



## **Literatur Review: Pendekatan *Hybrid SVM* Dan *KNN* Untuk Klasifikasi Penyakit Tiroid**

**Daksa Nizam Adila<sup>1</sup>, Muhammad Mustaqim<sup>2</sup>, Muhammad Bara Aksayeth<sup>3</sup>, Dicky Fahmi<sup>4</sup>, Perani Rosyani<sup>5\*</sup>**

<sup>1,2,3,4,5</sup>Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

Email: <sup>1</sup>[khml71024@gmail.com](mailto:khml71024@gmail.com), <sup>2</sup>[Muhhammadmustaqim471@gmail.com](mailto:Muhhammadmustaqim471@gmail.com), <sup>3</sup>[aksayeth.bara@gmail.com](mailto:aksayeth.bara@gmail.com), <sup>4</sup>[dickyfahmi08111999@gmail.com](mailto:dickyfahmi08111999@gmail.com), <sup>5</sup>[dosen00837@unpam.ac.id](mailto:dosen00837@unpam.ac.id)

(\* : coressponding author)

**Abstrak** - Klasifikasi penyakit tiroid merupakan aspek penting dalam diagnosis medis, mengingat dampaknya yang signifikan terhadap kesehatan pasien. Penelitian ini membahas pendekatan hybrid yang menggabungkan algoritma SVM dan KNN untuk meningkatkan akurasi klasifikasi penyakit tiroid. SVM dikenal efektif dalam membangun hyperplane yang memisahkan kelas secara optimal, sementara KNN memperkuat kemampuan analisis pada data lokal di sekitar garis keputusan. Studi ini bertujuan mengevaluasi efektivitas pendekatan hybrid SVM-KNN melalui analisis literatur yang relevan. Artikel-artikel ilmiah yang diterbitkan dalam periode tertentu dikaji untuk mengidentifikasi keunggulan serta tantangan dalam penerapan metode ini. Hasil kajian menunjukkan bahwa pendekatan hybrid mampu meningkatkan akurasi dan ketahanan model terhadap dataset yang tidak seimbang. Hal ini memperkaya kemampuan model dalam menghasilkan diagnosis yang lebih andal. Namun, terdapat beberapa kendala, seperti pemilihan parameter optimal dan kebutuhan komputasi yang lebih tinggi. Untuk mengatasi tantangan ini, disarankan peningkatan kualitas data serta pengoptimalan algoritma. Evaluasi berkala juga diperlukan untuk memastikan sistem tetap relevan dan dapat diandalkan dalam mendukung diagnosis penyakit tiroid.

**Kata Kunci:** Tiroid, SVM, KNN

**Abstract** - Classification of thyroid diseases is an important aspect in medical diagnosis, considering its significant impact on patient health. This study discusses a hybrid approach that combines SVM and KNN algorithms to improve the accuracy of thyroid disease classification. SVM is known to be effective in constructing hyperplanes that optimally separate classes, while KNN enhances the analysis capability on local data around the decision line. This study aims to evaluate the effectiveness of the SVM-KNN hybrid approach through an analysis of relevant literature. Scientific articles published in a certain period were reviewed to identify the advantages and challenges in implementing this method. The results of the study indicate that the hybrid approach is able to improve the accuracy and robustness of the model to imbalanced datasets. This enriches the model's ability to produce more reliable diagnoses. However, there are several obstacles, such as optimal parameter selection and higher computational requirements. To overcome these challenges, it is recommended to improve data quality and optimize the algorithm. Periodic evaluation is also needed to ensure the system remains relevant and reliable in supporting the diagnosis of thyroid diseases.

**Keywords:** Thyroid, SVM, KNN

### **1. PENDAHULUAN**

Penyakit tiroid merupakan salah satu gangguan kesehatan yang memengaruhi fungsi metabolisme tubuh. Diagnosis dini penyakit ini sangat penting untuk mencegah komplikasi serius. Dalam bidang medis, data diagnosis sering kali memiliki karakteristik non-linear dan kompleks, sehingga memerlukan metode klasifikasi yang mampu menangani berbagai pola data dengan akurasi tinggi. Dalam hal ini, Penelitian semakin berkonsentrasi pada pendekatan *hybrid SVM* dan *KNN* untuk klasifikasi penyakit tiroid.

Penyakit tiroid merupakan salah satu masalah kesehatan global yang signifikan, dengan prevalensi yang terus meningkat setiap tahunnya. Penyakit ini mencakup berbagai gangguan seperti hipotiroidisme, hipertiroidisme, dan tiroiditis, yang dapat berdampak serius pada kesehatan metabolik, kardiovaskular, hingga mental pasien. Diagnosis yang cepat dan akurat sangat penting untuk memastikan pengobatan yang tepat dan mencegah komplikasi lebih lanjut. Namun, diagnosis penyakit tiroid sering menghadapi tantangan karena kompleksitas data medis yang melibatkan berbagai parameter seperti kadar hormon, riwayat medis, dan faktor genetik.



Salah satu metode yang telah digunakan dalam pendekatan *hybrid SVM* dan *KNN* untuk klasifikasi penyakit tiroid dengan K-Nearest Neighbor atau KNN adalah metode algoritma supervised learning yang digunakan pada tugas pengklasifikasian. Pengklasifikasian dilakukan hanya berdasarkan pada memori berdasarkan training sample. Hasil instance query baru didapat dari kelas yang paling sering muncul. Algoritma ini mengklasifikasikan data berdasarkan kedekatan jaraknya pada titik-titik sejumlah-k disekitarnya (Sunge, (2024). Dan juga dengan Support Vector Machine (SVM) merupakan algoritma dari pembelajaran mesin atau machine learning yang diterapkan pada regresi dan klasifikasi. SVM berdimensi tinggi memiliki kemampuan generalisasi yang baik pada data yang bersifat linear. Sebaliknya pada data yang bersifat non-linear kemampuannya kurang baik sehingga dibutuhkan kernel. Kernel memungkinkan pemisahan data secara linear dan berfungsi untuk mentransformasi data sehingga ruang dimensinya menjadi tinggi. SVM dengan ruang kernel menjadi lebih unggul karena dapat mengklasifikasikan model hanya dengan menggunakan data yang terpilih saja (F. Liantoni and A. Santoso, 2020).

Melalui tinjauan literatur ini, kami bertujuan untuk mengevaluasi pendekatan hybrid antara Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) dalam klasifikasi penyakit tiroid. Penyakit tiroid merupakan masalah kesehatan global yang memengaruhi jutaan orang, dan deteksi dini sangat penting untuk mencegah komplikasi lebih lanjut. Meskipun banyak metode telah diterapkan untuk klasifikasi ini, keterbatasan akurasi dan efisiensi dalam skenario tertentu masih menjadi tantangan utama. SVM dikenal dengan keunggulannya dalam menangani data berdimensi tinggi dan memberikan keputusan klasifikasi yang baik, terutama di dekat batas antar kelas. Namun, dalam kasus tertentu, terutama di sekitar data yang kompleks atau noisy, performa SVM bisa menurun. Di sisi lain, KNN adalah metode non-parametrik yang sederhana dan efektif, terutama dalam memanfaatkan pola lokal untuk meningkatkan akurasi. Pendekatan hybrid SVM-KNN mencoba menggabungkan keunggulan kedua metode ini, di mana SVM digunakan untuk membentuk hyperplane awal, sementara KNN meningkatkan akurasi di area batas antar kelas dengan memanfaatkan tetangga terdekat (Liliya A. Demidova, 2021). Oleh karena itu, diharapkan penelitian ini dapat memberikan kontribusi yang signifikan terhadap terciptanya model klasifikasi yang andal dan robust dalam mendukung diagnosis penyakit tiroid, serta memperluas aplikasi machine learning di bidang kesehatan.

## **2. METODE**

### **2.1 Studi Literatur Review**

Literature review adalah Literature review adalah metode penelitian yang bertujuan untuk mengidentifikasi, mengevaluasi, dan menginterpretasi seluruh hasil penelitian yang relevan dengan pertanyaan penelitian, topik tertentu, atau fenomena yang menjadi fokus perhatian. (Kitchenham, 2004). Studi individu (individual study) merupakan jenis studi primer (primary study), sedangkan literature review tergolong sebagai studi sekunder (secondary study). Literature review sangat berguna untuk menyintesis berbagai hasil penelitian yang relevan, sehingga informasi yang disampaikan kepada pembuat kebijakan menjadi lebih lengkap dan seimbang.

### **2.2 Clustering**

Clustering, atau pengelompokan, adalah salah satu metode dalam data mining yang digunakan untuk mengelompokkan sejumlah besar data berdasarkan kemiripan antar nilai yang bersifat homogen. Proses ini melibatkan pengamatan terhadap objek atau data untuk menentukan cluster yang sesuai. Dalam praktiknya, proses clustering biasanya didukung oleh penerapan algoritma tertentu bertujuan untuk lebih memperinci dan target dalam pengelempokan data jauh lebih terstruktur dalam sistem matematis (Perani Rosyani et al., 2023).

### **2.3 Algoritma K-NN**

K-Nearest Neighbor (KNN) merupakan salah satu teknik klasifikasi yang sering digunakan dalam berbagai penelitian. Banyak studi sebelumnya telah memanfaatkan algoritma ini untuk berbagai keperluan, seperti pengenalan pola, pengenalan teks, hingga pemrosesan objek. Kelebihan



utama KNN terletak pada kesederhanaannya, terutama dalam menangani data pelatihan dan pengujian, bahkan dalam jumlah yang sangat besar. (Murni Handayani, 2023)

Algoritma K-Nearest Neighbor adalah teknik penambahan datad dan pembelajaran mesin yang digunakan untuk melakukan klasifikasi dan regresi. Metode ini bekerja dengan mencari nilai terdekat dengan titik data baru berdasarkan jarak Euclidean atau jarak lain dalam ruang fitur. Titik data baru kemudian diklasifikasikan ke dalam salah satu kelas atau prediksi regresi dibuat berdasarkan sebagian besar nilai tersebut. Kelebihan metode K-Nearest Neighbor antara lain pelatihan yang sangat cepat, mudah, dan efektif pada dataset yang besar. Namun metode K-nearest neighbour memiliki beberapa kelemahan: Kebutuhan untuk menentukan nilai k, kompleksitas komputasi, keterbatasan memori, dan kemungkinan pengaruh atributasing. Klasifikasi yang didasarkan pada sistem informasi merupakan teknik memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas yang sudah didefinisikan sebelumnya.

#### **2.4 SVM**

Klasifikasi adalah proses untuk menemukan fungsi atau model yang dapat menjelaskan perbedaan antar kelas dalam suatu data. Proses ini juga digunakan untuk membedakan metode atau konsep tertentu, dengan tujuan memprediksi kelas dari objek yang labelnya belum diketahui. (Agus Heri Yunial, 2020) Algoritma Support Vector Machine (SVM) diterapkan pada data nonlinear dengan memanfaatkan pendekatan kernel. Data berdimensi rendah awalnya dipetakan ke dimensi yang lebih tinggi menggunakan fungsi kernel. Dalam proses klasifikasinya, SVM memanfaatkan data historis berdasarkan skenario yang telah ditentukan untuk membagi data menjadi dua kategori: diterima dan tidak diterima. Oleh karena itu, data dikelompokkan menjadi data training dan data testing. SVM bertujuan menemukan hyperplane (garis pemisah) terbaik di antara berbagai kemungkinan untuk memisahkan kelas-kelas tersebut. Hyperplane optimal ditentukan berdasarkan margin terbesar, yaitu jarak maksimum antara titik-titik terdekat dari masing-masing kelas dengan hyperplane. (Agung Budi Susanto et al., 2023)

#### **2.5 Data Sekunder**

Data sekunder merupakan informasi yang diambil dari sumber yang sudah ada, seperti jurnal-jurnal ilmiah yang membahas metode SVM, KNN, serta pendekatan hybrid untuk klasifikasi penyakit. Data ini tidak diperoleh langsung oleh peneliti melainkan bersumber dari penelitian sebelumnya. Jurnal-jurnal tersebut diakses melalui <https://scholar.google.com/>, Semantic Scholar: dan dan digunakan untuk mendukung analisis serta landasan teori dalam penelitian ini.

### **3. PEMBAHASAN**

<b>NO</b>	<b>Nama Peneliti dan Tahun</b>	<b>Metode yang dibahas</b>	<b>Tujuan Penelitian</b>	<b>Hasil yang didapat</b>
1	Perani Rosyani et al., 2023	Klasifikasi SVM	Tujuan utama klasifikasi adalah menentukan kelas tertentu untuk setiap instance, sehingga memungkinkan kita memahami bagaimana data dapat digunakan untuk memprediksi perilaku pada kasus baru. Dalam konteks data mining, klasifikasi digunakan untuk menyelesaikan	Klasifikasi ini berfungsi untuk mengelompokkan data berdasarkan tingkat efektivitas dan efisiensi, serta membangun model klasifikasi menggunakan metode Support Vector Machine (SVM). Proses ini dilakukan dengan bantuan kernel polynomial pada MATLAB, yang



			<p>berbagai masalah dan mengelompokkan data. Algoritma SVM menggunakan fungsi atau hyperplane sebagai pemisah antara dua kelas pola, dengan fokus menemukan hyperplane optimal yang memaksimalkan jarak antara kedua kelas tersebut. Secara umum, SVM merupakan metode klasifikasi biner. Namun, untuk kasus multiclass, SVM dapat diterapkan dengan membagi dataset ke dalam beberapa klasifikasi biner.</p>	<p>memungkinkan pemetaan data ke dimensi yang lebih tinggi untuk meningkatkan akurasi dalam memisahkan kelas-kelas yang berbeda.</p>
2	Rina Silviany Tantika. (2022)	Klasifikasi SVM	<p>Tujuannya adalah Dalam data mining, klasifikasi digunakan untuk menyelesaikan berbagai permasalahan dengan memisahkan data ke dalam kategori tertentu. SVM (Support Vector Machine) melakukan klasifikasi dengan menggunakan hyperplane, yaitu fungsi yang memisahkan dua kelas pola. Algoritma ini berfokus pada pencarian hyperplane yang optimal, yaitu hyperplane yang memaksimalkan jarak antara kelas-kelas tersebut. SVM pada dasarnya dirancang untuk</p>	<p>didapatkan dari Metode SVM multiclass dengan pendekatan One Against One digunakan untuk memprediksi kelas tiroid pasien, yaitu apakah pasien berada dalam kondisi normal (euthyroidism), mengalami hypothyroidism, atau hyperthyroidism. Pendekatan ini memanfaatkan kernel linier untuk memisahkan data berdasarkan pola yang terdapat dalam setiap kategori.</p>



			klasifikasi biner, namun dapat diterapkan untuk klasifikasi multiclass dengan membagi dataset menjadi beberapa klasifikasi biner berdasarkan kelas-kelas yang ada.	
3	Supardianto et al., 2022	SVM dan KNN	<p>Tujuannya adalah Untuk menangani dataset yang mengandung missing values, salah satu upaya yang dilakukan adalah dengan menerapkan algoritma K-Nearest Neighbours Imputation (KNNImputer). Algoritma ini sangat efektif dalam mengisi nilai yang hilang pada dataset, dengan cara menghitung nilai rata-rata dari tetangga (neighbors) yang memiliki kemiripan paling dekat. Dengan menggunakan KNNImputer, data yang hilang dapat diimputasi secara akurat, yang pada gilirannya dapat meningkatkan performa model SVM dalam melakukan pemodelan data. Hal ini memungkinkan pembuatan basis pengetahuan yang lebih akurat untuk memprediksi penyakit tiroid.</p>	<p>Hasilnya yang didapati Dalam pengembangan model pengetahuan menggunakan machine learning, kualitas data yang baik dan terbebas dari nilai missing sangat mempengaruhi kinerja model. Pada penelitian ini, data yang digunakan mengandung nilai missing yang cukup tinggi, sehingga penerapan algoritma machine learning seperti Support Vector Machine (SVM) tidak dapat menghasilkan model yang optimal. Untuk menangani masalah tersebut, langkah yang diambil adalah dengan menerapkan algoritma K-Nearest Neighbours Imputation (KNNImputer), yang terbukti sangat efektif dalam menangani data missing values. Algoritma ini membantu mengimputasi nilai yang hilang dengan cara menghitung nilai berdasarkan rata-rata dari tetangga terdekat, sehingga memungkinkan penggunaan data</p>



				yang lebih lengkap dan meningkatkan kinerja model.
4	Angel, Dyah Erny Herwindiati., 2024	SVM dan KNN	membandingkan algoritma machine learning dalam melakukan klasifikasi kondisi thyroid berdasarkan ciri-ciri, gaya hidup, dan kondisi kesehatan pasien. Ada tiga buah algoritma machine learning yang akan dibandingkan yaitu Support Vector Machine (SVM), K-NN	SVM yang paling baik untuk melakukan klasifikasi tersebut. Agar dapat mengetahui metode yang lebih akurat, penelitian selanjutnya dapat menggunakan metode Random Forest dan Logistic Regression.
5	Balasree et al., 2022	Penelitian ini menggunakan berbagai metode machine learning seperti Decision Tree, Support Vector Machine (SVM), Naïve Bayes, dan k-Nearest Neighbors (k-NN) untuk klasifikasi penyakit tiroid.	Penelitian bertujuan mengembangkan model prediksi yang akurat untuk mendiagnosis penyakit tiroid dengan menggunakan algoritma machine learning.	SVM dan k-NN juga memberikan hasil akurasi tinggi, masing-masing 95,3% dan 94,2%.

#### **4. KESIMPULAN**

Penerapan metode hybrid SVM dan KNN dalam klasifikasi penyakit tiroid terbukti mampu meningkatkan akurasi dan efisiensi sistem klasifikasi. Pendekatan ini memanfaatkan keunggulan masing-masing algoritma, di mana SVM unggul dalam membangun model pemisahan data secara global, sedangkan KNN memperkuat analisis lokal pada area yang sulit. Meski ada tantangan dalam pengoptimalan parameter kedua algoritma, kombinasi ini secara signifikan meningkatkan kinerja model dalam mendeteksi penyakit tiroid. Sistem berbasis hybrid ini dapat mendukung tenaga medis dalam diagnosis yang lebih cepat dan akurat, memberikan kontribusi penting pada bidang kesehatan berbasis teknologi.

#### **DAFTAR PUSTAKA**

- Adzan Zuhri Seknun, Aditya Kusuma, Alivia Sabrina, Amanda Dwi Cahyani Putri, Muhammad Raehan, Perani Rosyani. (2023). Klasifikasi Tingkat Kematangan Buah Tomat dengan Variasi Model Warna Menggunakan Support Vector Machine. Logic: Jurnal Ilmu Komputer dan Pendidikan.
- Ahzril Pria Adisty, Novara Luthfyani, Perahim Tara, Richky Adriyan, Perani Rosyani. (2023) Klasterisasi Menggunakan Algoritma K-Means Clustering Untuk Memprediksi Kelulusan



- Mata Kuliah Mahasiswa. OKTAL: Jurnal Ilmu Komputer dan Sains
- Angel, Erny Herwindiati Dyah. (2024) Perbandingan Algoritma K-NN, SVM, dan Decision Tree dalam Klasifikasi Kelenjar Tiroid. IFTK
- Azis, A., Zy, A. T., & Sunge, A. S. (2024). Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor. Jurnal Teknologi Dan Sistem Informasi Bisnis, 6(1), 117-124
- Christanto Duta Arief, Susanto Agung Budi, Handayani Murni. (2023) Analisis Sistem Pembelajaran Menggunakan Google Classroom Untuk Meningkatkan Kompetensi Siswa Dengan Metode K-Nearest Neighbor (Studi Kasus: SMK Mawar Saron Taman Royal Kota Tangerang). Jurnal Penelitian Ilmu Komputer
- Dharmarajan K., Balasree K., Arunachalam A.S., Abirmai. K. (2020) Thyroid Disease Classification Using Decision Tree and SVM. IJPHRD
- F. Liantoni and A. Santoso, "Perbaikan Kontras Citra Mammogram Pada Klasifikasi Kanker Payudara Berdasarkan Fitur Gray-Level-Co-Occurrence Matrix," Science And Information Technology Journal, no. 3, pp. 26- 51, 2020.
- Kitchenham a Barbara, O Pearl Brereto, et.al. Systematic Literatur Reviews in Software Eigneering- ASiystematic Literatur Review. Article Departemen Of Computer Science, Durham University, Durham, Uk, (2008): 793.
- Liliya A. Demidova. (2021) Two-Stage Hybrid Data Classifiers Based on SVM and kNN Algorithms. *Symmetry*.
- Nazura Is Anin, Susanto Agung Budi, Sudarno. (2023). Analisis Siswa Paket C PKBM Siliwangi Pamulang Diterima Masuk Perguruan Tinggi Negeri Menggunakan Metode Algoritma *Support Vector Machine* dan *K-Nearest Neighbor* (Studi Kasus: Sekolah PKBM Siliwangi). Jurnal Penelitian Ilmu Komputer
- Tantika, Rina Silviany. (2022) Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid. Bandung Conference Series: Statistics
- Supardianto, Mutawali Lalu, Murniati Wafiah. (2022) Penerapan KNNImputer dalam Mengolah Missing Value untuk Meningkatkan Akurasi SVM dalam Klasifikasi Penyakit Tiroid. JINTEKS (Jurnal Informatika Teknologi dan Sains).
- Yunial Agus Heri. (2020). Analisis Optimasi Algoritma Klasifikasi Support Vector Machine, Decision Trees, dan Neural Network Menggunakan Adaboost dan Bagging. Jurnal Informatika Universitas Pamulang