

Algoritma *Supervised Machine Learning* Untuk Klasifikasi Diagnosa Penyakit Hipertensi

Bambang Eka Purnama¹, Yusuf Sutanto²

Politeknik Kesehatan Bhakti Mulia¹, Universitas Dharma AUB²
bambang@poltekkesbhaktimulia.ac.id¹, yusuf.sutanto@stie-aub.ac.id²

Diterima (22-11-2026)	Direvisi (08-01-2026)	Disetujui (25-02-2026)
--------------------------	--------------------------	---------------------------

Abstrak - Setiap tahun, prevalensi penderita hipertensi mengalami peningkatan signifikan, dan diperkirakan tahun 2025, total pasien hipertensi akan menyentuh 1,5 miliar jiwa. Salah satu metode dalam *data mining* yang sering digunakan untuk klasifikasi adalah *Support Vector Machine* (SVM). Metode SVM berupaya menemukan *hyperplane* atau fungsi batas keputusan terbaik yang mampu memisahkan dua kelas atau lebih dari data dalam ruang masukan. Penelitian ini bertujuan untuk menentukan hasil klasifikasi dan akurasi diagnosis hipertensi menggunakan metode SVM. Sebelas atribut yang digunakan meliputi usia, kebiasaan merokok, aktivitas fisik, mengkonsumsi gula, mengkonsumsi garam, mengkonsumsi lemak, mengkonsumsi alkohol, kurangnya mengkonsumsi buah dan sayur, serta tekanan darah sistolik dan diastolik. Penelitian ini akan memanfaatkan perangkat lunak Jupyter Notebook dan bahasa pemrograman Python sebagai instrumen penelitian. Metode SVM dilatih dengan berbagai atribut kernel dan hyperparameter untuk menghasilkan model terbaik. Hasil penelitian menunjukkan bahwa kernel RBF dengan parameter $C=100$ dan $\gamma=0,1$ menghasilkan akurasi sebesar 97,7%, menjadikannya model terbaik dalam mengklasifikasikan hipertensi. Dari hasil ini, dapat disimpulkan bahwa metode SVM mampu menghasilkan klasifikasi diagnosis hipertensi yang sangat baik dan dapat memberikan diagnosis untuk mendeteksi hipertensi secara dini. Lebih jauh lagi, luaran penelitian ini diharapkan berkontribusi dalam menekan angka mortalitas akibat hipertensi dan berpotensi untuk diaplikasikan lebih luas demi kemaslahatan masyarakat.

Kata Kunci : hipertensi, klasifikasi, *machine learning*, SVM

Abstract - Forecasts predict that total of hypertension sufferers will escalate to 1.5 billion people by 2025. For classification purposes, a widely recognized algorithm is Support Vector Machine (SVM). SVM algorithm seeks to establish the most effective hyperplane, which functions as a decision boundary to delineate between two or more classes of data in feature space. This research endeavors to ascertain classification outcomes and hypertension diagnosis accuracy by employing SVM method. The analysis is based on eleven attributes, which consist of age, smoking habits, physical activity, consumption of sugar, salt, and fat, alcohol intake, inadequate fruit and vegetable consumption, and both systolic and diastolic blood pressure. Jupyter Notebook and Python are utilized as the computational tools for this research. To achieve the most effective model, SVM method was subjected to training with different kernel functions and hyperparameter configurations. The findings revealed that RBF kernel, configured with parameters $C=100$ and $\gamma=0.1$, resulted in the best model for hypertension classification, achieving an accuracy of 97.7%. These results lead to the conclusion that SVM method offers a robust solution for hypertension diagnosis and can aid in its early identification. Moreover, the outcomes of this research are expected contribute to reducing mortality rates associated with hypertension and offers scalable applications for public health benefits.

Keywords: hypertension, classification, machine learning, SVM

I. PENDAHULUAN

Hipertensi, atau tekanan darah tinggi, adalah suatu kondisi kesehatan yang dikenal sebagai "*silent killer*" karena sering kali tidak disertai keluhan atau gejala (Ratwatte & Celermajer, 2024). Kondisi ini merupakan pembuluh darah yang tidak normal sehingga

mengganggu kiriman oksigen dan nutrisi yang dialirkan oleh darah ke setiap jaringan tubuh yang memerlukan. Peningkatan usia berbanding lurus dengan insidensi hipertensi (Benaired et al., 2024). Hipertensi menjadi isu kesehatan global yang serius dan menimbulkan dampak dalam waktu singkat maupun jangka panjang (Li

et al., 2024). Hipertensi menjadi salah satu faktor risiko utama yang menyebabkan tingginya angka kematian akibat penyakit tidak menular di banyak negara (Xu et al., 2023). Peningkatan jumlah penderita hipertensi terjadi secara signifikan dari tahun ke tahun dan proyeksi menunjukkan bahwa pada tahun 2026, angka tersebut akan mencapai 1,6 miliar (Tri Sutanti Puji Hartati & Emyr Reisha Isaura, 2023). Diperkirakan sekitar 9,5 juta orang meninggal setiap tahunnya akibat hipertensi beserta komplikasi yang menyertainya (Kreutz et al., 2024).

Data Mining (DM) berfungsi mengidentifikasi pola yang berguna mendukung pengambilan keputusan, klasifikasi, dan prediksi. Melalui penggalian data, informasi dapat diekstrak dari volume data yang besar menjadi format yang bermanfaat dan mudah diinterpretasi (Dohan et al., 2024). Dalam DM, klasifikasi merupakan suatu metode pengelompokan data ke dalam kelas-kelas atau kategori tertentu berdasarkan ciri-ciri yang dimilikinya (Setyadi et al., 2024). Salah satu algoritma yang dikenal luas dalam klasifikasi adalah *Support Vector Machine* (SVM). SVM dapat diterapkan untuk klasifikasi data linear maupun non-linear (Amriana et al., 2025). Sebagai algoritma pembelajaran terawasi (*supervised learning*), SVM bekerja dengan mengidentifikasi model berlabel yang terdapat dalam himpunan data (Biswas & Islam, 2023). Tujuan utama adalah mencari *hyperplane* atau fungsi batas keputusan optimal yang mampu memisahkan dua atau lebih kelas data secara efektif dalam ruang masukan (Jumanto et al., 2024). Pemilihan metode SVM dalam penelitian ini didasarkan pada konsepnya yang lugas, sistematis, dan konsisten jika dibandingkan dengan algoritma klasifikasi lain. SVM mampu memberikan nilai akurasi yang tinggi, bahkan ketika berhadapan dengan data yang tidak seimbang (Guido et al., 2024). Lebih lanjut, metode ini dapat meminimalkan kesalahan klasifikasi untuk sampel yang belum dikenal serta menunjukkan kemampuan generalisasi yang lebih baik dibandingkan dengan metode Jaringan Saraf Tiruan (*Artificial Neural Network / ANN*) dan ketahanan terhadap outlier (Nurrani et al., 2023).

Di tingkat global, Organisasi Kesehatan Dunia (WHO) memprediksi bahwa prevalensi hipertensi akan mencapai 33% pada tahun 2023, di mana 67% kasusnya terjadi di negara yang berpenghasilan sedang berkembang (Rochman et al., 2024). Jumlah total penderita hipertensi diproyeksikan terus bertambah, mencapai 1,5 miliar orang di seluruh dunia pada tahun 2025 (Kurniawan et al., 2023). Di

Indonesia, prevalensi hipertensi mencapai 36% pada tahun 2024. Angka prevalensi tertinggi, yakni 45,3%, ditemukan pada kelompok usia 45–54 tahun (Ardiansyah & Widowati, 2024). Berdasarkan hal tersebut, deteksi dan penanganan hipertensi sejak dini menjadi sangat krusial. Guna meningkatkan akurasi dan efisiensi dalam mengklasifikasikan suatu penyakit, diperlukan pertimbangan terhadap sistem klasifikasi yang dibangun melalui pendekatan penambangan data. Penelitian ini bertujuan untuk menyusun klasifikasi diagnosis hipertensi dan mengukur tingkat akurasi metode SVM dari hasil klasifikasi yang dihasilkan.

Beberapa penelitian sebelumnya juga telah mengklasifikasikan hipertensi. Salah satunya adalah penelitian yang dipublikasikan dalam sebuah jurnal, yang mengembangkan aplikasi berbasis web untuk diagnosis hipertensi menggunakan metode Naïve Bayes. Penelitian tersebut menggunakan himpunan data (dataset) dari *Framingham Heart Study* (FHS) yang melibatkan 4.434 partisipan dari *National Institutes of Health*. Dengan penyesuaian konfigurasi kecil, penelitian ini mampu menghasilkan klasifikasi biner untuk kelas prediksi. Akurasi hasil pengujian untuk klasifikasi biner adalah 77,15% sebelum diskretisasi dan 76,4% setelah diskretisasi. Akurasi model yang berhasil memprediksi klasifikasi hipertensi yang dikembangkan adalah 84,28% (Munali & Armansyah, 2024).

Dalam penelitian berbeda, sebuah sistem berbasis *Internet of Things* (IoT) dikembangkan untuk mengukur tekanan darah dan detak jantung menggunakan metode *K-Nearest Neighbors* (KNN). Sistem tersebut dapat mengukur tekanan darah dan mendeteksi tekanan darah sistolik, diastolik, serta detak jantung secara waktu nyata. Pemrosesan data dengan metode KNN mengklasifikasikan status ke dalam beberapa kategori, yaitu hipertensi optimal, normal, prahipertensi, hipertensi grade 1, dan grade 2. Dengan KNN, nilai yang dihasilkan konsisten dan akurat, sehingga metode ini dapat menunjang sistem untuk mengidentifikasi dan mengklasifikasikan kondisi hipertensi (Nurhadiva, 2024).

Pada penelitian lainnya, algoritma KNN, *Principal Component Analysis* (PCA), dan *t-Distributed Stochastic Neighbor Embedding* (t-SNE) diterapkan untuk mengklasifikasikan kondisi hipertensi. Klasifikasi yang dihasilkan mencakup Normal, Hipertensi, Hipertensi Stadium 1, dan Hipertensi Stadium 2. Data sekunder sebanyak 7.794 sampel digunakan dalam penelitian ini, yang bersumber dari Rumah Sakit Umum Daerah Labuang Baji,

Makassar. Sistem ini menggunakan atribut berupa usia, berat badan, serta tekanan darah sistolik dan diastolik. Hasil pengujian akurasi menunjukkan bahwa akurasi metode KNN adalah 99%, kombinasi KNN dan PCA mencapai 100%, dan kombinasi KNN dengan t-SNE menghasilkan akurasi 99%. Dari hasil pengujian akurasi ini, dapat disimpulkan bahwa algoritma KNN, terutama ketika dikombinasikan dengan t-SNE, mampu mengklasifikasikan struktur data non-linear dengan sangat akurat dan efektif (Resky et al., 2025).

Untuk klasifikasi diagnosis hipertensi, penelitian ini memilih metode SVM. Sebagai alat bantu, penelitian ini akan menggunakan Jupyter Notebook dan bahasa pemrograman Python. Tujuannya adalah untuk mengetahui hasil klasifikasi serta tingkat akurasi diagnosis hipertensi menggunakan metode SVM. Sebanyak sebelas atribut digunakan, yang meliputi: usia, kebiasaan merokok, aktivitas fisik, mengkonsumsi gula, mengkonsumsi garam, mengkonsumsi lemak, mengkonsumsi alkohol, kurangnya mengkonsumsi buah dan sayur, serta tekanan darah sistolik dan diastolik.

II. METODOLOGI PENELITIAN

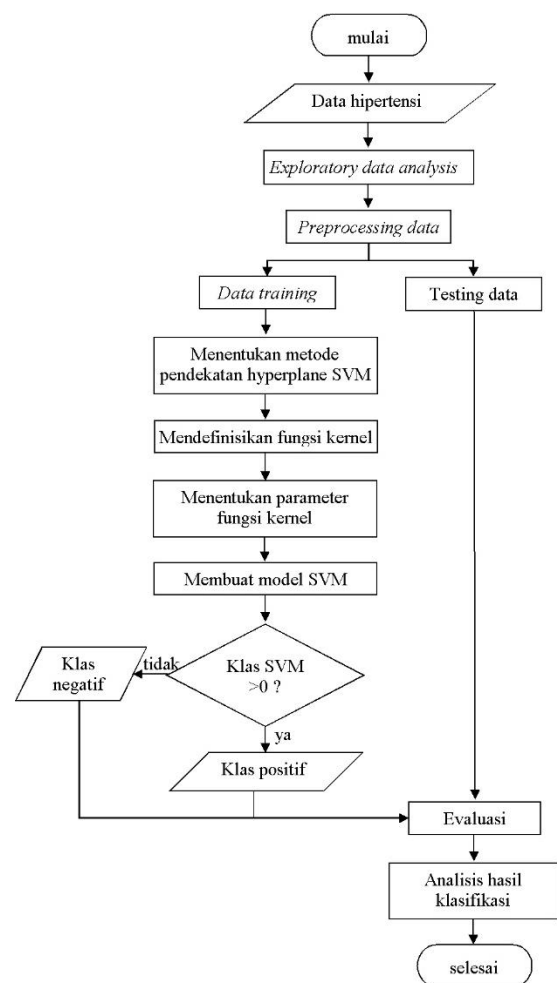
Penelitian ini menerapkan metode kuantitatif. Data kuantitatif yang digunakan berasal dari 1.373 data sekunder mengenai kasus hipertensi di Puskesmas Nusukan Surakarta. Atribut-atribut yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Table 1. Atribut Penelitian

Atribut	Keterangan
Usia	usia pasien
jenis kelamin	pria/wanita
kurang aktifitas (min. 30 menit/hari)	ya/tidak
mengkonsumsi gula berlebihan (>4 sendok makan/hari)	ya/tidak
mengkonsumsi garam berlebihan (>1 sendok tek/hari)	ya/tidak
mengkonsumsi lemak berlebihan (>4 potong/hari)	ya/tidak
kurang mengkonsumsi buah dan sayuran (<5 porsi/hari)	ya/tidak
mengkonsumsi alkohol	ya/tidak
tekanan darah sistolik	tekanan darah saat dipompa
tekanan darah diastolik	tekanan darah saat istirahat

Sumber: Hasil penelitian (2025)

Gambar 1 menyajikan diagram alir yang memuat tahapan-tahapan penelitian.



Sumber: Hasil penelitian (2025)

Gambar 1. Diagram Alir Penelitian

Tahapan yang dilakukan dalam penelitian ini terdiri dari:

1. Implementasi SVM dalam klasifikasi diagnosis hipertensi

- Pengumpulan data hipertensi yang terdiri dari variabel dependen dan independen.
- Melakukan analisis data eksploratif untuk menemukan informasi dan wawasan yang ada.
- Melakukan prapemrosesan data yang terdiri dari pembersihan data (*data cleaning*), transformasi data (*data transformation*), dan oversampling data (*data oversampling*).
- Membagi data menjadi dua bagian yaitu data pelatihan dan data pengujian dengan rasio 80:20.
- Melakukan proses klasifikasi dengan metode SVM yang meliputi: pemilihan metode untuk menentukan hyperplane SVM, pemilihan fungsi kernel, menentukan nilai parameter, dan pembentukan model SVM menggunakan

fungsi kernel. Kernel *Gaussian Radial Basis Function* (RBF) dapat dihitung dengan persamaan matematis sebagai berikut:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (1)$$

Masalah optimasi dual untuk kasus SVM non-linear dapat diformulasikan sebagai berikut:

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N y_i y_j \alpha_i \alpha_j \phi(x_i) \cdot \phi(x_j) \quad (2)$$

Nilai vektor pembobot (w) dalam Ruang Hilbert dapat dihitung dengan persamaan berikut:

$$w_i = \sum_{i=1}^N \alpha_i y_i \phi(x_i) \quad (3)$$

Persamaan untuk menghitung parameter bias b :

$$b = y_i - \sum_{i,j=1}^N \alpha_i y_i K(x_i, x_j) \quad (4)$$

Dengan proses penjumlahan dengan persamaan 4, didapatkan persamaan umum untuk menghitung *hyperplane*:

$$f(\phi(x)) = \text{sign}(\sum_{i,j=1}^N \alpha_i y_j K(x_i, x_j) + b) \quad (5)$$

(Sutanto et al., 2025)

2. Analisis Akurasi Metode SVM pada Klasifikasi Potensi Penyakit Hipertensi.

- Akurasi model klasifikasi metode SVM dievaluasi menggunakan validasi silang K-lipatan. Validasi silang K-lipatan membagi data secara acak menjadi K bagian.
- Analisis performa model klasifikasi SVM dalam memprediksi diagnosis hipertensi dengan matriks konfusi.

Ada empat persamaan yang digunakan untuk menguji kinerja model berdasarkan matriks konfusi:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (8)$$

$$F1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (9)$$

TP = *True Positive*, TN = *True Negative*
FP = *False Positive*, FN = *False Negative*

Atribut dengan tipe data numerik meliputi usia, tekanan darah sistolik, dan tekanan darah

diastolik. Tabel 2 menampilkan statistik deskriptif dari atribut-atribut numerik tersebut.

Tabel 2. Statistik Deskriptif Pada Atribut Numerik

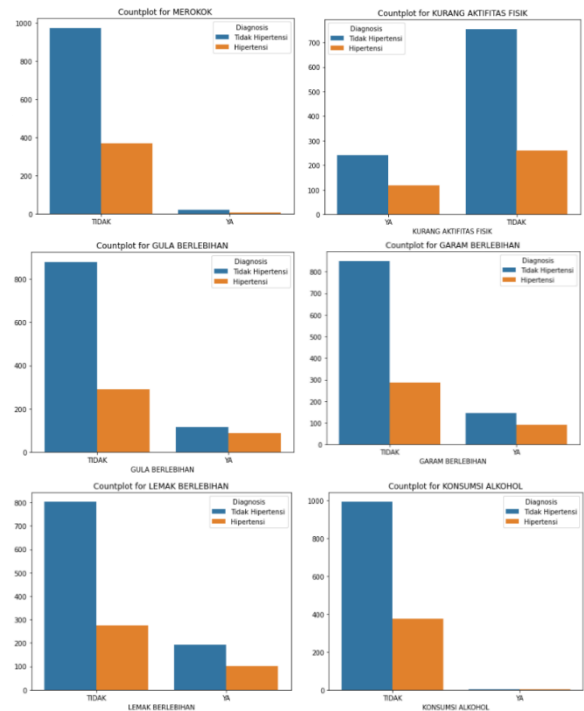
atribut	rerata	standar deviasi	min	max
usia	46,27	14,36	5	98
Sistolik	131,88	21,94	79	238
Diastolik	81,19	12,11	12	161

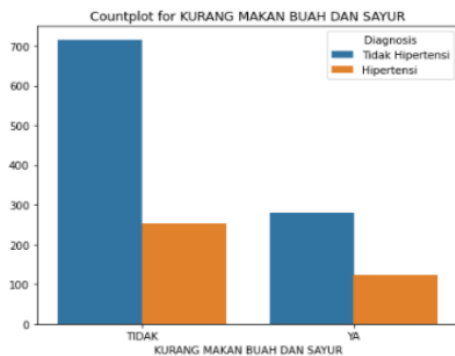
Sumber: Hasil penelitian (2025)

III. HASIL DAN PEMBAHASAN

1. Analisis Data Eklporasi

Atribut dengan tipe data kategorikal meliputi kebiasaan merokok, kurangnya aktivitas fisik, mengkonsumsi gula berlebih, mengkonsumsi garam berlebih, kurangnya asupan buah dan sayur, dan mengkonsumsi alkohol. Untuk atribut bertipe data kategorikal, dilakukan analisis bivariat menggunakan diagram batang guna mengetahui frekuensi pasien hipertensi terkait variabel dependen. Hubungan antara atribut kategorikal dan frekuensi atribut hipertensi ditunjukkan pada Gambar 2.





Sumber: Hasil penelitian (2025)

Gambar 2. Frekuensi Hubungan Antar Atribut

Berdasarkan Gambar 2 dapat diartikan sebagai berikut: orang yang tidak merokok berisiko lebih rendah terkena hipertensi, orang yang sering beraktivitas fisik mempunyai risiko lebih rendah terserang hipertensi, seseorang yang tidak sering mengonsumsi gula berlebih kecil kemungkinannya memiliki gejala hipertensi, seseorang yang tidak banyak mengonsumsi garam memiliki risiko lebih rendah menderita hipertensi, orang yang tidak terlalu banyak mengonsumsi lemak akan berisiko lebih rendah mengalami hipertensi, orang yang jarang mengonsumsi buah dan sayur mempunyai risiko lebih tinggi untuk menderita hipertensi, dan orang yang tidak mengonsumsi alkohol mempunyai risiko sangat rendah untuk menderita hipertensi.

2. Prapemrosesan Data

Table 3. Korelasi Antar Setiap Atribut

korelasi	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	1	-0.024	0.015	0.005	0.028	0.058	0.165	0.025	-0.018	-0.016
A2	-0.024	1	0.045	0.154	0.083	0.251	0.019	0.092	0.062	0.046
A3	0.015	0.045	1	0.204	0.180	0.156	-0.025	0.210	0.086	0.141
A4	0.005	0.154	0.204	1	0.190	0.171	0.036	0.210	0.111	0.121
A5	0.028	0.083	0.179	0.19	1	0.141	0.027	0.140	0.075	0.063
A6	0.058	0.251	0.156	0.170	0.141	1	0.014	0.083	0.025	0.059
A7	0.165	0.019	-0.025	0.036	0.027	0.014	1	0.033	0.017	0.022
A8	0.025	0.092	0.201	0.211	0.140	0.083	0.033	1	0.532	0.547
A9	-0.018	0.062	0.086	0.111	0.075	0.025	0.016	0.532	1	0.141
A10	-0.016	0.066	0.137	0.112	0.082	0.042	0.017	0.566	0.670	1

Sumber: Hasil penelitian (2025)

Seperti yang disajikan pada Tabel 3, hasil analisis korelasi menunjukkan temuan yang signifikan. Terdapat korelasi positif yang sangat kuat (0,670111) antara atribut tekanan darah diastolik dan diagnosis hipertensi. Di sisi lain, korelasi negatif yang paling kuat (-0,02499)

a. Transformasi Data

Untuk mengubah data dalam set data sesuai format yang dapat diproses oleh perangkat lunak yang digunakan. Pada tahap ini, dilakukan proses transformasi data atribut kategorikal menjadi numerik. Jawaban "ya" diubah menjadi angka "1" dan jawaban "tidak" diubah menjadi angka "0". Setelah semua atribut dikonversi menjadi numerik, korelasi antar variabel dapat dicari menggunakan korelasi Pearson. Hasil korelasi akan berupa nilai dengan rentang -1 hingga +1, dengan interpretasi sebagai berikut:

- 1) Jika $r = -1$, maka korelasinya negatif sempurna (hubungan berlawanan arah paling kuat).
- 2) Jika $r = 1$, maka korelasinya positif sempurna (hubungan searah paling kuat).
- 3) Jika koefisien korelasi menunjukkan 0, maka kedua variabel kemungkinan besar tidak saling memengaruhi.
- 4) Jika nilai r berada di antara -1 dan 0 atau 0 dan +1, maka hubungannya linear dengan kekuatan yang bervariasi (semakin dekat ke -1 atau +1, semakin kuat hubungannya).

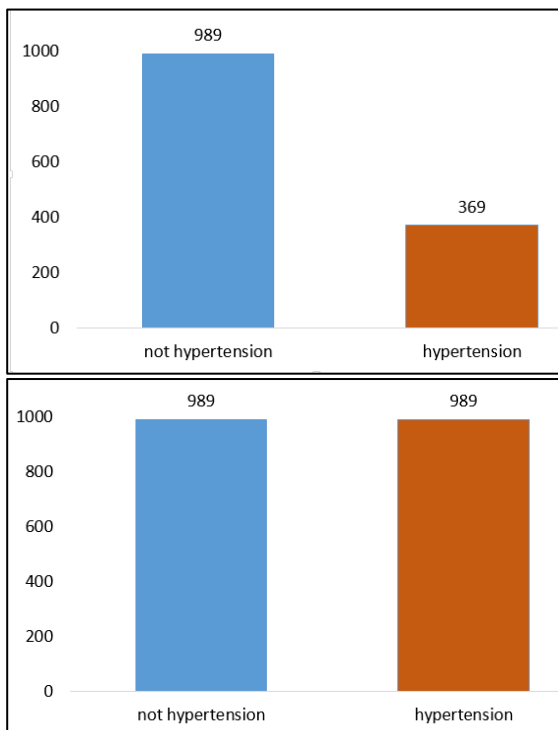
Hasil korelasi Pearson antara atribut yang digunakan dalam penelitian ini disajikan dalam Tabel 3. A1: kebiasaan merokok, A2: kurang aktivitas fisik, A3: konsumsi gula berlebihan, A4: konsumsi garam berlebihan, A5: konsumsi lemak berlebihan, A6: kurang konsumsi buah dan sayur, A7: konsumsi alkohol, A8: tekanan darah sistolik, A9: tekanan darah diastolik, A10: Diagnosis.

terjadi antara atribut konsumsi gula berlebihan dan konsumsi alkohol.

b. Data Oversampling

Menyamakan jumlah sampel data dari kedua kelas dengan meningkatkan data dari kelas minoritas agar setara dengan jumlah kelas mayoritas. SMOTE memilih sampel dari kelas minoritas dalam dataset. Kemudian,

untuk setiap sampel minoritas yang terpilih, SMOTE mencari tetangga terdekatnya dalam ruang fitur, umumnya menggunakan metode seperti *K-Nearest Neighbors* (KNN). SMOTE membuat sampel sintetis baru di antara sampel minoritas dan tetangga terdekatnya. Sampel-sampel sintetis ini berada pada garis yang menghubungkan kedua sampel tersebut dalam ruang fitur. Terakhir, sampel sintetis yang baru ini ditambahkan ke dataset, sehingga meningkatkan jumlah sampel di kelas minoritas hingga mencapai level yang diinginkan atau seimbang dengan kelas mayoritas. Gambaran pelabelan kelas data sebelum dan sesudah proses oversampling disajikan pada Gambar 3.



Sumber: Hasil penelitian (2025)

Gambar 3. Tampilan Data Sebelum dan Sesudah *Oversampling*

Dari Gambar 3, dapat dilihat bahwa sebelum dilakukan *oversampling* data, terdapat 989 data non-hipertensi dan 369 data hipertensi. Setelah proses *oversampling* data, kedua kelas memiliki jumlah data yang sama, yaitu 989 data untuk setiap kelas.

c. Klasifikasi Dengan Metode SVM

Proses klasifikasi diawali dengan membagi data menjadi dua, yaitu data latih dan data uji, dengan rasio 80:20. Rasio ini dipilih berdasarkan hasil uji coba yang telah dilakukan, di mana model yang dilatih

dengan rasio ini terbukti mencapai tingkat akurasi yang tinggi. Data latih, yang terdiri dari beberapa variabel dan kelas target, digunakan sebagai data sampel untuk membentuk model SVM. Sementara itu, data uji digunakan untuk mengevaluasi akurasi model klasifikasi berdasarkan *confusion matrix* yang dihasilkan. Dalam perhitungan manual, klasifikasi SVM dilakukan menggunakan kernel RBF yang didefinisikan dalam Persamaan (1) dengan parameter γ (gamma) = 1. Contoh perhitungan manual ini menggunakan 4 data yang disajikan pada Tabel 4.

Table 4. *Sample Data* Untuk Perhitungan Manual

X_1	X_2	Y
1	0	0
0	1	1
0	0	1
0	1	0

Sumber: Hasil penelitian (2025)

Yang dilakukan pertama kali dalam menentukan klasifikasi SVM adalah menghitung nilai matriks kernel K. Perhitungan nilai kernel K menggunakan dimensi $m \times m$, dan m adalah jumlah data.

$$K(x_1, x_1) = \exp\left(-1\left(\sqrt{(1-1)^2 + (0-0)^2}\right)^2\right) = \exp(0) = 1$$

$$K(x_1, x_2) = \exp\left(-1\left(\sqrt{(1-0)^2 + (0-1)^2}\right)^2\right) = \exp(-2) = 0.1353$$

$$K(x_1, x_3) = \exp\left(-1\left(\sqrt{(1-0)^2 + (0-0)^2}\right)^2\right) = \exp(-3) = 0.3679$$

$$K(x_1, x_4) = \exp\left(-1\left(\sqrt{(1-0)^2 + (0-1)^2}\right)^2\right) = \exp(-2) = 0.1353$$

Perhitungan ini dilanjutkan hingga nilai x_4, x_4 untuk menghasilkan matriks kernel K:

$$K = \begin{bmatrix} 1 & 0,1353 & 0,3679 & 0,1353 \\ 0,1353 & 1 & 0,3679 & 1 \\ 0,3679 & 0,3679 & 1 & 0,3679 \\ 0,1353 & 1 & 0,3679 & 1 \end{bmatrix}$$

Dengan menggunakan kernel K sebagai substitusi untuk hasil perkalian titik (*dot-product*) $\phi(x_i) \cdot \phi(x_j)$ dalam persamaan dualitas Lagrange pada Persamaan (2):

$$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}(\alpha_1\alpha_2 + \alpha_2^2 + 0.3679\alpha_2\alpha_3 + \alpha_3^2)$$

Dengan syarat $-\alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 = 0$ dan $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \geq 0$

Dalam fungsi objektif, suku kedua dikalikan dengan $y_1 y_j$. Persamaan tersebut telah memenuhi standar *Quadratic Programming* maka dari itu pemecah komersial dapat menyelesaikan persamaan tersebut. Dengan memanfaatkan *tool* tambahan, dapat diperoleh hasil: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$. Hasil ini membuktikan bahwa setiap data dalam merupakan vektor pendukung, karena $\alpha \neq 0$. Berikutnya, mencari nilai w (bobot) dengan menggunakan Persamaan (3) sehingga mendapatkan hasil yang disajikan pada Tabel 5.

Tabel 5. Nilai Bobot

w_i	Bobot
w_1	-1,639
w_2	2,49
w_3	2,08
w_4	-2,49

Sumber: Hasil penelitian (2025)

Persamaan (4) digunakan untuk menghitung nilai b (bias). Dari hasil perhitungan diperoleh nilai $b = -0,115$. Setelah nilai w dan b diketahui maka sebuah model SVM dapat tercipta dan diterapkan dalam proses klasifikasi dengan menggunakan Persamaan (6) sebagai berikut:

$$f(\phi(x)) = \text{sign} \left(\sum_{i,j=1}^N \alpha_i y_j K(x_i, x_j) - 0,115 \right)$$

Perhitungan untuk proses klasifikasi data pertama pada Tabel 4, dilakukan sebagai berikut:

$$f(\phi(1)) = \text{sign}((1 \times -1 \times 1 - 0,115) + (1 \times -1 \times 1,353 - 0,115) + (1 \times -1 \times 0,3679 - 0,115) + (1 \times -1 \times 0,1353 - 0,115))$$

$$f(\phi(1)) = \text{sign}(-2,11) = -1$$

Berdasarkan perhitungan tersebut, data pertama dapat diklasifikasikan sebagai kelas negatif. Untuk menghasilkan model SVM yang optimal dalam proses klasifikasi hipertensi maka proses perhitungan tersebut akan digunakan pada semua data latih. Kernel RBF digunakan untuk klasifikasi SVM dengan parameter $C = 0,01, 0,1, 1, 10$, dan

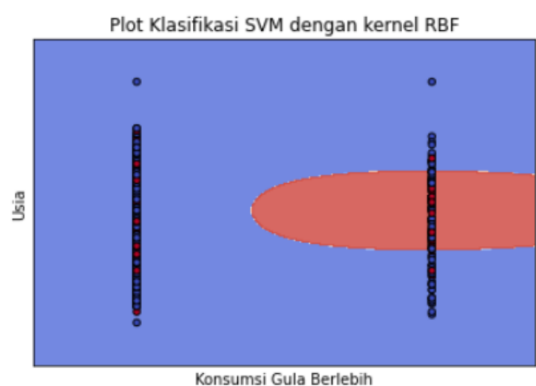
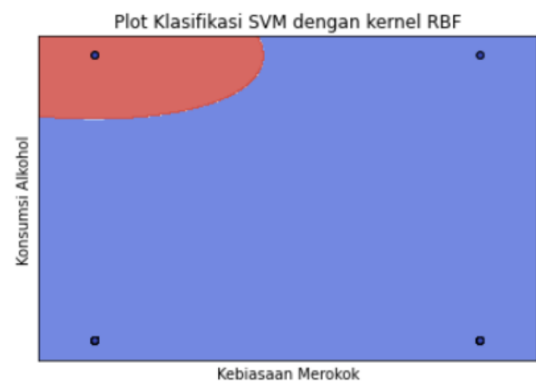
100 serta parameter γ (gamma) = 0,01, 0,1, 1, 10, dan 100. Percobaan dilakukan hanya satu kali menggunakan 25 parameter kernel RBF. Dengan terbentuknya model SVM menggunakan data latih, maka nilai akurasi model didapat dengan menggunakan data uji parameter kernel RBF seperti pada Tabel 6.

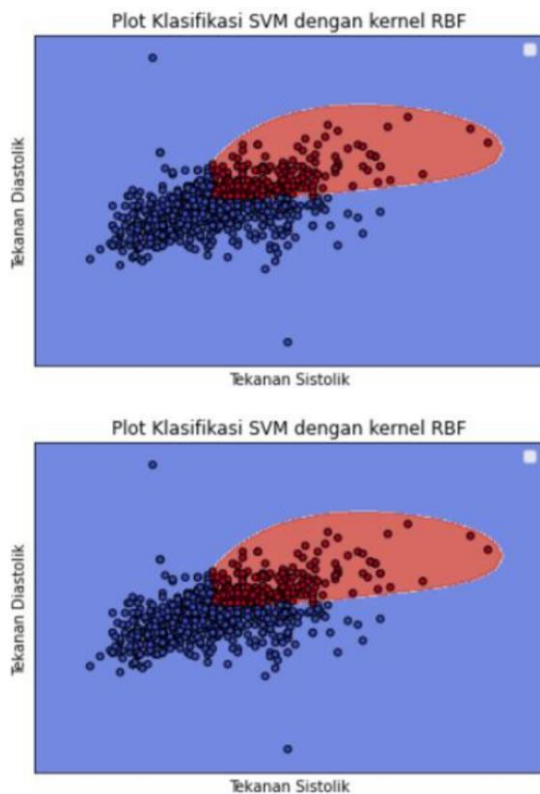
Tabel 6. Akurasi Parameter Model Kernel RBF SVM

Parameter	Akurasi			
	$\gamma = 0,01$	$\gamma = 0,1$	$\gamma = 1$	$\gamma = 10$
$C = 0,01$	75,75%	90,25%	70,25%	50,25%
$C = 0,1$	91,01%	92,75%	90,75%	70,75%
$C = 1$	93,75%	94,75%	96,25%	90,01%
$C = 10$	95,01%	96,25%	97,00%	91,25%
$C = 100$	95,75%	97,75%	97,25%	91,25%

Sumber: Hasil penelitian (2025)

Dari Tabel 5 dan Tabel 6 dapat diketahui bahwa kernel RBF dengan parameter $C = 100$ dan γ (gamma) = 0,1 merupakan akurasi tertinggi yang bernilai 97,74%, dan jumlah vektor pendukung dari model ini adalah 133 titik. Model klasifikasi SVM yang terbentuk dapat divisualisasikan dalam bentuk sebuah plot. Plot tersebut dibentuk oleh kombinasi dua variabel independen. Berdasarkan kombinasi dua variabel independen tersebut, secara keseluruhan didapatkan plot-plot seperti yang ditunjukkan pada Gambar 4.





Sumber: Hasil penelitian (2025)

Gambar 4. Plot Klasifikasi SVM Kernel RBF dengan $\gamma = 0,1$ dan $C = 100$

d. Pengujian

Akurasi hasil model klasifikasi dapat ditentukan menggunakan K-Fold Cross Validation dengan nilai k=5. Setiap bagian data akan digunakan sebagai data uji satu kali, sementara data lainnya digunakan sebagai data latih. Dengan demikian, model klasifikasi akan dilatih 5 kali dan diuji 5 kali. Berikut adalah hasil akurasi, presisi, dan recall yang diperoleh menggunakan 5-Fold Cross Validation, seperti pada Tabel 7.

Tabel 7. Hasil Pengujian menggunakan 5-Fold Cross Validation

Paramater k	Presisi	Recall	Akurasi
k = 1	0,96099	0,93901	0,97001
k = 2	0,97111	0,96012	0,89981
k = 3	0,96991	0,94599	0,97751
k = 4	1,00000	0,93658	0,94502
k = 5	1,00000	0,96499	0,96503
Rerata	0,98018	0,94929	0,95144

Sumber: Hasil penelitian (2025)

Dari Tabel 7 dapat diketahui bahwa rerata presisi adalah 0,98018, rerata recall sebesar 0,94929, dan rerata akurasinya 0,95144. Dari

hasil tersebut dapat disimpulkan bahwa kinerja model SVM dengan kernel RBF dalam mengklasifikasikan data termasuk kategori baik.

e. Analisis Kinerja Model Klasifikasi

Untuk menghitung akurasi klasifikasi diagnosis penyakit hipertensi dengan metode SVM dan kernel RBF menggunakan parameter $C = 100$ dan $\gamma = 0,1$ dapat dianalisis dengan membuat matriks konfusi pada Tabel 8.

Tabel 8. Hasil Confusion Matrix

Data	Aktual	
	Non hipertensi	Hipertensi
prediksi Non hipertensi	196	7
prediksi Hipertensi	2	203

Sumber: Hasil penelitian (2025)

Dari Tabel 8 matriks konfusi, diketahui bahwa 196 pasien non hipertensi dapat diprediksi dengan benar sebagai non-hipertensi, 7 pasien hipertensi diprediksi salah sebagai non-hipertensi, 203 pasien hipertensi berhasil diprediksi secara tepat sebagai hipertensi, sedangkan 2 pasien non hipertensi diprediksi salah sebagai hipertensi. Nilai akurasi, presisi, recall, dan F1-Score dihitung sebagai berikut:

$$Akurasi = \frac{196 + 203}{196 + 203 + 7 + 2} \times 100\% = 97,8\%$$

$$Presisi = \frac{196}{196 + 7} \times 100\% = 96,5\%$$

$$Recall = \frac{196}{196 + 2} \times 100\% = 98,9\%$$

$$F1\ Score = 2 \times \frac{96,5\% \times 98,9\%}{96,5\% + 98,9\%} \times 100\% = 97,7\%$$

Berdasarkan hasil perhitungan yang telah dilakukan, dapat disimpulkan bahwa akurasi untuk klasifikasi diagnosis hipertensi menggunakan SVM adalah 97,7%. Hal tersebut dapat sebagai bukti bahwa metode SVM berhasil mengklasifikasikan diagnosis hipertensi dengan sangat baik. Nilai presisi yang dihasilkan adalah 96,5%, artinya dari 201 pasien yang diprediksi oleh metode SVM sebagai non hipertensi, 196 pasien diantaranya memang benar sebagai non hipertensi. Nilai recall sebesar 98,9%, berarti dari 198 pasien non hipertensi, 196 pasien berhasil diidentifikasi dengan tepat oleh metode SVM. Nilai F1-Score yang dihasilkan adalah 97,7%, yang menunjukkan bahwa metode SVM mampu

menyeimbangkan presisi dan *recall* dengan baik.

IV. KESIMPULAN

Hasil klasifikasi metode SVM untuk diagnosis klasifikasi hipertensi menunjukkan bahwa model terbaik menggunakan kernel RBF. Nilai parameter yang digunakan adalah $C = 100$ dan γ (gamma) = 0,1. Analisis hasil klasifikasi menggunakan matriks *confusion* membuktikan bahwa 196 pasien non hipertensi diprediksi secara tepat sebagai non hipertensi, 7 pasien hipertensi salah diprediksi sebagai non hipertensi, 203 pasien hipertensi diprediksi sesuai sebagai hipertensi, dan 2 pasien non hipertensi salah diprediksi sebagai hipertensi. Tingkat akurasi yang dihasilkan dari diagnosis klasifikasi hipertensi menggunakan metode SVM kernel RBF dengan parameter $C = 100$ dan $\gamma=0,1$ adalah 97,7%. Nilai tersebut menunjukkan bahwa metode SVM mampu memberikan klasifikasi diagnosis hipertensi dengan sangat baik. Berdasarkan analisis dan perhatiungan dari hasil penelitian ini, dapat dinyatakan bahwa metode SVM dapat menjadi alternatif metode klasifikasi yang digunakan untuk menghasilkan klasifikasi diagnosis hipertensi.

V. REFERENSI

- Amriana, A., Ilham, A. A., Achmad, A., & Yusran, Y. (2025). Optimization of Herbal Plant Classification Using Hybrid Method Particle Swarm Optimization with Support Vector Machine. *International Journal on Informatics Visualization*, 9(1), 396–406. <https://doi.org/10.62527/joiv.9.1.2576>
- Ardiansyah, M. Z., & Widowati, E. (2024). Hubungan Kebisingan dan Karakteristik Individu dengan Kejadian Hipertensi pada Pekerja Rigid Packaging. *HIGEIA (Journal of Public Health Research and Development)*, 8(1), 141–151. <https://doi.org/10.15294/higeia.v8i1.75362>
- Benaired, N., Meghraoui, M. H., & Benselama, Z. A. (2024). Traitement du Signal Hypertension Management via Photoplethysmography: An Ensemble Learning-Based Approach for Classification of Blood Pressure Using Fourier Synchrosqueezed Transform. *Traitement Du Signal*, 41(5), 2263–2278. <https://doi.org/https://doi.org/10.18280/ts.410504>
- Biswas, A., & Islam, M. S. (2023). A Hybrid Deep CNN-SVM Approach for Brain Tumor Classification. *Journal of Information Systems Engineering and Business Intelligence*, 9(1), 1–15. <https://doi.org/10.20473/jisebi.9.1.1-15>
- Dohan, M., Mohammed, R. B., Gwad, W. H., Khalaf, M., & Othman, K. M. Z. (2024). Predicting Vehicle Driver Preference from the Analysis of In-Vehicle Coupon Recommendation Data. *Journal of Soft Computing and Data Mining*, 5(2), 274–282. <https://doi.org/10.30880/jscdm.2024.05.02.020>
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information (Switzerland)*, 15(4), 2–36. <https://doi.org/10.3390/info15040235>
- Jumanto, Rofik, Sugiharti, E., Alamsyah, Arifudin, R., Prasetyo, B., & Muslim, M. A. (2024). Optimizing Support Vector Machine Performance for Parkinson's Disease Diagnosis Using GridSearchCV and PCA-Based Feature Extraction. *Journal of Information Systems Engineering and Business Intelligence*, 10(1), 38–50. <https://doi.org/10.20473/jisebi.10.1.38-50>
- Kreutz, R., Brunström, M., Burnier, M., Grassi, G., Januszewicz, A., Muiesan, M. L., Tsioufis, K., de Pinho, R. M., Albin, F. L., Boivin, J. M., Doumas, M., Nemcsik, J., Rodilla, E., Agabiti-Rosei, E., Algharably, E. A. E., Agnelli, G., Benetos, A., Hitij, J. B., Cifková, R., ... Mancía, G. (2024). 2024 European Society of Hypertension clinical practice guidelines for the management of arterial hypertension. *European Journal of Internal Medicine*, 126(May), 1–15. <https://doi.org/10.1016/j.ejim.2024.05.033>
- Kurniawan, R., Utomo, B., Siregar, K. N., Ramli, K., Besral, Suhatri, R. J., & Pratiwi, O. A. (2023). Hypertension prediction using machine learning algorithm among Indonesian adults. *IAES International Journal of Artificial Intelligence*, 12(2), 776–784. <https://doi.org/10.11591/ijai.v12.i2.pp776-784>
- Li, J., Bi, J., Yang, S., Wang, S., Yang, S., Chen, S., Han, K., Luo, S., Jiang, Q., Liu, M., & He, Y. (2024). Analysis of Related Factors Influencing Hypertension Classification among Centenarians in Hainan, China. *Reviews in Cardiovascular Medicine*, 25(7), 1–10. <https://doi.org/10.31083/j.rcm2507235>
- Munali, Y., & Armansyah. (2024). Classification of Hypertension Using Naïve Bayes Method

- with Data Discretization Approach Risk Factors. *Jurnal Sistem Cerdas*, 7(1), 1–12. <https://doi.org/10.37396/jsc.v7i1.381>
- Nurhadiva, S. S. (2024). Blood Pressure and Heart Rate Measurement for Hypertension Classification Using the K-Nearest Neighbors Method Based on IoT. *Piksel*, 12(225), 373–382. <https://doi.org/10.33558/piksel.v12i2.9824>
- Nurrani, H., Andi Kurniawan Nugroho, & Sri Heranurweni. (2023). Image Classification of Vegetable Quality using Support Vector Machine based on Convolutional Neural Network. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 7(1), 168–178. <https://doi.org/10.29207/resti.v7i1.4715>
- Ratwate, S., & Celermajer, D. S. (2024). The latest definition and classification of pulmonary hypertension. *International Journal of Cardiology Congenital Heart Disease*, 17(July), 100534. <https://doi.org/10.1016/j.ijcchd.2024.100534>
- Resky, A. A. C., Lapendy, J. C., Risal, A. A. N., Surianto, D. F., & Wahid, A. (2025). PCA and t-SNE Implementation for KNN Hypertension Classification. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 9(1), 175–184. <https://doi.org/https://doi.org/10.29207/resti.v9i1.6208>
- Rochman, E. M. S., Indriawati, N., Suzanti, I. O., Setiawan, W., Husni, Ma'arof, M. I. N., & Rachmad, A. (2024). Classification of hypertension disease using Artificial Neural Network (ANN) backpropagation method case study in mitigating health risk: UPT Modopuro Mojokerto Health Center. *BIO Web of Conferences*, 146, 1–8. <https://doi.org/10.1051/bioconf/202414601083>
- Setyadi, H. A., Supriyanta, Nurohim, G. S., Widodo, P., & Sutanto, Y. (2024). Knowledge-Based Intelligent System for Diagnosing Three-Wheeled Motorcycle Engine Faults. *International Journal on Informatics Visualization*, 8(4), 2472–2478. <https://doi.org/10.62527/joiv.8.4.2487>
- Sutanto, Y., Setyadi, H. A., Nugroho, W., & Amin, B. Al. (2025). Extreme Learning Machine Method Application to Forecasting Coffee Beverage Sales. *Journal of Applied Informatics and Computing (JAIC)*, 9(5), 2461–2467. <https://doi.org/https://doi.org/10.30871/jaic.v9i5.10465>
- Tri Sutanti Puji Hartati, & Emyr Reisha Isaura. (2023). Three Body Mass Index Classification Comparison In Predicting Hypertension Among Middle-Aged Indonesians. *Media Gizi Indonesia*, 18(1), 38–48. <https://doi.org/10.20473/mgi.v18i1.38-48>
- Xu, C., Li, M., Meng, W., Han, J., Zhao, S., Tang, J., Yang, H., Maimaitiaili, R., Teliewubai, J., Yu, S., Chi, C., Fan, X., Xiong, J., Zhao, Y., Xu, Y., & Zhang, Y. (2023). Etiological Diagnosis and Personalized Therapy for Hypertension: A Hypothesis of the REASOH Classification. *Journal of Personalized Medicine*, 13(2). <https://doi.org/10.3390/jpm13020261>