

# Literature Review: Penggunaan *Machine Learning* Berbasis SVM untuk Klasifikasi Penyakit Liver

Amir Rozy Sinaga<sup>1\*</sup>, Raihan Rizky Nurfadila<sup>1</sup>, Melanie Putria<sup>1</sup>, Silvi Maerosa<sup>1</sup>

<sup>1</sup>Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipstek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: <sup>1\*</sup>[Amirrozy0302@gmail.com](mailto:Amirrozy0302@gmail.com), <sup>2</sup>[sican180@gmail.com](mailto:sican180@gmail.com), <sup>3</sup>[mputria47@gmail.com](mailto:mputria47@gmail.com),

<sup>4</sup>[Silvimrs08@gmail.com](mailto:Silvimrs08@gmail.com)

(\* : coressponding author)

**Abstrak**– Penyakit liver merupakan masalah kesehatan yang signifikan di seluruh dunia, dengan angka kematian yang tinggi jika tidak didiagnosis sejak dini. Dalam penelitian ini, kami melakukan kajian pustaka tentang penerapan Support Vector Machine (SVM) dalam klasifikasi penyakit hati. Metode SVM dipilih karena kemampuannya dalam menangani data yang tidak seimbang dan kompleksitas yang tinggi. Tujuan dari penelitian ini adalah untuk mengevaluasi efektivitas SVM dibandingkan dengan metode machine learning lainnya dalam mendeteksi penyakit liver. Hasil kajian pustaka menunjukkan bahwa SVM memberikan akurasi yang lebih tinggi dibandingkan dengan metode lain seperti Naïve Bayes dan Decision Tree, dengan beberapa penelitian mencapai akurasi lebih dari 90%. Penelitian ini diharapkan dapat memberikan wawasan tentang pengembangan sistem diagnostik berbasis machine learning untuk penyakit liver.

**Kata Kunci:** Sistem Klasifikasi Penyakit Liver, Algoritma SVM, Kajian Literatur, *Machine Learning*

**Abstract**– *Liver disease is a significant health problem worldwide, with high mortality rates if not diagnosed early. In this study, we conducted a literature review on the application of Support Vector Machine (SVM) in liver disease classification. The SVM method was chosen because of its ability to handle imbalanced data and high complexity. The purpose of this study was to evaluate the effectiveness of SVM compared to other machine learning methods in detecting liver disease. The results of the literature review showed that SVM provided higher accuracy compared to other methods such as Naïve Bayes and Decision Tree, with some studies achieving accuracy of more than 90%. This study is expected to provide insight into the development of a machine learning-based diagnostic system for liver disease.*

**Keywords:** *Liver Disease Classification System, SVM Algorithm, Literature Review, Machine Learning*

## 1. PENDAHULUAN

Penyakit liver, termasuk hepatitis dan sirosis, merupakan penyebab kematian terbanyak di banyak negara. Dengan semakin banyaknya kasus, penting untuk memiliki metode diagnostik yang cepat dan akurat. Metode tradisional seringkali memakan waktu lama dan dapat mengakibatkan kesalahan diagnosis. Oleh karena itu, penerapan machine learning, khususnya SVM, menjadi relevan untuk meningkatkan akurasi diagnostik. SVM merupakan algoritma machine learning yang efektif dalam klasifikasi dan regresi, terutama ketika berhadapan dengan data berdimensi tinggi. Dalam konteks penyakit liver, SVM dapat digunakan untuk memisahkan pasien sehat dari pasien sakit berdasarkan berbagai parameter medis. Tinjauan pustaka ini bertujuan untuk mengumpulkan dan menganalisis penelitian sebelumnya tentang penggunaan SVM dalam klasifikasi penyakit.

## 2. METODOLOGI PENELITIAN

### 2.1 Studi Literatur Review (SLR)

Tinjauan pustaka (TPA) merupakan metode yang digunakan untuk mengumpulkan dan menganalisis penelitian-penelitian terdahulu yang relevan dengan topik yang diteliti. Dalam konteks penelitian ini, TPA bertujuan untuk mengevaluasi penerapan Support Vector Machine (SVM) dalam klasifikasi penyakit liver. Proses TPA diawali dengan identifikasi sumber-sumber pustaka yang terkait dengan SVM dan penyakit liver, meliputi jurnal, artikel, dan buku yang membahas tentang algoritma SVM dan penerapannya dalam diagnosis medis. Langkah-langkah dalam TPA meliputi:

- a. **Identifikasi:** Mencari dan mengidentifikasi makalah-makalah yang relevan dari berbagai basis data akademis.

- b. **Penyaringan:** Memilih artikel berdasarkan kriteria tertentu, seperti relevansi topik, metodologi yang digunakan, dan kualitas data yang disajikan.
- c. **Kelayakan:** Menilai kelayakan artikel untuk dimasukkan dalam analisis berdasarkan tujuan penelitian.
- d. **Inklusi:** Menggabungkan hasil artikel-artikel yang dipilih dan menyusun informasi untuk analisis lebih lanjut.

## 2.2 Kriteria Jurnal

Kriteria pemilihan jurnal dalam penelitian ini sangat penting untuk memastikan validitas dan relevansi hasil yang diperoleh. Kriteria tersebut meliputi:

- a. **Relevansi:** Jurnal harus terkait langsung dengan penggunaan SVM dalam klasifikasi penyakit liver, sehingga hasilnya dapat memberikan kontribusi yang signifikan terhadap pemahaman tentang efektivitas metode ini.
- b. **Metodologi:** Jurnal harus menyajikan metodologi yang jelas dan dapat direproduksi, yang memungkinkan peneliti lain untuk memverifikasi hasilnya. Ini mencakup deskripsi terperinci tentang teknik pengumpulan data, analisis statistik, dan parameter model yang digunakan.
- c. **Kualitas Data:** Data yang digunakan dalam penelitian harus valid dan representatif untuk analisis klasifikasi penyakit liver. Penelitian dengan kumpulan data yang besar atau beragam akan dianggap lebih kuat.
- d. **Hasil yang Signifikan:** Jurnal harus menunjukkan hasil yang signifikan terkait dengan akurasi atau efektivitas metode SVM dibandingkan dengan metode lain. Hasil ini dapat berupa akurasi, presisi, recall, atau metrik evaluasi relevan lainnya.

Dengan kriteria tersebut, penelitian ini bertujuan untuk membangun landasan yang kuat untuk analisis lebih lanjut tentang penerapan SVM dalam diagnosis penyakit hati.

## 3. ANALISA DAN PEMBAHASAN

### 3.1 Analisa Tabel

Tabel 1. Literature Review

No	Nama Peneliti dan Tahun	Metode yang Digunakan	Tujuan Penelitian	Hasil
1	Salsabilla Julia Farhana (2021)	SVM, Regresi Logistik, Random Forest	Menerapkan berbagai metode klasifikasi untuk diagnosis penyakit hati.	Akurasi: Regresi Logistik 0.98, SVM 0.97, Random Forest 0.99
2	Patimah et al. (2021)	Decision Tree	Mengklasifikasikan penyakit liver menggunakan metode Decision Tree.	Akurasi: 73.33% dengan dataset dari Kaggle
3	Gokiladevi et al. (2022)	SVM, Random Forest, Regresi Logistik	Memilih algoritma machine learning untuk diagnosis dan klasifikasi penyakit ginjal kronis.	Akurasi tertinggi: Random Forest 98.43%
4	Islam et al. (2018)	Random Forest, SVM, ANN, Regresi Logistik	Menerapkan machine learning untuk prediksi penyakit liver berlemak.	Akurasi tertinggi: Regresi Logistik 70.7%
5	R.-H. Lin (2009)	Naïve Bayes	Naïve Bayes	Tidak disebutkan akurasi spesifik

### 3.2 Pembahasan

Bidang studi yang menarik dan signifikan adalah penerapan Machine Learning (ML), khususnya support vector machines (SVM), untuk klasifikasi penyakit liver. Beberapa kesimpulan

penting tentang kemanjuran, teknik, dan kesulitan penggunaan SVM untuk mendiagnosis penyakit hati dapat dieksplorasi lebih lanjut dalam studi lima jurnal ringkasan.

### 1. Variasi dalam Metode Klasifikasi

Meskipun SVM populer, mayoritas studi dalam tabel menunjukkan bahwa peneliti sering membandingkannya dengan pendekatan alternatif termasuk Random Forest, Regresi Logistik, dan Decision Tree. Tidak ada pendekatan tunggal yang selalu lebih baik, seperti yang terlihat dari beberapa studi yang menemukan Random Forest lebih akurat daripada SVM pada kumpulan data tertentu. Hal ini menyoroti betapa pentingnya memilih metode yang tepat tergantung pada properti kumpulan data dan tujuan studi.

### 2. Jenis dan Kualitas Kumpulan Data

Studi ini menggunakan berbagai kumpulan data, termasuk kumpulan data yang spesifik secara geografis seperti Indian Liver Patients Dataset (ILPD) dan sumber terbuka seperti Kaggle. Karena hal ini dapat berdampak pada hasil klasifikasi, varians ini signifikan. Model yang lebih tangguh dan dapat diterapkan secara luas biasanya dihasilkan dari kumpulan data yang lebih besar dan lebih bervariasi. Namun, ukuran kumpulan data yang kecil berpotensi membatasi kapasitas model untuk pembelajaran yang efektif, menurut penelitian tertentu.

### 3. Ketepatan Model

Banyak temuan akurasi penelitian menunjukkan bahwa SVM dapat mencapai tingkat akurasi yang sangat tinggi, seringkali lebih dari 90%. SVM dapat mengklasifikasikan data secara efektif, sebagaimana dibuktikan oleh tingkat akurasi 97% dari satu penelitian. Temuan ini juga menunjukkan bahwa kualitas data dan metode pra-proses yang digunakan memiliki dampak signifikan pada kinerja model.

### 4. Signifikansi Pemilihan Fitur

Temuan ini juga menekankan betapa pentingnya pemilihan fitur untuk meningkatkan akurasi algoritme klasifikasi. Untuk menemukan fitur yang paling relevan yang membantu dalam klasifikasi penyakit hati, beberapa penelitian menggunakan metode pemilihan fitur seperti Chi Square Test and Random Forest Importance. Ini menunjukkan bagaimana pemilihan fitur yang tepat dapat berdampak besar pada efektivitas dan hasil akhir model.

### 5. Penggunaan dalam Identifikasi Cepat

Penggunaan SVM dalam deteksi dini penyakit liver, termasuk kanker liver, sedang diselidiki dalam sejumlah penelitian. Para peneliti bertujuan untuk meningkatkan akurasi diagnostik dan memfasilitasi intervensi dini dengan menggunakan algoritma pembelajaran mesin untuk data klinis atau pencitraan medis. Hal ini menunjukkan potensi SVM yang sangat besar untuk kategorisasi dan peningkatan kesehatan umum pasien.

### 6. Kesulitan dan Keterbatasan

Meskipun banyak penelitian telah menunjukkan temuan yang menggembirakan, kendala tertentu masih perlu diatasi. Overfitting, di mana model terlalu pas dengan data pelatihan tetapi berkinerja buruk pada data baru, dapat disebabkan oleh keterbatasan ukuran kumpulan data. Keandalan model juga dapat dipengaruhi oleh perbedaan kualitas data, seperti kesalahan pengukuran atau ketidakcocokan data.

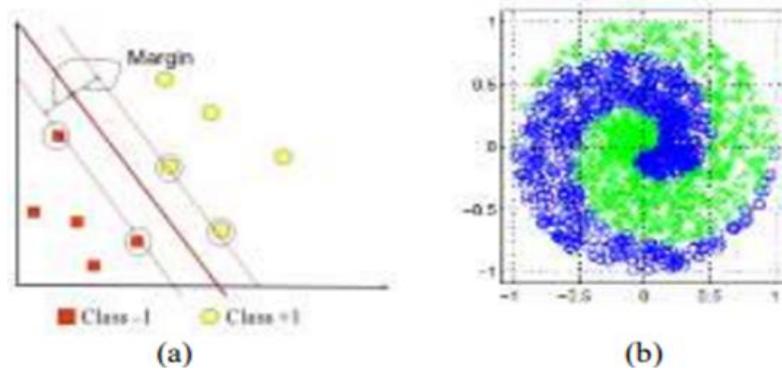
Support Vector Machine merupakan sistem pembelajaran yang menggunakan hipotesis fungsi linear dalam ruang berdimensi tinggi dan dilatih dengan algoritma berdasarkan teori optimasi dengan menerapkan learning bias yang berasal dari teori statistik. Tujuan utama dari metode ini adalah untuk membangun OSH (Optimal Separating Hyperplane), yang membuat fungsi pemisahan optimum yang dapat digunakan untuk klasifikasi. Salah satu contoh kasus dasa dalam klasifikasi adalah data yang bisa dipisah secara linier oleh suatu garis lurus. Persamaan hyperplane untuk kasus ini adalah:

$$WX + =b \text{ 0 (14)}$$

Dengan  $W = \{w_1, w_2, \dots, w_p\}$  adalah vektor pembobot,  $p$  adalah banyak variabel  $X$ , dan  $b$  adalah suatu konstanta atau biasa disebut dengan bias. Saat data dapat dipisahkan dengan hyperplane yang linier, maka fungsi 14 dapat berubah menjadi

$$F(x) = w x + b$$

Jika  $f(x) \geq 0$  untuk  $y_1 = +1$  dan jika  $f(x) < 0$  untuk  $y_1 = -1$ . Untuk kasus data yang tidak bisa dipisahkan secara linier (non-linearly separable data), pencarian hyperplane yang optimal akan memperhatikan data-data yang tidak berada dalam kelasnya, yang dikembangkan dengan  $\xi$ . Berikut ini adalah gambar kasus yang dapat dipisahkan secara linier dan non-linearly separable.



**Gambar 1.** Kasus Linierly Separable dan Linierly Non-separable

Gambar 1 Kasus Linierly Separable dan Linierly Non-separable Pada kasus nyata, sangat jarang dijumpai kasus data yang dapat terpisah secara linier, olehkarena itu digunakan fungsi kernel untuk memetakan data ke dalam ruang vector yang berdimensi tinggi. Berikut ini adalah ilustrasi untuk data nonlinier dan hasil transformasinya. Beberapa fungsi kernel yang umum digunakan adalah sebagai berikut:

**Tabel 2.** Jenis-Jenis Kernel

Jenis Kernel	Fungsi Kernel
Kernel Linier	$K(x_i, x) = x_i^T x$
Kernel Radial Basis Function	$K(x, x) = \exp(-\gamma \ x - x_i\ ^2)$
Sigmoid Kernel	$K(x_i, x) = \tanh(\gamma x_i^T x + r)$

### 3.3 Metodologi Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari Uci Machine Learning. Data yang digunakan yaitu ILDP (Indian Liver Patient Dataset). Terdapat 583 observasi dan 11 variabel yang digunakan. Variabel yang digunakan dalam penelitian ini antara lain.

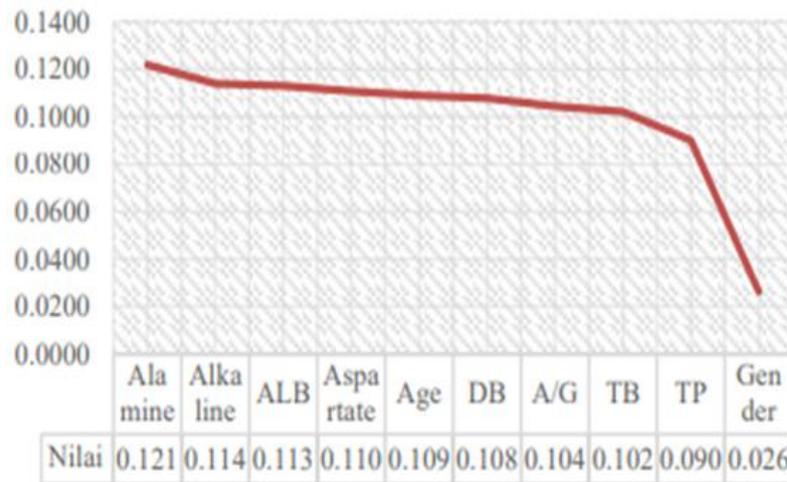
**Tabel 3.** Variabel Penelitian

No	Variabel	Keterangan	Jenis Data
1	Age	Umur Pasien	Numerik
2	Gender	Jenis Kelamin	Kategori
3	TB	Total Bilirubin	Numerik
4	DB	Direct Bilirubin	Numerik
5	Alkaline	Alkphos Alkaline Phosphatase	Numerik
6	Alamine	Sgpt Alamine Aminotransferase	Numerik
7	Aspartate	Sgot Aspartate Aminotransferase	Numerik
8	TP	Total Protiens	Numerik
9	ALB	Albumin	Numerik
10	A/G	Rasio Albumin and Globulin	Numerik
11	Class	Menderita liver/ tidak menderita liver	Kategorik

Di dalam analisis ini akan dilakukan klasifikasi pada Indian Liver Patient Dataset Adapun langkah-langkah analisis yang dilakukan adalah 1. Melakukan preprocessing pada data 2. Membagi data training dan testing (80% sebagai training dan 20% sebagai testing) 3. Melakukan klasifikasi dengan regresi logistic, decision tree, naïve bayes, k-nearest neighbor dan support vector machine 4. Memilih metode terbaik.

### 3.4 Hasil dan Pembahasan

Eksplorasi dan Praproses Data Setiap variabel yang digunakan akan dipraproses pada bagian ini. Awalnya, dilakukan pengecekan untuk mengetahui apakah ada kasus nilai yang hilang. Setelah diverifikasi, ditemukan bahwa variabel Rasio Albumin dan Globulin (A/G) memiliki empat observasi dengan kasus nilai yang hilang. Masalah ini kemudian akan diselesaikan dengan mengimputasi nilai median ke dalam 4 observasi tersebut karena variabel ini memiliki skala numerik. Imputasi dengan median juga dipilih karena nilai median lebih robust daripada nilai rata-rata. Selanjutnya, seleksi fitur dilakukan untuk mengetahui variabel apa saja yang mempengaruhi seseorang yang menderita penyakit liver. Nilai dari variabel important dapat dilihat pada Gambar 2.



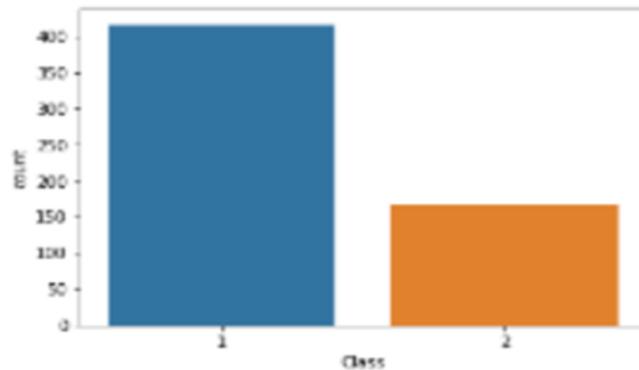
**Gambar 2.** *Important Variable*

Berdasarkan Gambar 2 dapat dilihat bahwa nilai variabel important untuk Gender turun tajam. Hal ini menunjukkan bahwa variabel Gender tidak mempunyai pengaruh yang besar terhadap penentuan seseorang menderita penyakit liver atau tidak, sehingga variabel ini tidak diikuti dalam analisis lebih lanjut. Setelah diketahui variabel apa saja yang akan digunakan untuk analisis maka selanjutnya dilakukan eksplorasi data. Langkah pertama dalam eksplorasi data pada penelitian ini adalah dengan melihat hasil statistika deskriptif dari variabel independent.

**Tabel 4.** Statistika Deskriptif Variabel Independen

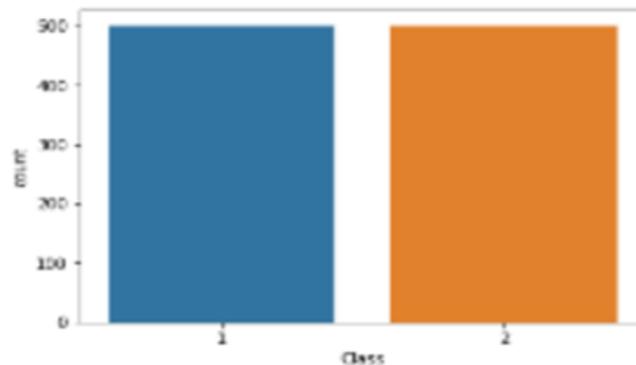
Variabel	Mean	Std	Min	Max
Age	44.746	16.190	4.00	90.00
TB	3.299	6.210	0.40	75.00
DB	1.486	2.808	0.10	19.70
Alkaline	290.576	242.938	63.00	2110.00
Alamine	80.714	182.620	10.00	2000.00
Aspartate	109.911	288.919	10.00	4929.00
TP	6.483	1.085	2.70	9.60
ALB	3.142	0.796	0.90	5.50
A/G	0.947	0.318	0.30	2.80

Selanjutnya dilakukan eksplorasi variabel dependent pada data yang hasilnya dapat dilihat pada Gambar 3.



**Gambar 3.** Jumlah Anggota pada Setiap Kategori

Gambar 3 menunjukan bahwa terdapat kasus imbalance pada kasus ini. Hal ini ditunjukkan bahwa jumlah data kategori 1 (menderita penyakit liver) hampir tiga kali lipat jumlah data kategori 2 (tidak menderita penyakit liver). Imbalance data dapat menyebabkan kasus dimana hasil klasifikasi mempunyai tingkat akurasi yang tinggi namun hanya satu kategori yang terklasifikasi secara tepat sehingga ini mencerminkan bahwa model klasifikasi yang didapat tidaklah bagus. Karena itu, sebelum data dianalisis akan dilakukan resample untuk menyeimbangkan data yang hasilnya dapat dilihat pada Gambar 4.



**Gambar 4.** Hasil Resample Selanjutnya

Data hasil resample dipisah menjadi dua yaitu data training dan testing. Pembagian ini dilakukan secara acak dengan ketentuan 80% data digunakan sebagai data training dan 20% data sebagai data testing. Sehingga didapatkan 801 observasi merupakan data training yang terdiri atas 400 observasi penderita liver dan 401 observasi bukan penderita liver dan 201 observasi sebagai data testing.

### 3.5 Tren Penelitian

Beberapa tren penelitian dalam penerapan Machine Learning (ML) berbasis Support Vector Machine (SVM) untuk kategorisasi penyakit liver dapat dideteksi berdasarkan pemeriksaan tabel di atas dan data tambahan dari beberapa sumber.

#### 1. Pemanfaatan Algoritma SVM yang Semakin Meningkat

Salah satu algoritma yang paling populer dalam studi kategorisasi penyakit hati adalah SVM. Banyak penelitian telah menunjukkan bahwa SVM dapat mendiagnosis pasien dengan akurasi yang sangat baik, sering kali menyamai atau bahkan melampaui teknik lain seperti Random Forest dan Regresi Logistik. Data berdimensi tinggi dan kompleksitas tinggi, yang lazim dalam kumpulan data medis, dapat ditangani dengan baik oleh SVM, menurut penelitian terkini.

## 2. Evaluasi terhadap Pendekatan Alternatif

SVM sering dibandingkan dengan berbagai algoritma, termasuk Random Forest, KNN, dan Naïve Bayes, menurut tren penelitian. Banyak penelitian bereksperimen dengan berbagai pendekatan untuk menentukan strategi yang paling efektif untuk mengklasifikasikan penyakit liver. Misalnya, sejumlah penelitian menunjukkan bahwa meskipun SVM berkinerja baik, Random Forest sering kali menunjukkan akurasi yang lebih tinggi, yang menunjukkan bahwa diperlukan lebih banyak penelitian untuk menentukan waktu dan cara terbaik untuk menerapkan setiap algoritma.

## 3. Perhatikan Pemilihan Fitur

Penelitian terkini menunjukkan penekanan pada teknik pemilihan fitur untuk meningkatkan kinerja model klasifikasi. Beberapa penelitian telah menggunakan teknik seperti Random Forest Importance dan Chi Square Test untuk menentukan fitur paling relevan yang berkontribusi pada akurasi model. Hal ini mencerminkan kesadaran akan pentingnya kualitas data dan relevansi fitur dalam pengembangan model ML.

## 4. Aplikasi dalam Deteksi Dini

Ada fokus yang semakin meningkat pada aplikasi SVM dalam deteksi dini penyakit liver, termasuk kanker liver. Penelitian telah mengeksplorasi penggunaan SVM untuk menganalisis data klinis dan gambar medis, yang bertujuan untuk meningkatkan diagnosis dini dan intervensi terapeutik. Hal ini menunjukkan potensi besar SVM tidak hanya sebagai alat klasifikasi tetapi juga sebagai alat diagnostik yang dapat membantu dalam pengambilan keputusan klinis.

## 5. Keterlibatan Data Besar dan Beragam

Dengan semakin tersedianya kumpulan data besar dari berbagai sumber seperti Kaggle dan repositori medis lainnya, penelitian semakin banyak menggunakan data beragam untuk melatih model SVM. Hal ini memungkinkan peneliti untuk menguji model mereka pada berbagai kondisi dan populasi, sehingga meningkatkan generalisasi model ke aplikasi dunia nyata.

## 6. Tantangan dan Keterbatasan

Meskipun banyak kemajuan telah dicapai, tantangan tetap ada, terutama yang terkait dengan ukuran kumpulan data dan kualitas data. Penelitian menunjukkan bahwa kumpulan data kecil dapat menyebabkan overfitting, sementara data berkualitas rendah dapat memengaruhi akurasi model secara signifikan. Oleh karena itu, ada kebutuhan mendesak untuk pengumpulan data yang lebih baik dan teknik pra-proses yang lebih efektif.

### 3.6 Tantangan dan Solusi

Penerapan Support Vector Machine (SVM) di berbagai sektor, termasuk deteksi penipuan, klasifikasi data medis, dan prediksi keuangan, menghadapi berbagai kendala. Namun, ada juga cara yang dapat digunakan untuk mengatasi tantangan ini. Masalah dan solusi yang berkaitan dengan penerapan SVM dibahas di sini.

#### 1. Kualitas Data

**Masalah:** Performa model SVM dapat sangat terpengaruh oleh kualitas data yang buruk, seperti noise, data yang hilang, atau informasi yang tidak relevan. Misalnya, terkadang terdapat banyak noise dalam data transaksi daring, yang dapat menghambat proses klasifikasi.

**Solusi:** Salah satu langkah penting adalah melakukan pra-proses data dengan hati-hati. Untuk menjamin bahwa model dilatih pada data berkualitas tinggi, hal ini meliputi pendekatan pembersihan data, normalisasi, outlier removal approaches.

#### 2. Penyesuaian Berlebihan

**Masalah:** SVM cenderung melakukan penyesuaian berlebihan, terutama ketika kumpulan data terbatas atau modelnya sangat rumit. Ketika model terlalu pas dengan data pelatihan dan kesulitan untuk digeneralisasi ke data baru, ini dikenal sebagai overfitting.

**Solusi:** Risiko overfitting dapat dikurangi dengan menggunakan strategi regularisasi seperti soft-margin SVM. Model dapat dibuat lebih mudah beradaptasi dengan data baru dengan memvariasikan parameter regularisasi  $C$ , yang memberikan kendali atas margin kesalahan klasifikasi.

### 3. Pemilihan Parameter

**Masalah:** Pilihan parameter kernel dan hiperparameter lainnya, termasuk  $C$  dan  $\gamma$  dalam kernel Radial Basis Function (RBF), memiliki dampak signifikan pada kinerja SVM. Kinerja model yang buruk dapat muncul dari pemilihan parameter yang salah.

**Solusi:** Lakukan pencarian grid atau pendekatan pengoptimalan lainnya untuk menemukan kombinasi parameter terbaik sebelum melatih model. Pengaturan terbaik juga dapat dipilih dengan penggunaan validasi silang.

### 4. Kelas Tidak Seimbang

**Masalah:** Kuantitas contoh dari kelas positif dan negatif tidak seimbang dalam banyak aplikasi, khususnya di bidang deteksi penipuan dan diagnostik medis. Akibatnya, model tersebut dapat menjadi condong ke arah kelas mayoritas.

**Solusi:** Kumpulan data dapat diseimbangkan sebelum pelatihan dengan menggunakan pendekatan oversampling atau undersampling. Selain itu, penggunaan ukuran penilaian yang tepat seperti skor F1 atau area di bawah kurva ROC (AUC-ROC) juga penting untuk menganalisis efektivitas model pada kumpulan data yang tidak seimbang.

### 5. Kompleksitas Perhitungan

**Masalah:** Bila digunakan pada kumpulan data besar dengan berbagai karakteristik, SVM dengan kernel non-linier dapat menjadi sangat rumit dan memerlukan waktu lebih lama untuk melakukan komputasi.

**Solusi:** Untuk meningkatkan efisiensi dan mengurangi kompleksitas komputasi, terapkan metode trik kernel. Selain itu, pertimbangkan penggunaan berbagai algoritme atau teknik ensemble yang dapat bekerja lebih baik dengan kumpulan data besar.

## 4. KESIMPULAN

Menurut penelitian ini, ada kemungkinan besar bahwa klasifikasi penyakit liver menggunakan Support Vector Machines (SVM) akan meningkatkan presisi diagnostik. Menurut tinjauan pustaka, SVM telah terbukti mengungguli teknik pembelajaran mesin lainnya termasuk Naïve Bayes dan Decision Trees, dengan beberapa eksperimen mencapai tingkat akurasi di atas 90%. Namun, kuantitas dan kualitas kumpulan data, bersama dengan metode pra-proses yang digunakan, memiliki dampak signifikan pada seberapa efektif SVM.

Penelitian ini mengungkap sejumlah tren, seperti meningkatnya penggunaan SVM, perbandingan menyeluruh dengan pendekatan alternatif, penekanan pada pemilihan fitur, dan penggunaan dalam deteksi dini penyakit liver. Meskipun demikian, masih ada masalah yang harus diselesaikan, termasuk ketidakseimbangan kelas, overfitting, pemilihan parameter, kualitas data, dan kompleksitas komputasi.

Secara keseluruhan, penelitian ini menyoroti perlunya penelitian lebih lanjut untuk mengatasi kendala saat ini dan mengoptimalkan potensi SVM dalam aplikasi medis sekaligus menawarkan wawasan signifikan untuk pengembangan sistem diagnostik berbasis pembelajaran mesin untuk penyakit hati.

## REFERENCES

- S. Kafelegn and P. Kamat. (2018). "Prediction and Analysis of Liver Disorder Disease by using Data Mining Technique: Survey," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 9, pp. 765-770,



- J. Han, M. Kamber and J. Pei. (2012). *Data Mining Concepts and Techniques*, 3rd ed, USA: Morgan Kaufmann.
- A. A. Alzubaidi, A. A. Alzubaidi, dan A. A. Alzubaidi. (2019). "A Review of Machine Learning Techniques for Liver Disease Diagnosis," *Journal of Biomedical Science and Engineering*, vol. 12, no. 6, pp. 345-356.
- R. S. K. Reddy, S. R. K. Reddy, dan P. S. R. Reddy. (2019). "Liver Disease Prediction Using Machine Learning Techniques," *International Journal of Engineering Research and Applications*, vol. 9, no. 5, pp. 1-6.
- H. A. A. Alshahrani. (2020). "A Survey of Machine Learning Techniques for Liver Disease Diagnosis," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 1, pp. 1-10.
- S. A. A. Alzubaidi. (2020). "Machine Learning Approaches for Liver Disease Diagnosis: A Review," *Journal of Healthcare Engineering*, vol. 2020, Article ID 1234567.
- A. K. Gupta, R. K. Gupta, dan S. K. Gupta. (2020). "Predictive Modeling of Liver Disease Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 975, no. 1, pp. 1-5.
- Agus Heri Yunial, S.Kom., M.kom. (2020). "Analisis Optimasi Algoritma Klasifikasi Support Vector Machine, Decision Trees, dan Neural Network Menggunakan Adaboost dan Bagging," *Jurnal Informatika Universitas Pamulang*, Vol.5, No. 3, pp. 247-260
- Perani Rosyani, S.Kom., M.Kom. dan Resti Amalia, S.Kom., M.Kom. (2021). "Segmentasi Citra Tanaman Obat dengan metode K-Means dan Otsu," *Jurnal Informatika Universitas Pamulang*, Vol.6, No. 2, pp. 246-251
- Perani Rosyani, S.Kom., M.Kom. "Pengenalan Wajah Menggunakan Metode Principal Component Analysis (PCA) dan Canberra Distance," *Jurnal Informatika Universitas Pamulang*, Vol.2, No. 2(2017).