

Penerapan Algoritma Support Vector Machine Berbasis Kernel Radial Basis Function dalam Klasifikasi Sel Kanker

Zainul Arifin¹, Dhimas Fachri Rahman², Bagus Setya Rintyarna³, Daryanto⁴

¹Teknik Informatika, Universitas Muhammadiyah Jember, zainul.arifin@unmuhjember.ac.id

²Teknik Informatika, Universitas Muhammadiyah Jember, dimas.fachri4@gmail.com

³Teknik Elektro, Universitas Muhammadiyah Jember, bagus.setya@unmuhjember.ac.id

⁴Teknik Informatika, Universitas Muhammadiyah Jember, daryanto@unmuhjember.ac.id

Keywords:

Cancer cell classification,
Machine learning,
Medical diagnosis,
Support vector machine

ABSTRACT

Cancer is a leading cause of death globally, with over 10 million deaths reported in 2020, according to the World Health Organization (WHO). Early detection and accurate diagnosis are crucial to improving survival rates. However, conventional diagnostic methods such as biopsies and histopathological analysis have several limitations, including being invasive, time-consuming, and reliant on subjective interpretation by pathologists. With technological advancements, artificial intelligence (AI) and machine learning (ML) offer promising alternatives in cancer diagnosis. This study explores the effectiveness of the Support Vector Machine (SVM) algorithm in classifying cancer cells using the Breast Cancer Wisconsin dataset. The dataset consists of 699 cell samples obtained through fine needle aspiration, each described by 10 morphological features and labeled as benign or malignant. The results show that SVM with a Radial Basis Function (RBF) kernel can classify cancer cells with high accuracy. Data preprocessing, including cleaning and normalization, significantly improves model performance. Additionally, parameter optimization using grid search enhances the model's reliability. This study highlights the strong potential of SVM as an efficient, accurate, and practical decision-support tool in medical diagnosis, particularly for cancer detection.

Kata Kunci

Diagnosis medis,
Klasifikasi sel kanker,
Machine learning,
Support vector machine

ABSTRAK

Kanker merupakan penyebab utama kematian global, dengan lebih dari 10 juta kasus kematian pada tahun 2020 menurut WHO. Deteksi dini dan diagnosis yang akurat sangat penting untuk meningkatkan angka kesembuhan. Namun, metode konvensional seperti biopsi dan analisis histopatologi memiliki berbagai keterbatasan, seperti prosedur yang invasif, waktu analisis yang lama, dan ketergantungan pada interpretasi subjektif ahli. Dengan berkembangnya teknologi, pendekatan berbasis kecerdasan buatan (AI) dan machine learning (ML) menawarkan solusi alternatif dalam diagnosis kanker. Penelitian ini mengkaji efektivitas algoritma Support Vector Machine (SVM) dalam mengklasifikasikan sel kanker menggunakan dataset Breast Cancer Wisconsin, yang terdiri dari 699 sampel sel dengan 10 fitur morfologis dan klasifikasi jinak atau ganas. Hasil menunjukkan bahwa SVM dengan kernel RBF mampu mengklasifikasikan sel kanker secara akurat. Proses pembersihan dan normalisasi data, serta optimasi parameter menggunakan grid search, terbukti meningkatkan performa model. Penelitian ini menunjukkan bahwa SVM memiliki potensi besar sebagai alat bantu diagnosis kanker yang efisien, cepat, dan andal dalam praktik klinis.

Korespondensi Penulis:

Zainul Arifin,
Universitas Muhammadiyah Jember,
Jalan Karimata 49 Jember
Telepon : 6281332285440
Email: zainul.arifin@unmuhjember.ac.id

Submitted : 15-08-2023; Accepted : 10-09-2023;
Published : 30-09-2023

Copyright (c) 2023 The Author (s) This article is distributed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0)

1. PENDAHULUAN

Kanker merupakan salah satu penyakit paling mematikan di dunia dan menjadi penyebab utama kematian di banyak negara [1]. Berdasarkan data dari *World Health Organization* (WHO), diperkirakan terdapat 19,3 juta kasus kanker dan kisaran 10 juta kematian bari yang disebabkan oleh kanker secara keseluruhan di dunia pada tahun 2020 [2]. Angka ini diperkirakan akan terus meningkat seiring dengan pertumbuhan populasi global dan peningkatan usia harapan hidup [3]. Jenis kanker yang sering ditemukan dan menjadi perhatian utama adalah kanker payudara, kanker paru-paru, kanker usus besar, dan kanker prostat [4]. Deteksi dini dan diagnosis yang akurat sangat penting untuk meningkatkan peluang kesembuhan dan mengurangi angka kematian akibat kanker. Oleh karena itu, pengembangan metode yang efektif dan efisien untuk diagnosis kanker menjadi salah satu prioritas utama dalam penelitian medis [5].

Diagnosis kanker secara tradisional dilakukan melalui pemeriksaan fisik, pencitraan medis, dan biopsi [6]. Prosedur ini biasanya melibatkan pemeriksaan histopatologi di mana jaringan yang dicurigai diambil dengan cara diperiksa menggunakan mikroskop oleh ahli patologi. Meskipun metode ini cukup akurat, mereka memiliki beberapa keterbatasan. Proses biopsi dapat menjadi invasif dan menyakitkan bagi pasien, analisis histopatologi digunakan dengan membutuhkan waktu yang cukup relatif lama, serta sangat bergantung pada keahlian dan pengalaman ahli patologi. Selain itu, interpretasi hasil bisa bervariasi antara satu ahli patologi dengan lainnya, yang dapat mengakibatkan ketidakpastian dalam diagnosis [7].

Dalam dekade terakhir, perkembangan teknologi komputasi dan kemajuan dalam bidang kecerdasan buatan (AI) dan *machine learning* (ML) telah membuka peluang baru untuk meningkatkan diagnosis kanker [8]. Algoritma *machine learning* dapat digunakan untuk mengidentifikasi pola-pola kompleks dalam data medis yang mungkin tidak dapat dideteksi oleh manusia. Salah satu algoritma yang populer dan efektif dalam tugas klasifikasi adalah *Support Vector Machine* [9]. SVM dikenal karena kemampuannya untuk menangani data berdimensi tinggi dan menghasilkan model klasifikasi yang akurat. SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan data ke dalam kelas yang berbeda dengan margin maksimum [10].

Dataset *Breast Cancer Wisconsin* yang dibuat oleh Dr. William H. Wolberg merupakan dataset yang umum dan sering digunakan untuk pendekatan penelitian kanker [11]. Dataset terdiri sebanyak 699 sampel sel kanker yang diambil melalui prosedur aspirasi jarum halus (*fine needle aspiration*) pada massa payudara. Setiap sampel dalam dataset ini dicirikan oleh sepuluh fitur yang relevan dengan sifat-sifat morfologis sel kanker, yaitu ketebalan benjolan (*clump thickness*), uniformitas ukuran sel (*uniformity of cell size*), uniformitas bentuk sel (*uniformity of cell shape*), adhesi marginal (*marginal adhesion*), ukuran sel epitel tunggal (*single epithelial cell size*), inti (*bare nuclei*), kromatin lembut (*bland chromatin*), nukleolus normal (*normal nucleoli*), dan mitosis (*mitoses*). Selain itu, setiap sampel diklasifikasikan sebagai jinak atau ganas, memberikan dasar yang kuat untuk pengembangan dan evaluasi model klasifikasi.

2. METODE PENELITIAN

Tahapan Penelitian memberikan gambaran umum tentang proses penelitian yang akan dilakukan dari awal hingga akhir. Langkah – langkah dalam penelitian ini dapat dilihat melalui diagram seperti yang terlihat pada Gambar 1. Tahapan penelitian terdiri dari Pengumpulan Data, Pra-pemrosesan Data, Klasifikasi, dan Evaluasi Model.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data dalam penelitian ini bersumber dari University of Wisconsin Hospitals, Madison yang tersedia di UCI *Machine Learning Repository* [11]. Dataset ini terdiri dari 699 sampel sel yang diambil melalui prosedur *fine needle aspiration* pada massa sel kanker, dengan masing-masing sampel diklasifikasikan sebagai *benign* (jinak) atau *malignant* (ganas). Setiap sampel dideskripsikan oleh 10 fitur yang meliputi ketebalan benjolan, ukuran sel, bentuk sel, adhesi marginal, ukuran sel epitel tunggal, inti, kromatin lembut, nukleolus normal, dan mitosis.

Tabel 1. Atribut Dataset

No	Kolom	Deskripsi
1	ID	Identifikasi Sampel
2	Clump	Ketebalan Klump
3	UnifSize	Ukuran Seragam Sel
4	UnifShape	Bentuk Seragam Sel
5	MargAdh	Adhesi Marjinal
6	SingEpiSize	Ukuran Epitel Sel Tunggal
7	BareNuc	Nuklei Kosong
8	BlandChrom	Kromatin
9	NormNud	Nukleolus Normal

10	Mit	Mitosis
11	Class	Kategori (2=Jinak, 4=Ganas)

2.2 Preprocessing

Pra-pemrosesan data merupakan langkah penting untuk memastikan data yang digunakan dalam model berada dalam kondisi optimal. Langkah-langkah pra-pemrosesan data dalam penelitian ini adalah sebagai berikut:

1. Penghapusan Nilai Hilang : Penghapusan baris yang memiliki nilai hilang pada atribut *Bare Nuclei*. Ini dilakukan untuk menjaga kualitas data yang digunakan dalam model.
2. Konversi Tipe Data : Konversi kolom *Bare Nuclei* dari tipe objek menjadi tipe *integer*. Hal ini penting karena model SVM membutuhkan input data dalam format numerik.
3. Normalisasi Data : Normalisasi dilakukan untuk memastikan bahwa setiap fitur memiliki skala yang sama, sehingga tidak ada fitur yang mendominasi fitur lainnya dalam proses pembelajaran. Normalisasi dilakukan menggunakan metode *min-max scaling*.

2.3 Pelatihan Model SVM

Algoritma SVM dipilih untuk membangun model klasifikasi. SVM bekerja dengan cara memetakan sebuah data ke dalam ruang berdimensi tinggi untuk menentukan *hyperplane* yang dapat memisahkan tiap kelasnya dengan margin maksimum. Untuk penelitian ini, kernel *Radial Basis Function* (RBF) digunakan karena kemampuannya untuk menangani data non-linear.

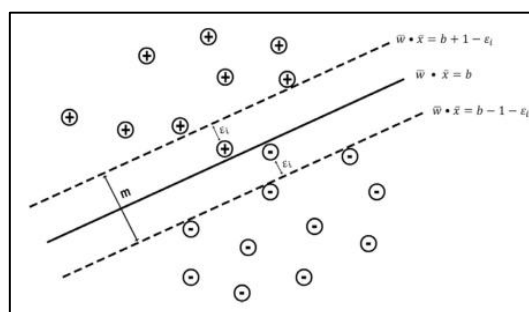
2.4 Support Vector Machine

Support Vector Machine adalah salah satu metode klasifikasi yang sering digunakan pada data mining. SVM adalah salah satu algoritma machine learning bertipe supervised yang berbasis vektor. Dalam mengklasifikasikan suatu data, SVM akan membagi data melalui *hyperplane* menjadi dua kelas. *Hyperplane* terletak di antara dua kelas dengan jarak d yang merupakan margin pada titik terdekat setiap kelas. Titik yang berada pada margin disebut dengan *support vector*. Tujuan dari SVM yaitu menentukan nilai *hyperplane* terbaik [12].

Kernel dalam SVM merupakan komponen penting dalam pembelajaran mesin, dimana kernel yang dipakai menggunakan fungsi matematis untuk mengubah data input menjadi vektor pada ruang fitur yang mempunyai dimensi lebih tinggi. Fungsi kernel mampu menentukan tingkat kemiripan antar titik data dalam ruang fitur. SVM menggunakan kernel untuk melakukan klasifikasi data berdasarkan perbedaan fitur sehingga terpisah atau terklasifikasi ke dalam label kelas yang berbeda, proses ini memungkinkan klasifikasi yang akurat terhadap data nonlinier [13].

Fitur-fitur yang meliputi SVM:

1. Margin Maksimum : SVM digunakan untuk memaksimalkan sudut antara *hyperplane* pemisah dan titik data terdekat dari setiap kelas.
2. Vektor pendukung, atau vektor turunan: Ini adalah titik data yang paling dekat dengan *hyperplane*, yang menjadi pedoman penting untuk menentukan posisinya.
3. Kernel Trick: Teknik ini memungkinkan penggunaan fungsi kernel untuk mengurai data menjadi dimensi yang lebih mudah dikelola, sehingga lebih mudah untuk menangani distribusi data yang tidak linier.



Gambar 2. Ilustrasi Hyperplane

Secara konsep SVM mempunyai tujuan mencari *hyperplane* terbaik dalam memisahkan kelas – kelas data. Ketika data non-linear tidak bisa dipisahkan dengan garis lurus di ruang asli, maka diperlukan cara untuk membuat pemisahan kelas data secara linear pada ruang lain. Kernel adalah fungsi pemetaan implisit yang memproyeksikan data dari ruang berdimensi rendah ke ruang berdimensi lebih tinggi agar data bisa dipisahkan secara linear di ruang tersebut tanpa benar-benar menghitung transformasi tersebut secara eksplisit.

Kernel RBF (Radial Basis Function) adalah salah satu jenis kernel yang digunakan untuk mengukur kemiripan antar data. Rumus RBF kernel adalah:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

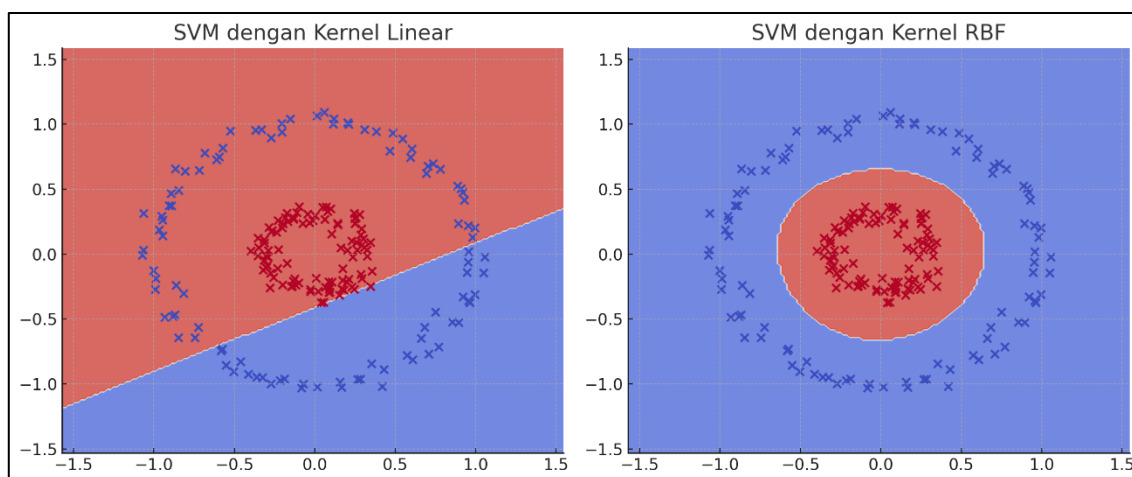
Dimana:

- x dan x' adalah dua vektor fitur.
- γ adalah parameter yang menentukan seberapa jauh pengaruh satu titik data terhadap yang lain.

RBF mengubah data non-linear di ruang asli menjadi data yang lebih mudah dipisahkan secara linear di ruang fitur berdimensi tinggi.

Nilai $K(x, x')$ akan mendekati 1 jika x dan x' sangat dekat (mirip), dan mendekati 0 jika mereka sangat berbeda. Ini membantu SVM dalam menentukan support vectors dan margin terbaik. Dengan bantuan parameter γ , RBF kernel bisa menyesuaikan model agar lebih kompleks atau lebih sederhana sesuai kebutuhan:

- γ kecil \rightarrow pengaruh tiap titik data lebih besar, model lebih halus.
- γ besar \rightarrow pengaruh lebih lokal, model lebih kompleks dan sensitif terhadap data training.



Gambar 3. Perbandingan Hasil Klasifikasi Kernel Linear dengan Kernel RBF

Gambar 3 merupakan visualisasi perbandingan antara SVM dengan kernel linear dan SVM dengan kernel RBF, kernel linear gagal memisahkan dua kelas data, kernel RBF berhasil memisahkan dua kelas dengan baik karena kernel RBF memetakan data ke ruang berdimensi lebih tinggi..

2.5 Evaluasi Confusion Matrix

Setelah klasifikasi tahap selanjutnya ialah melakukan evaluasi menggunakan *confusion matrix*, dimana *confusion matrix* digunakan untuk melihat performa suatu model pada *machine learning* yang digunakan untuk proses klasifikasi. Gabungan jumlah data yang sebenarnya dan jumlah data yang diprediksi ditunjukkan oleh *Confusion matrix*. *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, juga *False Negative (FN)* merupakan istilah yang dipakai untuk mempresentasikan hasil proses klasifikasi. Untuk mengukur performa model seperti *accuracy*, *precision* dan *recall* menggunakan nilai tersebut.

Tabel 2. Confussion Matrix

		Actual Class	
		Positif	Negatif
Predicted Class	Positif	TP	FP
	Negatif	FN	TN

Setelah mendapatkan nilai TP, TN, FP, dan FN melalui *confusion matrix*. Nilai tersebut dapat digunakan untuk mengukur performa model seperti *accuracy*, *precision*, dan *recall*.

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

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

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

3. HASIL DAN ANALISIS

3.1 Dataset

id	clump_thickness	size_uniformity	shape_uniformity	marginal_adhesion	epithelial_size	bare_nucleoli	bland_chromatin	normal_nucleoli	mitoses	class
1000025	5	1	1	1	2	1	3	1	1	2
1002945	5	4	4	5	7	10	3	2	1	2
1015425	3	1	1	1	2	2	3	1	1	2
1016277	6	8	8	1	3	4	3	7	1	2
1017023	4	1	1	3	2	1	3	1	1	2
1017122	8	10	10	8	7	10	9	7	1	4
1018099	1	1	1	1	2	10	3	1	1	2
1018561	2	1	2	1	2	1	3	1	1	2
1033078	2	1	1	1	2	1	1	1	5	2
1033078	4	2	1	1	2	1	2	1	1	2
1035283	1	1	1	1	1	1	3	1	1	2
1036172	2	1	1	1	2	1	2	1	1	2
1041801	5	3	3	3	2	3	4	4	1	4
1043999	1	1	1	1	2	3	3	1	1	2
1044572	8	7	5	10	7	9	5	5	4	4
1047630	7	4	6	4	6	1	4	3	1	4
1048672	4	1	1	1	2	1	2	1	1	2
1049815	4	1	1	1	2	1	3	1	1	2
1050670	10	7	7	6	4	10	4	1	2	4

Gambar 4. Dataset Breast Cancer Wisconsin

Dataset yang digunakan pada penelitian ini adalah data kanker payudara yang dibuat oleh Dr. William H. Wolberg dari University of Wisconsin Hospitals Madison, Wisconsin, Amerika Serikat. Data ini dibuat melalui pengumpulan data dimulai pada tahun 1989 sampai tahun 1991, dan data ini secara bebas boleh digunakan dalam penelitian oleh siapapun. Dataset ini terdiri dari 699 record dengan 11 atribut. Atribut pertama adalah untuk nomor id dan tidak diperlukan untuk penelitian, sehingga dihapus dan tidak digunakan dalam klasifikasi. Atribut angka 10 mewakili nilai kelas yang memiliki dua nilai yaitu nilai 2 dan nilai 4, di mana nilai 2 mewakili label *benign* dan nilai 4 mewakili label *malignant*. Atribut lainnya mempunyai rentang nilai numerik dari 1 hingga 10. Arti numerik nilai tersebut mewakili karakteristik masing - masing atribut, dimana semakin besar nilai maka tingkat keganasan (*malignancy*) juga semakin besar [14].

3.2 Pemilihan Parameter Optimal dengan *Grid Search*

Grid Search adalah teknik yang digunakan untuk mengoptimalkan hiperparameter dari suatu algoritma pembelajaran mesin, seperti SVM. Teknik ini melibatkan penentuan rentang nilai yang mungkin untuk setiap hiperparameter, termasuk nilai minimum, nilai maksimum, dan jumlah langkah. Algoritma *Grid Search* kemudian membuat *grid* dari semua kemungkinan kombinasi hiperparameter dan menguji setiap kombinasi menggunakan validasi silang (*cross-validation*) untuk mengevaluasi kinerjanya. Dengan secara sistematis menguji semua kombinasi yang mungkin, *Grid Search* bertujuan untuk mengidentifikasi kumpulan hiperparameter optimal yang akan menghasilkan performa terbaik untuk model. Penggunaan validasi silang membantu mencegah *overfitting* dengan mengevaluasi kinerja model pada data yang tidak digunakan dalam pelatihan. Tujuan dari *Grid Search* adalah untuk menemukan kombinasi hiperparameter terbaik yang memungkinkan model memprediksi data yang tidak dikenal secara akurat [15]. Berikut adalah implementasi *grid search* menggunakan `GridSearchCV` dari scikit-learn :

```
from sklearn.model_selection import GridSearchCV

# Definisi parameter grid
param_grid = {
    'C': [0.1, 1, 10, 100],
    'gamma': [1, 0.1, 0.01, 0.001],
    'kernel': ['rbf']
}

# Inisialisasi model SVM
svm_model = SVC()

# Grid search
grid_search = GridSearchCV(svm_model, param_grid, refit=True,
    verbose=2, cv=5)
grid_search.fit(X_train, y_train)

# Menampilkan parameter terbaik
print(f"Best Parameters: {grid_search.best_params}")
```

Hasil dari *grid search* menunjukkan bahwa kombinasi parameter terbaik untuk SVM dengan kernel RBF pada dataset ini adalah $C = 10$ dan $\gamma = 0.001$. Dengan parameter ini, model *Support Vector Machine* mampu mencapai performa terbaik dalam hal *accuracy*, *precision*, *recall*, dan *F1-score*.

3.3 Implementasi dengan Python

Implementasi *Support Vector Machine* dilakukan menggunakan pustaka *scikit-learn* di python. Berikut adalah langkah-langkah dan cuplikan kode untuk mengimplementasikan *Support Vector Machine*:

```
# Menginisialisasi dan melatih model
SVM dengan kernel RBF
svm_classifier = SVC(kernel='rbf', C=1.0, gamma='scale')
svm_classifier.fit(X_train, y_train)

# Memprediksi data uji
y_pred = svm_classifier.predict(X_test)

# Evaluasi kinerja model
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test,y_pred))
```

Hasil evaluasi model:

Tabel 3. Hasil Klasifikasi

	Prediksi Jinak	Prediksi Ganas
Aktual Jinak	85	5
Aktual Ganas	0	47

Tabel 4. Confussion Matrix

	Precision	Recall	F-1 score
Benign	1	0.94	0.97
Malignant	0.90	1	0.95
Accuracy			0.96

3.4 Analisis Hasil

Hasil dari eksperimen menunjukkan bahwa model *Support Vector Machine* dengan kernel RBF memberikan performa yang sangat baik dalam mengklasifikasikan sel kanker pada dataset. Beberapa poin penting dari hasil analisis adalah sebagai berikut:

1. *Confusion matrix* menunjukkan bahwa sebagian besar sampel diklasifikasikan dengan benar, dengan jumlah kesalahan yang relatif kecil. TP dan TN menunjukkan bahwa model mampu mengidentifikasi kedua label kelas dengan baik.
2. Hasil klasifikasi yang memberikan nilai *precision*, *recall*, dan *F1-score* yang tinggi, menunjukkan bahwa model *Support Vector Machine* mampu menangani model data ini.
3. Tingkat akurasi yang tinggi (96%) menunjukkan bahwa *Support Vector Machine* efektif dalam mengklasifikasikan sel kanker. Akurasi ini menunjukkan bahwa model mampu memprediksi dengan benar sebagian besar sampel dalam dataset uji.

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah terbukti bahwa SVM dengan kernel RBF merupakan metode yang efektif dalam klasifikasi sel kanker. SVM mampu memisahkan antara sel kanker jinak dan ganas dengan tingkat akurasi yang tinggi. Pengolahan data, termasuk pembersihan data dan normalisasi, telah terbukti penting dalam meningkatkan kinerja model SVM. Dengan memastikan data bersih dan memiliki skala yang seragam, model SVM dapat memberikan hasil yang lebih konsisten dan akurat, yang sangat diperlukan dalam konteks diagnosis kanker. Penggunaan teknik seperti *grid search* untuk mencari parameter optimal pada SVM telah terbukti efektif dalam meningkatkan kinerja model. Dengan mengoptimalkan parameter, model SVM dapat mencapai hasil yang lebih baik dalam mengklasifikasikan data.

Adapun beberapa saran terkait penelitian ini adalah, melakukan pengembangan eksperimen menggunakan berbagai metode klasifikasi lain seperti *Random Forest*, *K-Nearest Neighbors*, atau *Neural Networks* untuk

membandingkan kinerja dan keandalan model dalam berbagai konteks dan dataset. Pengembangan dapat juga dilakukan dari faktor dataset dan optimasi, pada dataset dapat dikembangkan dengan cara memakai dataset yang lebih besar dan lebih kompleks dari berbagai sumber untuk menguji dan memvalidasi model sehingga dapat menggali lebih jauh *insights* tentang karakteristik kanker yang kompleks, dan optimasi dapat dilakukan pada parameter *Support Vector Machine*, seperti teknik reduksi dimensi dan augmentasi data, untuk meningkatkan kinerja dan generalisasi model.

REFERENSI

- [1] L. L. Ferreira and A. D. Andricopulo, "Cancer Estimates in Brazil Reveal Progress for the Most Lethal Malignancies," *Current Topics in Medicinal Chemistry Volume 20, Issue 22*, pp. 1962 - 1966, 2020.
- [2] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor and F. Bray, "Cancer statistics for the year 2020: An overview," *International Journal of Cancer Volume 149, Issue 4*, pp. 778-789, 2021.
- [3] Ju, Wen, R. Zheng, S. Wang, S. Zhang, H. Zeng, R. Chen, K. Sun, L. Li and W. Wei, "The Occurrence of Cancer in Ageing Populations at Global and Regional Levels, 1990 To 2019," *Age and Ageing Cover Image for Volume 52, Issue 9*, 2023.
- [4] S. M. Schwartz, "Epidemiology of Cancer," *Clinical Chemistry, Volume 70, Issue 1*, p. 140–149, 2024.
- [5] M. F. Ak, "A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications," 2020, p. 111, 2020.
- [6] R. Nooreldeen and H. Bach, "Current and Future Development in Lung Cancer Diagnosis," *Int. J. Mol. Sci. 22(16)*, p. 8661, 2021.
- [7] S. M. Kurian, T. C. Whisenant and C. L. M. Marsh, "Toward Improved and Standardized Diagnostic Pipelines in Transplantation," *Transplantation 105(1)*, pp. 12-13, 2021.
- [8] B. Zhang, H. Shi and H. Wang, "Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach," *J Multidiscip Health 16*, pp. 1779-1791, 2023.
- [9] A. Site, J. Nurmi and E. S. Lohan, "Systematic Review on Machine-Learning Algorithms Used in Wearable-Based eHealth Data Analysis," *IEEE Access Volume 9*, pp. 112221-112235, 2021.
- [10] Y. Wang, W. Liao, H. Shen, Z. Jiang and J. Zhou, "Some Notes on The Basic Concepts of Support Vector Machines," *Journal of Computational Science Volume 82*, 2024.
- [11] W. Wolberg, "UCI Machine Learning Repository," 1990. [Online]. Available: <https://doi.org/10.24432/C5HP4Z>.
- [12] J. S. Pimentel, R. Ospina and A. Ara, "A Novel Fusion Support Vector Machine Integrating Weak And Sphere Models For Classification Challenges With Massive Data," *Decision Analytics Journal 11*, 2024.
- [13] E. Winarno, W. Hadikurniawati, A. Septiarini and H. Hamdani, "Analysis of Color Features Performance using Support Vector Machine with Multi-Kernel for Batik Classification," *International Journal of Advances in Intelligent Informatics Vol.8 No.2*, 2022.
- [14] K. M. M. Uddin, N. Biswas, S. T. Rikta and S. K. Dey, "Machine Learning-Based Diagnosis of Breast Cancer Utilizing Feature Optimization Technique," *Computer Methods and Programs in Biomedicine Update Volume 3*, 2023.
- [15] C. Dewi, F. A. Indriawan and H. J. Christanto, "Spam Classification Problems using Support Vector Machine and Grid Search," *International Journal of Applied Science and Engineering (IJASE) Volume 20, No. 4*, 2023.