JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8747
Hal 414-421
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Performansi Model Machine Learning dalam Klasifikasi Penyakit Diabetes Tipe 2

Ryan Hidayatulloh*, Wahyu Aji Eko Prabowo[2]

Fakultas Ilmu Komputer, Program Studi PJJ Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]118202200002@mhs.dinus.ac.id, [2]prabowo@dsn.dinus.ac.id
Email Penulis Korespondensi: 118202200002@mhs.dinus.ac.id
Submitted **15-06-2025;** Accepted **10-07-2025;** Published **14-08-2025**

Abstrak

Diabetes Mellitus tipe 2 merupakan penyakit kronis yang berkembang secara bertahap dan dapat menyebabkan komplikasi serius seperti penyakit jantung, gagal ginjal, dan kebutaan apabila jika tidak terdeteksi sejak dini. Tujuan penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja empat algoritma pembelajaran mesin, yaitu *Logistic Regression, Random Forest, Multilayer Perceptron,* dan *Deep Neural Network*, dalam memprediksi risiko diabetes tipe 2 berdasarkan data medis. *Dataset* yang digunakan adalah *Pima Indians Diabetes* yang terdiri dari 9538 data pasien dan 16 fitur prediktor. Data dibagi menjadi data latih dan data uji dengan perbandingan 80 banding 20. Pelatihan mencakup *hyperparameter tunning* dengan menggunakan *Grid Search* dan validasi silang. *Accuracy, precision, recall, F1-score, Mean Squared Error, Root Mean Squared Error, R-Squared,* dan analisis *overfitting* digunakan untuk mengevaluasi performa model. Hasil penelitian menunjukkan bahwa model *Random Forest* memiliki performa terbaik dengan akurasi 100%, tidak ada kesalahan klasifikasi, nilai kesalahan prediksi mendekati nol, dan tidak ada gejala *overfitting*. *Logistic Regression* juga menunjukkan performa yang sangat baik, meskipun sedikit dibawah *Random Forest*. Sedangkan *Multilayer Perceptron* dan *Deep Neural Network* cenderung mengalami *overfitting* kecil dan tingkat *False Negative* yang lebih tinggi. Berdasarkan hasil tersebut, Model *Random Forest* direkomendasikan sebagai model yang paling andal untuk sistem prediksi dini *Diabetes Mellitus* tipe 2.

Kata Kunci: Diabetes Mellitus; Pembelajaran Mesin; Random Forest; Prediksi Penyakit; Evaluasi Model

Abstract

Type 2 Diabetes Mellitus is a chronic disease that develops gradually and can lead to serious complications—such as heart disease, kidney failure, and blindness—if not detected early. This study aims to evaluate and compare the performance of four machine learning algorithms—Logistic Regression, Random Forest, Multilayer Perceptron, and Deep Neural Network—in predicting the risk of type 2 diabetes based on medical data. The analysis uses the Pima Indians Diabetes dataset, which contains 9.538 patient records and 16 predictor variables. We split the data into training and testing sets using an 80:20 ratio. During training, we performed hyperparameter tuning using Grid Search combined with cross-validation. To evaluate model performance, we applied several metrics, including accuracy, precision, recall, F1-score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2, and an analysis of overfitting. The results indicate that the Random Forest model outperformed the others, achieving 100% accuracy, zero classification errors, near-zero prediction error values, and no signs of overfitting. Logistic Regression also performed well, though slightly below the Random Forest. In contrast, the Multilayer Perceptron and Deep Neural Network models showed mild overfitting and higher false negative rates. Based on these findings, we recommend the Random Forest model as the most reliable option for early prediction systems in type 2 diabetes mellitus.

Keywords: Diabetes Mellitus; Machine Learning; Random Forest; Disease Prediction; Model Evaluation

1. PENDAHULUAN

Diabetes mellitus tipe 2 (DMT2) adalah salah satu penyakit metabolik yang berkembang pesat di seluruh dunia. Resistensi insulin menyebabkan tubuh tidak dapat menggunakan insulin dengan baik, yang menyebabkan kadar glukosa darah tinggi. DMT2 dapat menyebabkan berbagai komplikasi jangka panjang, seperti penyakit jantung, gangguan ginjal, kebutaan, dan amputasi, yang akan menurunkan kualitas hidup pasien. Di Indonesia, DMT2 sangat umum, dengan lebih dari 10 juta orang yang didiagnosis pada tahun 2020. Meskipun penyakit ini dapat disembuhkan dengan pengobatan dan perubahan gaya hidup, mendeteksi orang yang berisiko sangat penting untuk mencegah komplikasi [1]. Karena tidak menunjukkan gejala yang jelas di awal, masih banyak orang yang tidak terdiagnosis sampai akhirnya mengalami stadium lanjut. Oleh karena itu, pengembangan teknik yang dapat memprediksi risiko DMT2 dengan lebih akurat dan efektif sangat penting.

Implementasi *screening* klinis tradisional seringkali sulit karena waktu dan biaya yang tinggi. Hal ini menunjukkan bahwa metode alternatif berbasis teknologi, terutama pembelajaran mesin, harus digunakan. Pembelajaran mesin dapat melakukan prediksi risiko DMT2 dengan lebih cepat dan akurat menggunakan data medis yang tersedia. Algoritma pembelajaran mesin memiliki kemampuan untuk mengolah data medis yang kompleks dan menemukan pola tersembunyi yang tidak dapat dikenali oleh teknik konvensional. *Random Forest* (RF) adalah salah satu algoritma yang paling banyak digunakan dalam klasifikasi penyakit karena bekerja dengan membentuk sejumlah pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi dan mengurangi kemungkinan overfitting. *Logistic Regression* (LR), Deep Neural Network (DNN), dan *Multilayer Perceptron* (MLP) adalah beberapa algoritma lain yang biasa digunakan untuk memprediksi DMT2 [2][3][4]. Andini dkk. (2022) melakukan penelitian yang menunjukkan bahwa algoritma Random Forest lebih baik dalam mengklasifikasikan pasien diabetes dengan akurasi 88,5% [5]. Penelitian ini juga menekankan betapa pentingnya memilih fitur penting seperti kadar glukosa, tekanan darah, dan indeks massa tubuh untuk meningkatkan akurasi model. Jannah et al. (2022) membandingkan algoritma DNN dan MLP untuk klasifikasi diabetes dan menemukan bahwa meskipun DNN lebih akurat, ia juga lebih sering mengalami *overfitting* dan

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8747
Hal 414-421
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

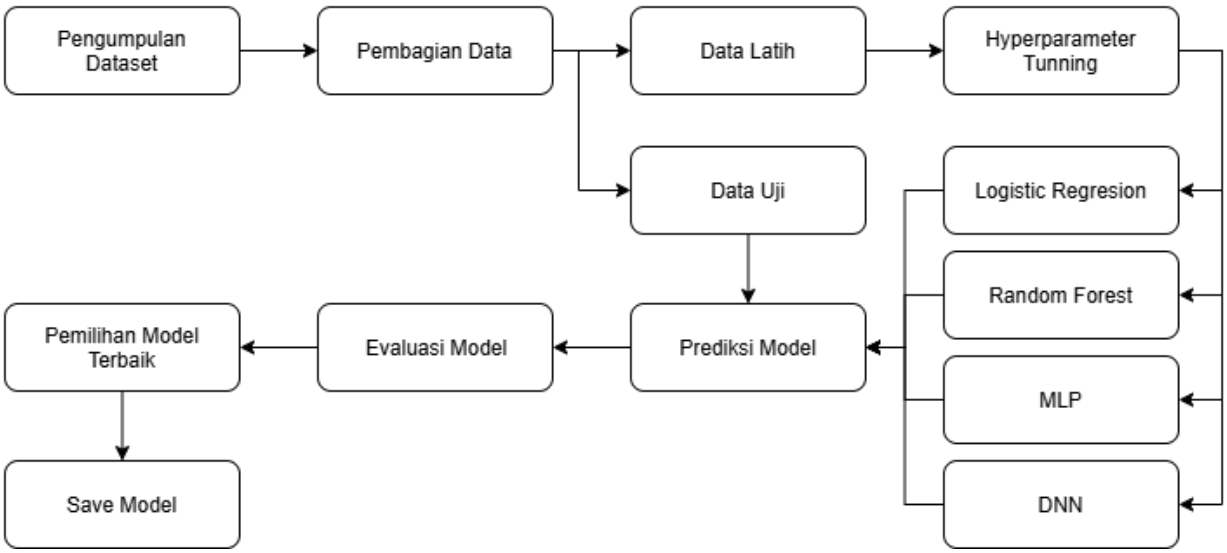
membutuhkan waktu pelatihan yang lebih lama [6]. Studi Saputri dkk. (2022) menemukan bahwa *Logistic Regression* dapat digunakan sebagai dasar model yang sederhana namun efektif untuk klasifikasi biner, meskipun tidak seakurat *Random Forest* dan DNN [6]. Di sisi lain, Nugroho dkk. (2021) mengevaluasi MLP dan menemukan bahwa meskipun stabil, mereka kurang presisi dan *recall* daripada RF [6].

Sebagian besar penelitian sebelumnya hanya membandingkan dua atau tiga algoritma dan tidak menggunakan evaluasi metrik yang lengkap seperti MSE, RMSE, atau R^2. Selain itu, masalah *overfitting* belum dipelajari secara menyeluruh, dan analisis performa generalisasi model terhadap data uji seringkali terbatas. Ini menjadi gap penelitian yang penting untuk dijawab. Berdasarkan latar belakang di atas, maka penelitian ini berusaha mengisi celah tersebut dengan melakukan perbandingan menyeluruh antara empat algoritma pembelajaran mesin yang digunakan untuk memprediksi diabetes melitus tipe 2, yaitu dengan *Random Forest, Logistic Regression, Deep Neural Network, dan Multilayer Perceptron*. Tujuan utama penelitian ini adalah untuk mengevaluasi dan membandingkan performa keempat algoritma dalam klasifikasi DMT2 berdasarkan data medis, serta menganalisis potensi overfitting dan efektivitas generalisasi model. Penelitian ini diharapkan dapat membantu mengembangkan sistem pendukung keputusan berbasis kecerdasan buatan yang dapat mendeteksi diabetes secara lebih akurat dan efektif. Penelitian ini juga diharapkan dapat mengungkap kelebihan dan kekurangan setiap algoritma dalam data medis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam penelitian ini dijelaskan secara umum pada gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menggambarkan alur tahapan penelitian yang dimulai dari proses pengumpulan *dataset Pima Indians Diabetes Database*, yang terdiri dari 9538 sampel dengan 16 atribut fitur dan 1 atribut target. Setelah *dataset* dikumpulkan dan dilakukan *preprocessing* data, *dataset* kemudian dibagi menjadi dua bagian, yaitu data *training* (80%) dan data *testing* (20%). Empat algoritma *machine learning* digunakan pada tahap pelatihan model: *Logistic Regression (LR), Random Forest (RF), Multi-Layer Perceptron (MLP), dan Deep Neural Network (DNN).* Algoritma tersebut dipilih karena kemampuannya dalam menangani data kesehatan dan tugas klasifikasi [4][7][8]. Setelah model dasar dibuat, pengoptimalan *hyperparameter tuning* dilakukan menggunakan teknik *Grid Search* dan *Cross-Validation* untuk mengoptimalkan kinerja algoritma [9][10][11]. Langkah ini memungkinkan untuk menemukan kombinasi parameter terbaik untuk masing-masing model. Selanjutnya model yang telah dioptimalkan digunakan untuk memprediksi status diabetes pada data tes. Berbagai metrik, termasuk *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), *R-squared* (R^2), *Confusion Matrix,* dan *Analysis of Overfitting Gap,* digunakan untuk evaluasi performa model. Metrik-metrik ini memberikan gambaran menyeluruh tentang akurasi, presisi, dan generalisasi model.

Model dengan performa terbaik kemudian disimpan untuk digunakan dan diterapkan di masa depan. Metode sistematis ini menjamin pembuatan model prediksi DMT2 yang akurat dan dapat diandalkan. Metode ini memiliki potensi besar untuk digunakan dalam skrining dini dan manajemen penyakit.

2.2 Data Penelitian

Penelitian ini menggunakan *Dataset* yang disebut *Pima Indians Diabetes,* yang awalnya berasal dari *National Institute of Diabetes and Digestive and Kidney Diseases* [12]. Tujuan dari *dataset* ini adalah untuk menggunakan sejumlah pengukuran medis untuk memprediksi apakah seorang pasien menderita diabetes atau tidak. *Dataset* ini diperoleh melalui

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8747
Hal 414-421
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

platform Kaggle. *Dataset* ini unik karena semua entri datanya adalah perempuan dari kelompok etnis Pima Indian di Amerika Serikat yang berusia minimal 21 tahun. Data ini terdiri dari 9538 baris data dengan 16 fitur yang digunakan sebagai variabel prediktor. Selain itu, ada variabel target (*Outcome*) yang menunjukkan nilai 1 untuk pasien yang menderita diabetes dan nilai 0 untuk pasien yang tidak diabetes. *Dataset* ini dipilih karena banyak digunakan dalam penelitian sebelumnya dan bersifat terbuka. Selain itu, *dataset* ini memiliki fitur medis yang relevan untuk tugas klasifikasi yang diberikan oleh algoritma pembelajaran mesin. Tabel 1 menunjukkan *dataset* yang digunakan dalam penelitian ini.

Tabel 1. Sampel *Dataset*

Age	*Pregnancies*	*BMI*	*Glucose*	*BloodPressure*	*HbA1c*	*LDL*	*HDL*	*WHR*
69	5	28,39	130,1	77	5,4	130,4	44	0,84
32	1	26,49	116,5	72	4,5	87,4	54,2	1,39
89	13	25,34	101	82	4,9	112,5	56,8	0,79
78	13	29,91	146	104	5,7	50,7	39,1	0,99
38	8	24,56	103,2	74	4,7	102,5	29,1	0,91
…..	…..	…..	…..	…..	…..	…..	…..	…..
19	0	31,11	89,4	75	4,3	135,6	62,8	0,75
21	11	28,45	86,4	87	4	125,2	48,4	1,46
35	8	28,26	104,1	73	4,4	107,9	50,2	0,86
79	14	24,59	121,8	89	4,9	36,8	58,9	1,13
32	4	28,34	50	86	4	64,7	43,4	0,96

Tabel 1 menyajikan representasi sebagian data dari dataset yang digunakan dalam penelitian ini. Tabel tersebut menampilkan sejumlah fitur utama yang dianggap relevan dalam klasifikasi diabetes mellitus tipe 2, seperti usia, jumlah kehamilan, indeks massa tubuh (BMI), kadar glukosa, tekanan darah, HbA1c, LDL, HDL, dan rasio pinggang-pinggul (WHR). Secara keseluruhan, dataset ini terdiri dari 16 fitur numerik yang digunakan sebagai variabel input dalam proses pelatihan dan evaluasi model klasifikasi.

2.3 Pembagian Data

Pada tahap ini, *dataset* dibagi menjadi dua bagian utama, yaitu data latih dan data uji. Data latih digunakan untuk melatih model agar mampu mempelajari pola dari data, sedangkan data uji digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya (*unseen*). Proporsi pembagian *dataset* tersebut disajikan pada Tabel 2.

Tabel 2. Pembagian *Dataset*

Jenis Data	Jumlah Data	Persentase (%)
Data Latih	7630	80
Data Uji	1908	20
Total	**9538**	**100**

Sebagaimana ditunjukkan pada Tabel 2, sebanyak 80% data dialokasikan sebagai data latih, dan 20% sisanya sebagai data uji untuk evaluasi akhir model. Proses pembagian ini dilakukan secara acak, namun menggunakan parameter *random_state* tertentu agar hasilnya dapat direproduksi. Teknik pembagian semacam ini umum digunakan dalam pembelajaran mesin untuk memastikan bahwa model dapat mengenali pola dari data pelatihan dan mampu melakukan generalisasi terhadap data baru.

2.4 *Hyperparameter Tuning*

Proses pencarian kombinasi *hyperparameter* terbaik untuk meningkatkan kinerja model statistik yang dikenal sebagai *hyperparameter tuning*. Ini dilakukan dengan melatih model dengan nilai *hyperparameter* yang berbeda dalam rentang yang telah ditentukan. Kemudian, berdasarkan matrik evaluasi yang dipilih, *hyperparameter* dengan hasil prediksi terbaik terhadap data diluar sampel (*Out-of-sample*) akan digunakan sebagai konfigurasi akhir model [13]. Dalam penelitian ini untuk algoritma klasik seperti *Logistik Regression* (LR) dan *Random Forest* (RF), Penyesuaian dilakukan melalui teknik *Grid Search Cross Validation (GridSearchCV)* dari pustaka *scikit-learn* [14][15][16][17]. Teknik ini menguji setiap kombinasi nilai *hyperparameter* yang telah ditentukan secara eksplisit, dan setiap kombinasi diuji melalui validasi silang [18]. Dalam kasus ini, lima kali validasi silang sehingga hasil penyesuaian tidak hanya bergantung pada kinerja satu subset data saja, melainkan diuji secara konsisten pada beberapa bagian data latih. Metode ini memungkinkan untuk mendapatkan kombinasi *hyperparameter* yang paling akurat dan stabil saat diuji pada data validasi. Kemudian, kombinasi ini digunakan untuk melatih ulang model sebelum dilakukan evaluasi terhadap data uji. Berbeda dengan model klasik, algoritma berbasis jaringan saraf seperti *Deep Neural Network* (DNN) dan *Multilayer Perceptron* (MLP) membutuhkan *tuning* yang lebih kompleks karena memiliki lebih banyak parameter dan arsitektur [19][20][21]. Untuk mengoptimalkan kinerja klasifikasi, penelitian ini menggunakan aturan *hyperparameter* pada algoritma MLP. Model yang digunakan memiliki dua lapisan tersembunyi (*Hidden layer*) yang memiliki ukuran neuron 50 dan 30, serta fungsi aktivasi ReLU,

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8747
Hal 414-421
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

solver Adam, dan jumlah maksimum iterasi 200 [22][23][24][25]. Struktur ini dipilih karena jika menggunakan tiga lapisan tersembunyi atau kurang menghasilkan akurasi yang tinggi sambil mempertahankan waktu pelatihan yang efisien [26]. Model cenderung mengalami *overfitting* jika terlalu banyak lapisan tersembunyi (*Hidden Layer*) digunakan, sementara terlalu sedikit dapat menyebabkan *underfitting*, terutama pada masalah klasifikasi kompleks seperti diagnosis penyakit. Fungsi *ReLu* dipilih karena sangat efektif dan banyak digunakan dalam data tabular. Meskipun ada kekurangan, seperti kehilangan warna, *ReLU* tetap memiliki kinerja yang stabil dibandingkan dengan varian lain [27]. Sementara itu *optimizer Adam* digunakan karena mampu menyesuaikan proses pelatihan secara otomatis [28]. Dalam penelitian ini, kombinasi *ReLU*, dan Adam dianggap efektif sebagai konfigurasi awal model MLP. Pada model DNN *hyperparameter tuning* dilakukan dengan membuat tiga lapisan: Dua *Hidden layer* dengan 16 dan 8 neuron, dan satu lapisan *output* memiliki aktivasi *sigmoid.* Model dikompilasi menggunakan *optimizer* Adam dan *Loss Function inary_crossentropy,* sesuai dengan prosedur umum untuk klasifikasi biner [29]. Proses pelatihan dilakukan selama 50 *epoch* dengan *batch_size* sebesar 16 dan *validation_split* sebesar 20%. Proses pelatihan dilakukan dengan *early stopping* dengan *patience* 5 dan *restore_best_weight True* untuk mencegah terjadinya *overfitting* saat validasi *Loss* tidak membaik [30]. Ditunjukkan bahwa kombinasi pengaturan ini akan mempertahankan stabilitas dan akurasi pelatihan serta meningkatkan *generalisasi* model terhadap data uji [31].

2.5 Pelatihan Model

Pada tahap ini, hasil terbaik dari proses *hyperparameter tuning* digunakan untuk melatih semua model. *Logistic Regression* (LR) adalah salah satu metode klasifikasi biner yang paling umum digunakan dalam bidang kesehatan, karena sederhana, interpretasi yang mudah, serta kemampuan untuk menangani data dengan label terbatas [32][33][34]. Teknik validasi silang digunakan selama proses pelatihan untuk menghindari *overfitting*. LR masih kompetitif dalam klasifikasi data kesehatan terutama ketika dikombinasikan dengan teknik regularisasi yang tepat serta pemilihan fitur yang tepat [35]. Namun, linearitas yang tidak diuji dan penanganan prediktor kontinu yang tidak tepat menyebabkan banyak kesalahan implementasi LR dalam prediksi klinis [35][36]. Oleh karena itu, penelitian ini menggunakan penalti L2 (*ridge*) dan *tuning* terhadap parameter C untuk melakukan regularisasi. *GridSearchCV* digunakan untuk menemukan kombinasi parameter yang paling cocok berdasarkan skor validasi.

Random Forest (RF) adalah algoritma *ensemble* berbasis pohon keputusan yang digunakan untuk klasifikasi karena memiliki kemampuan untuk menangani data yang sangat besar dan terhindar dari *overfitting* [37]. Parameter yang diatur termasuk jumlah pohon, kedalaman maksimum, dan jumlah sampel minimum untuk pembagian . *GridSearchCV* digunakan untuk mencari parameter terbaik selama proses pelatihan. Selain itu RF terbukti efektif dalam kondisi data yang kompleks dan bising, seperti saat tingkat gangguan tinggi, dan menunjukkan stabilitas dan akurasi yang tinggi dalam berbagai jenis data, yang membuatnya cocok untuk tugas klasifikasi prediktif seperti prediksi diabetes [37].

Model *Multilayer Perceptron* (MLP) dilatih menggunakan dua lapisan tersembunyi (*hidden layer*) dengan 50 dan 30 *neuron*. Selain itu, ada fungsi aktivasi *ReLU* dan *optimizer Adam* dengan jumlah iterasi maksimal sebanyak 200. *ReLU* digunakan karena efisiensi pelatihannya dan kemampuan untuk menghindari *vanishing gradient*, sementara *Adam* mempercepat konvergensi dan menghasilkan akurasi tinggi [38]. Kombinasi ini biasanya digunakan dalam arsitektur jaringan tiga lapis dengan *learning rate* rendah seperti 0,001, yang secara empiris mampu mencapai akurasi 97% [38]. Pendekatan berbasis regularisasi yang diterapkan pada *framework* DRIM-MLP telah terbukti meningkatkan akurasi klasifikasi secara signifikan, meskipun MLP konvensional memiliki keterbatasan dalam membedakan *input* yang tumpang tindih antar kelas [39]. Penelitian ini memilih konfigurasi dua lapisan tersembunyi (*hidden layer*) sebagai kompromi antara kompleksitas dan efisiensi. Konfigurasi ini juga sesuai untuk tugas klasifikasi biner pada data medis.

Model *Deep Neural Network* (DNN) dilatih dengan arsitektur tiga lapis. Ini terdiri dari dua lapisan tersembunyi (*hidden layer*) yaitu dengan jumlah masing-masing 16 dan 8 *neuron*, dan satu lapisan *output* yang memiliki fungsi aktivasi *sigmoid* untuk klasifikasi biner. Lapisan tersembunyi (*hidden layer*) menggunakan aktivasi *ReLU* karena kemampuan untuk menghindari masalah *vanishing gradient* dan mempercepat konvergensi. *Optimizer Adam* dipilih karena mampu mempercepat proses pelatihan dan stabil dalam pembaruan bobot. Proses pelatihan dapat berlangsung selama maksimal 50 *epoch* dengan ukuran *batch* 16, dan menggunakan mekanisme *early stopping* yang berbasis pada nilai *validation loss* untuk mencegah *overfitting*. Penggunaan *ReLU* dan *Adam* ini sesuai dengan praktik umum dalam pelatihan DNN kontemporer yang menyeimbangkan akurasi konvergensi, dan efisiensi sumber daya komputasi [40], [41], [42].

2.6 Evaluasi Model

Pada tahap ini, model dievaluasi menggunakan metrik utama seperti *confusion matrix, accuracy, precision, recall,* dan *F1-score*. *Confusion matrix* mengidentifikasi kesalahan klasifikasi, sementara *F1-score* memberikan penilaian seimbang antara *precision* dan *recall*, terutama pada data yang tidak seimbang [43]. Selain itu, metrik regresi tambahan seperti MSE, RMSE, dan *R-squared* (R^2) juga digunakan untuk mengukur tingkat kesalahan prediksi model. R^2 menunjukkan seberapa baik model menjelaskan *varians* dari data target [44]; semakin mendekati nilai 1, semakin baik kemampuan model dalam menjelaskan *varians* tersebut. Untuk mengidentifikasi *overfitting* digunakan metode seperti validasi silang dan *early stopping*, dengan membandingkan performa antara data latih dan data uji. Tabel 3 menyajikan struktur *confusion matrix* yang digunakan sebagai dasar evaluasi performa model klasifikasi. Struktur confusion matrix dan rumus matematis untuk menghitung metrik evaluasi performa model disajikan dalam Tabel 3. Komponen utama seperti TP, TN, FP, dan FN menjadi dasar perhitungan untuk metrik klasifikasi seperti *accuracy, precision, recall, dan F1-score*. Selain itu, MSE,

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8747
Hal 414-421
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

RMSE, dan R^2 juga ditampilkan sebagai metrik tambahan untuk mengukur deviasi prediksi serta seberapa baik model menjelaskan variasi data aktual.

Tabel 3. Struktur *Confusion Matrix*

	Predicted Positif	**Predicted Negatif**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (5)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (7)$$

Simbol $n, y_i, \hat{y}_i$, dan $\bar{y}$ masing-masing mempresentasikan jumlah total data uji, nilai aktual, nilai hasil prediksi, serta nilai rata-rata dari seluruh nilai aktual y_i.

2.7 Pemilihan Model Terbaik

Pilihan model terbaik didasarkan pada hasil evaluasi yang dilakukan pada data uji menggunakan metrik klasifikasi seperti *accuracy, precision, recall, F1-score*, serta metrik tambahan seperti MSE, RMSE, dan R^2. Ini karena dalam klasifikasi diabetes, kesalahan dalam mendeteksi kasus positif (*False Negative*) harus diminimalkan. Selain performa metrik, analisis *overfitting* dilakukan dengan membandingkan hasil evaluasi pada data uji dan data latih. Model yang menunjukkan kinerja yang sangat baik pada data latih tetapi sangat buruk pada data uji dianggap mengalami *overfitting*. Ini adalah indikator yang dikenal sebagai *overfitting gap*. Model dengan *gap* yang rendah dianggap lebih umum dan dapat diandalkan ketika dihadapkan pada data baru. Setelah menemukan model terbaik, model tersebut kemudian disimpan untuk digunakan kembali tanpa perlu dilatih ulang. Penyimpanan ini memungkinkan integrasi ke dalam aplikasi pendukung keputusan medis atau sistem prediksi.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dan analisis performa keempat model yang digunakan, yaitu *Logistic Regression* (LR), *Random Forest* (RF), *Multilayer Perceptron* (MLP), dan *Deep Neural Network* (DNN). Sebelum evaluasi performa dilakukan, seluruh model dilatih menggunakan data latih sebanyak 80% dari total *dataset*, sedangkan 20% sisanya digunakan sebagai data uji. LR dan RF dioptimasi menggunakan teknik *GridSearchCV* dengan 5-*fold cross-validation*. MLP dilatih menggunakan dua *hidden* layer (50 dan 30 neuron) dengan fungsi aktivasi ReLU dan *solver* Adam. Model DNN dibangun menggunakan *TensorFlow* dengan dua *hidden layer* (16 dan 8 neuron) dilatih selama 50 *epoch* dengan *batch size* 16, serta menerapkan teknik *early stopping* dan *validation split* sebesar 20%. Evaluasi performa dilakukan pada data uji yang benar-benar terjaga (*unseen*) oleh data latih menggunakan metrik klasifikasi dan regresi seperti *accuracy, precision, recall, F1-score*, MSE, RMSE, dan R^2 untuk menilai akurasi model, kestabilan prediksi, serta potensi *overfitting*.

3.1 Hasil Evaluasi Model

Dengan menggunakan data uji, keempat model penelitian ini dievaluasi untuk menilai performanya dalam klasifikasi diabetes *mellitus* tipe 2. *Precision, Accuracy, Recall,* dan *F1-score* adalah metrik klasifikasi utama yang digunakan untuk evaluasi model. Selain itu, untuk mengukur sejauh mana prediksi probabilistik model menyimpang dari nilai aktual, terutama pada model yang menghasilkan *output* probabilitas seperti DNN dan *Logistic Regression*, digunakan pula metrik *error* numerik seperti *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE). Hasil lengkap dari evaluasi performa keempat model disajikan pada Tabel 4.

JURIKOM (Jurnal Riset Komputer), Vol. 12 No. 4, Agustus 2025
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v12i4.8747
Hal 414-421
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

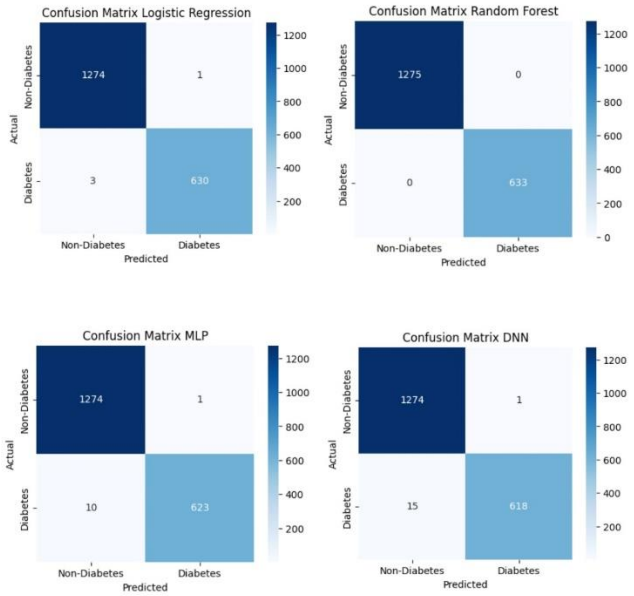
Tabel 4. Hasil Evaluasi Performansi Model

Model	Accuracy	Precision	Recall	F1-Score	MSE	RMSE	Overfitting gap	R^2
Logistic Regression	1.00	1.00	1.00	1.00	0.0021	0.0458	0.0018	0.9972
Random Forest	1.00	1.00	1.00	1.00	0.0000	0.0000	0.0000	1.0000
MLP	0.99	0.9920	0.9810	0.9865	0.0058	0.0759	0.0031	0.9740
DNN	0.99	0.9936	0.9794	0.9864	0.0084	0.0916	0.0061	0.9622

Berdasarkan tabel 4, model *Logistic Regression* dan *Random Forest* menunjukkan performa yang sangat baik, dengan nilai *accuracy* dan *F1-score* mencapai 1.00 atau mendekati sempurna. Khususnya, *Random Forest* mencatat nilai MSE dan RMSE sebesar 0.0000, yang menunjukkan tidak adanya *deviasi* antara hasil prediksi dan nilai aktual. Temuan ini sejalan dengan hasil penelitian Wang et al. (2021) yang menyatakan bahwa *Random Forest* mampu menghasilkan klasifikasi diabetes dengan akurasi tinggi melalui pendekatan *preprocessing* kompleks seperti kombinasi *SVM-SMOTE* dan *LASSO*. Hasil maksimal yang dicapai dalam penelitian ini, bahkan tanpa tahapan *preprocessing* yang rumit, memperkuat posisi *Random Forest* sebagai model yang andal dalam bidang prediksi kesehatan [45]. *Logistic Regression* juga menunjukkan performa luar biasa, dengan nilai MSE sebesar 0.0021 dan nilai RMSE sebesar 0.0458. Sementara itu, model MLP dan DNN tetap menunjukkan hasil yang sangat baik meskipun sedikit di bawah dua model pertama. DNN mencatat nilai MSE tertinggi (0.0084) dan nilai R^2 terendah (0.9622), yang mengindikasikan adanya kecenderungan *overfitting* ringan. Meskipun demikian, seluruh nilai evaluasi tetap berada dalam kategori sangat baik dan menunjukkan bahwa kedua model masih layak untuk digunakan dalam konteks klasifikasi medis.

3.2 Visualisasi *Confusion Matrix*

Gambar 2 menunjukkan *confusion matrix* dari keempat model yang digunakan dalam penelitian ini, yaitu *Logistic Regression, Random Forest, Multilayer Perceptron, dan Deep Neural Network*. Setiap *confusion matrix* memperlihatkan jumlah prediksi benar dan salah untuk kedua kelas, yaitu Non-Diabetes dan Diabetes.



Gambar 2. *Confusion matrix* untuk Keempat Model: LR, RF, MLP, dan DNN

Berdasarkan visualisasi pada Gambar 2, model *Random Forest* menunjukkan kinerja sempurna dengan nilai *True Positive* (TP) dan *True Negative* (TN) yang maksimal tanpa kesalahan klasifikasi. *Logistic Regression* juga memberikan hasil yang sangat baik, dengan hanya satu kesalahan kelas negatif (FN) dan tiga kesalahan pada kelas positif (FP). Kesalahan klasifikasi yang lebih tinggi terjadi pada model MLP dan DNN, terutama pada kelas positif, yang menunjukkan bahwa sebagian kecil pasien diabetes tidak dapat diidentifikasi dengan benar sehingga berpotensi menjadi risiko dalam konteks prediksi medis. Secara keseluruhan, visualisasi ini mendukung hasil analisis numerik sebelumnya, di mana model *Random Forest* unggul dalam akurasi, kestabilan, dan ketepatan klasifikasi dalam semua aspek.

3.3 Pemilihan Model Terbaik

Secara keseluruhan, model *Random Forest* adalah yang terbaik. Model ini berhasil memprediksi semua data uji dengan benar, tanpa kesalahan klasifikasi, serta nilai *accuracy, precision, F1-score,* dan R^2 yang semuanya mencapai nilai 1.00. Dengan nilai *False Positive* dan *False Negative* sama dengan nol, hal ini juga dikonfirmasi oleh *Confusion Matrix*. Selain itu, nilai MSE dan RMSE 0.0000 menunjukkan bahwa prediksi model secara numerik sangat dekat dengan nilai aktual.

Stabilitas model juga diperkuat dengan tidak adanya indikasi *overfitting*. Model *Logistic Regression* juga menunjukkan kinerja sangat tinggi dengan hanya empat kesalahan klasifikasi dan nilai *F1-Score* sebesar 1.00. Namun dalam hal keakuratan klasifikasi data uji, *Random Forest* masih jauh dibawahnya. Model ini bagus untuk sistem pendukung keputusan dasar dan pendukung keputusan karena efisiensi komputasi dan interpretabilitasnya.

Model MLP dan DNN menghasilkan performa yang baik, namun dengan sedikit penurunan pada nilai *Recall* dan *F1-Score* karena munculnya *False Negative* yang tinggi. Hal ini menunjukkan bahwa meskipun model memiliki tingkat akurasi yang tinggi, tetapi sedikit kurang sensitif untuk mendeteksi diabetes. Selain itu keputusan untuk tidak memilih MLP dan DNN sebagai model utama didukung oleh nilai MSE dan RMSE yang tinggi dibandingkan dua model lainnya, serta kemungkinan *overfitting* yang tampak dari performa *train-test gap*.

4. KESIMPULAN

Penelitian ini membandingkan empat algoritma *machine learning* (*Logistic Regression, Random Forest, Multilayer Perceptron, dan Deep Neural Network*) untuk klasifikasi diabetes *mellitus* tipe 2 menggunakan data medis dari Kaggle. Hasil evaluasi menunjukkan bahwa model *Random Forest* (RF) memberikan performa terbaik dengan klasifikasi sangat baik dan stabil tanpa indikasi *overfitting*. Meskipun *Logistic Regression* juga cukup baik, MLP dan DNN menunjukkan kelemahan pada *False Negative* yang relative tinggi. Setelah mempertimbangkan seluruh hasil evaluasi, maka Model *Random Forest* dipilih sebagai model terbaik. Selanjutnya, model ini dapat digunakan pada tahap implementasi sistem prediksi diabetes *mellitus* tipe 2 karena konsistensinya, akurasi tinggi, ketahanan terhadap *overfitting*, dan kemampuan untuk menangani data numerik dan klasifikasi dengan baik. Penelitian ini terbatas pada jumlah fitur medis yang sedikit dan hanya menggunakan satu dataset, sehingga disarankan agar penelitian lanjutan menggunakan data yang lebih beragam, fitur klinis yang lebih kompleks, serta teknik interpretasi model untuk meningkatkan keandalan sistem dalam praktik medis.

REFERENCES

[1] H. Yaribeygi, T. Sathyapalan, S. L. Atkin, and A. Sahebkar, "Molecular Mechanisms Linking Oxidative Stress and Diabetes Mellitus," *Oxid. Med. Cell. Longev.*, vol. 2020, 2020, doi: 10.1155/2020/8609213.

[2] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, "Retraction Note: Classification of stroke disease using machine learning algorithms (Neural Computing and Applications, (2020), 32, 3, (817-828), 10.1007/s00521-019-04041-y)," *Neural Comput. Appl.*, vol. 36, no. 17, p. 10379, 2024, doi: 10.1007/s00521-024-09844-2.

[3] P. Parhi, R. Bisoi, and P. K. Dash, "An Integrated Nature-Inspired Algorithm Hybridized Adaptive Broad Learning System for Disease Classification," *IEEE Access*, vol. 11, no. March, pp. 31636–31656, 2023, doi: 10.1109/ACCESS.2023.3262167.

[4] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, no. Ml, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[5] M. Rizky, A. Pramuntadi, D. Prastowo, and D. Hardan Gutama, "Implementation of Deep Neural Network Method on Classification of Type 2 Diabetes Mellitus Disease Implementasi Metode Deep Neural Network pada Klasifikasi Penyakit Diabetes Melitus Tipe 2," *MALCOM Indones. J. Mach. Learn. Comput. Sci.* , vol. 4, no. 3, pp. 1043–1050, 2024.

[6] Erlin, Yulvia Nora Marlim, Junadhi, Laili Suryati, and Nova Agustina, "Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 88–96, 2022, doi: 10.22146/jnteti.v11i2.3586.

[7] Y. Zeng and F. Cheng, "Medical and Health Data Classification Method Based on Machine Learning," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/2722854.

[8] Q. An, S. Rahman, J. Zhou, and J. J. Kang, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," *Sensors*, vol. 23, no. 9, 2023, doi: 10.3390/s23094178.

[9] M. Adnan, A. A. S. Alarood, M. I. Uddin, and I. ur Rehman, "Utilizing grid search cross-validation with adaptive boosting for augmenting performance of machine learning models," *PeerJ Comput. Sci.*, vol. 8, no. Ml, pp. 1–29, 2022, doi: 10.7717/PEERJ-CS.803.

[10] A. F. D. Putra, M. N. Azmi, H. Wijayanto, S. Utama, and I. G. P. W. Wedashwara Wirawan, "Optimizing Rain Prediction Model Using Random Forest and Grid Search Cross-Validation for Agriculture Sector," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 23, no. 3, pp. 519–530, 2024, doi: 10.30812/matrik.v23i3.3891.

[11] T. Yan, S. L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm," *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1292–1303, 2022, doi: 10.1016/j.jrmge.2022.03.002.

[12] "Predict Diabetes From Medical Records." Accessed: Jun. 14, 2025. [Online]. Available: https://www.kaggle.com/code/paultimothymooney/predict-diabetes-from-medical-records/input

[13] O. A. Montesinos López, A. Montesinos López, and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. 2022. doi: 10.1007/978-3-030-89010-0.

[14] M. Arambakam and J. Beel, "Federated Meta-Learning: Democratizing Algorithm Selection Across Disciplines and Software Libraries," *7th ICML Work. Autom. Mach. Learn.*, 2020, [Online]. Available: https://github.com/mukeshmk/scikit-learn

[15] G. N. Ahmad, H. Fatima, Shafiullah, A. Salah Saidi, and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV," *IEEE Access*, vol. 10, no. April, pp. 80151–80173, 2022, doi: 10.1109/ACCESS.2022.3165792.

[16] M. F. El-Amin, B. Alwated, and H. A. Hoteit, "Machine Learning Prediction of Nanoparticle Transport with Two-Phase Flow in Porous Media," *Energies*, vol. 16, no. 2, 2023, doi: 10.3390/en16020678.

[17] M. Castejón-Limas, L. Fernández-Robles, H. Alaiz-Moretón, J. Cifuentes-Rodriguez, and C. Fernández-Llamas, "A Framework for the Optimization of Complex Cyber-Physical Systems via Directed Acyclic Graph," *Sensors*, vol. 22, no. 4, 2022, doi: 10.3390/s22041490.
[18] F. M. Javed Mehedi Shamrat, P. Ghosh, M. H. Sadek, M. A. Kazi, and S. Shultana, "Implementation of Machine Learning Algorithms to Detect the Prognosis Rate of Kidney Disease," *2020 IEEE Int. Conf. Innov. Technol. INOCON 2020*, no. December, 2020, doi: 10.1109/INOCON50539.2020.9298026.
[19] T. Zhan, "Hyper-Parameter Tuning in Deep Neural Network Learning," pp. 95–101, 2022, doi: 10.5121/csit.2022.121809.
[20] D. H. Shirazi and H. Toosi, "Deep Multilayer Perceptron Neural Network for the Prediction of Iranian Dam Project Delay Risks," *J. Constr. Eng. Manag.*, vol. 149, no. 4, p. 04023011, Apr. 2023, doi: 10.1061/JCEMD4.COENG-12367;JOURNAL:JOURNAL:JCEMD4;WGROUP:STRING:PUBLICATION.
[21] S. Afzal, B. M. Ziapour, A. Shokri, H. Shakibi, and B. Sobhani, "Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms," *Energy*, vol. 282, p. 128446, Nov. 2023, doi: 10.1016/J.ENERGY.2023.128446.
[22] S. B. Chavan and S. V. Shinde, "An Experimental Investigation of U-Net-Based Deep Learning Segmentation for Histopathology Images," *2023 1st DMIHER Int. Conf. Artif. Intell. Educ. Ind. 4.0, IDICAIEI 2023*, 2023, doi: 10.1109/IDICAIEI58380.2023.10406634.
[23] X. Wu, Y. Ji, and X. Li, "High-accuracy handwriting recognition based on improved CNN algorithm," *2021 IEEE 3rd Int. Conf. Commun. Inf. Syst. Comput. Eng. CISCE 2021*, pp. 344–348, May 2021, doi: 10.1109/CISCE52179.2021.9445924.
[24] A. Barbu, "Training a Two-Layer ReLU Network Analytically," *Sensors*, vol. 23, no. 8, 2023, doi: 10.3390/s23084072.
[25] D. A. Kristiyanti, W. B. N. Pramudya, and S. A. Sanjaya, "How can we predict transportation stock prices using artificial intelligence? Findings from experiments with Long Short-Term Memory based algorithms," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, p. 100293, 2024, doi: 10.1016/j.jjimei.2024.100293.
[26] M. Uzair and N. Jamil, "Effects of Hidden Layers on the Efficiency of Neural networks," *Proc. - 2020 23rd IEEE Int. Multi-Topic Conf. INMIC 2020*, pp. 1–6, 2020, doi: 10.1109/INMIC50486.2020.9318195.
[27] S. Mastromichalakis, "ALReLU: A different approach on Leaky ReLU activation function to improve Neural Networks Performance," pp. 1–10, 2020, [Online]. Available: http://arxiv.org/abs/2012.07564
[28] Y. Nawaz, M. U. Hashmi, M. Ali, H. Bibi, M. Ali, and A. Manan, "Performance Evaluation of Various Optimizers for Breast Cancer Diagnosis Using Neural Networks," vol. 08, no. 01, 2024.
[29] S. A. Abdullah and A. Al-Ashoor, "An artificial deep neural network for the binary classification of network traffic," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 402–408, 2020, doi: 10.14569/ijacsa.2020.0110150.
[30] Y. Bai *et al.*, "Understanding and Improving Early Stopping for Learning with Noisy Labels," *Adv. Neural Inf. Process. Syst.*, vol. 29, no. NeurIPS, pp. 24392–24403, 2021.
[31] I. Kandel and M. Castelli, "The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset," *ICT Express*, vol. 6, no. 4, pp. 312–315, 2020, doi: 10.1016/j.icte.2020.04.010.
[32] K. J. Olowe, N. L. Edoh, S. Jean, C. Zouo, J. Olamijuwon, and I. Researcher, "Comprehensive review of logistic regression techniques in predicting health outcomes and trends," 2024.
[33] L. Pinheiro-Guedes, C. Martinho, and M. R. O Martins, "Logistic Regression: Limitations in the Estimation of Measures of Association with Binary Health Outcomes," *Acta Med. Port.*, vol. 37, no. 10, pp. 697–705, 2024, doi: 10.20344/amp.21435.
[34] P. Schober and T. R. Vetter, "Statistical Minute," *Int. Anesth. Res. Soc.*, vol. 129, no. 2, p. 2019, 2019.
[35] Y. Guo, Y. Ge, Y. C. Yang, M. A. Al-Garadi, and A. Sarker, "Comparison of Pretraining Models and Strategies for Health-Related Social Media Text Classification," *Healthc.*, vol. 10, no. 8, pp. 1–15, 2022, doi: 10.3390/healthcare10081478.
[36] E. Adeli, X. Li, D. Kwon, Y. Zhang, and K. M. Pohl, "Logistic Regression Confined by Cardinality-Constrained Sample and Feature Selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 7, pp. 1713–1728, Jul. 2020, doi: 10.1109/TPAMI.2019.2901688.
[37] J. Hatwell, M. M. Gaber, and R. M. A. Azad, *CHIRPS: Explaining random forest classification*, vol. 53, no. 8. Springer Netherlands, 2020. doi: 10.1007/s10462-020-09833-6.
[38] Y. R. Youn and J. Hong, "Optimization of Model based on Relu Activation Function in MLP Neural Network Model," vol. 13, no. 2, pp. 80–87, 2024.
[39] R. Mondal, T. Pal, and P. Dey, "Discriminative Regularized Input Manifold for multilayer perceptron," *Pattern Recognit.*, vol. 151, p. 110421, Jul. 2024, doi: 10.1016/J.PATCOG.2024.110421.
[40] X. Hu, L. Chu, J. Pei, W. Liu, and J. Bian, "Model complexity of deep learning: a survey," *Knowl. Inf. Syst.*, vol. 63, no. 10, pp. 2585–2619, 2021, doi: 10.1007/s10115-021-01605-0.
[41] S. M. Nabavinejad, S. Reda, and M. Ebrahimi, "BatchSizer: Power-Performance Trade-off for DNN Inference," *Proc. Asia South Pacific Des. Autom. Conf. ASP-DAC*, pp. 819–824, 2021, doi: 10.1145/3394885.3431535.
[42] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, 2023, doi: 10.1007/s00521-023-08568-z.
[43] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.
[44] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.
[45] X. Wang *et al.*, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–14, 2021, doi: 10.1186/s12911-021-01471-4.