1), pp. 66–96. doi: 10.1007/s00357-018-9297-3.
- Tawhid, M. A. and Ibrahim, A. M. (2020) 'Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm', *International Journal of Machine Learning and Cybernetics*, 11(3), pp. 573–602. doi: 10.1007/s13042-019-00996-5.
- Yunita, F. (2016) 'Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-NN)', *Bappeda*, 2, pp. 223–230.
- Yusnanda, F., Rochadi, R. K. and Maas, L. T. (2019) 'Pengaruh Riwayat Keturunan terhadap Kejadian Diabetes Mellitus pada Pra Lansia di BLUD RSUD Meuraxa Kota Banda Aceh Tahun 2017', *Journal of Healthcare Technology and Medicine*, 4(1), p. 18. doi: 10.33143/jhtm.v4i1.163.