

# PENDEKATAN PEMBELAJARAN MESIN UNTUK PAYUDARA DETEKSI DINI KANKER

MUAWIA A.ELSADIG

Universitas Imam Abdulrahman Bin Faisal, PO Box 1982, Dammam, Arab Saudi

Email: muawiasadig66@gmail.com

## ABSTRAK

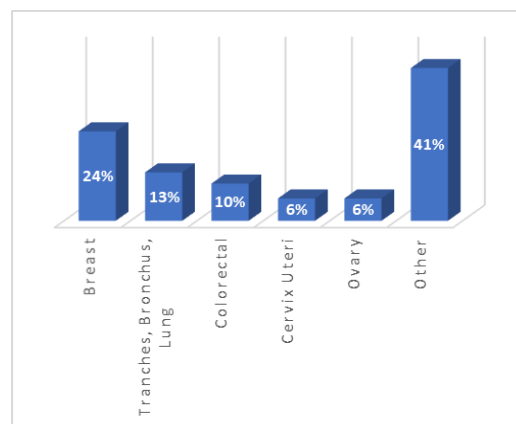
Peningkatan pesat dalam kejadian kanker payudara jelas terlihat. Penyebab penyakit ini tidak jelas dan alasan di balik peningkatan kejadian tidak diidentifikasi dengan baik. Selain itu, metode untuk mencegah terjadinya belum ditemukan. Oleh karena itu, deteksi dini memainkan peran utama dalam proses pengobatan dan membantu dalam mencapai tingkat kelangsungan hidup yang dapat diterima. Faktanya, ada banyak metode, berdasarkan pembelajaran mesin atau pendekatan statistik, untuk membedakan antara gambar jinak dan ganas. Namun, kebanyakan dari mereka tidak mencapai akurasi yang diinginkan, karena penggunaan fitur yang tidak akurat, tidak adanya pengklasifikasi yang tepat, atau kumpulan data yang tidak efisien. Karena itu, penelitian ini memperkenalkan pendekatan pengklasifikasi yang efektif berdasarkan support vector machine (SVM) dengan metode pemilihan fitur yang memadai yang hanya mempertimbangkan fitur dengan pengaruh tinggi dan mengabaikan fitur lainnya. Skenario ini berpotensi meningkatkan akurasi klasifikasi dan mengurangi overhead komputasi. Selain itu, penggunaan kumpulan data terpercaya dan penerapan metode validasi yang tepat tercermin dalam hasil yang andal dan terpercaya. Pemilihan SVM dilakukan setelah melakukan skenario eksperimental nyata untuk tujuh pengklasifikasi terkemuka di bidang diagnosis kanker payudara. Hasil eksperimen menunjukkan bahwa pendekatan classifier berbasis SVM mengungguli classifier lainnya dengan akurasi tertinggi mencapai 97,4%. Kontribusi makalah ini termasuk memperkenalkan pendekatan SVM yang efisien untuk memprediksi kanker payudara dan menyajikan studi perbandingan untuk tujuh pengklasifikasi populer di bidang ini. Hasil kami telah divalidasi secara menyeluruh untuk menominasikan SVM sebagai pengklasifikasi terbaik untuk deteksi kanker payudara.

Kata kunci: Penambangan data, Pembelajaran Mesin, Pembelajaran mendalam, Klasifikasi Gambar Medis, Kanker Payudara Deteksi, Diagnosis Kanker Payudara.

## 1. PERKENALAN

Saat ini, kanker payudara dianggap sebagai salah satu penyakit utama di dunia. Banyak survei telah menjelaskan jumlah orang yang meninggal karena kanker payudara [1]. Kematian akibat kanker payudara sangat tinggi dibandingkan dengan jenis kanker lainnya [2]. Pada tahun 2012, survei pemerintah AS menunjukkan bahwa 40 ribu orang telah meninggal karena penyakit ini [1]. Selain itu, 2 juta kasus baru didiagnosis pada tahun 2018 [3]. Alasan utama rendahnya tingkat kelangsungan hidup adalah karena keterlambatan penemuan penyakit ini dan komplikasi dari proses diagnosisnya. Oleh karena itu, deteksi dini memiliki dampak penting dalam mengurangi risiko dengan mencegah perkembangan penyakit dan mengurangi tingkat morbiditasnya [4, 5]. Hal ini memungkinkan pasien untuk mendapatkan pengobatan yang tepat [6]. Dubey dkk. telah jelas menyatakan bahwa deteksi dini kanker payudara dapat meningkatkan dan meningkatkan tingkat kelangsungan hidup hingga 98% untuk kasus tumor kecil dan 73% untuk kasus tumor besar [3]. Namun, deteksi dini seringkali merupakan tugas yang sulit karena tidak adanya gejala di awal [7].

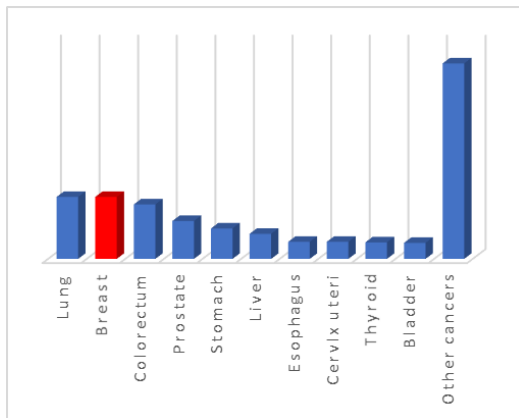
Kumari dkk. [1] menyatakan bahwa kanker payudara lebih dominan dibandingkan dengan jenis lainnya, seperti yang digambarkan pada Gambar 1, yang memberikan persentase kanker payudara dibandingkan dengan jenis kanker lainnya [1]. Ini mencerminkan penyebaran penyakit yang luas.



Gambar 1: Jenis Kanker [1]

Di sisi lain, laporan terbaru [8] menunjukkan jumlah kasus baru dan kematian pada tahun 2018 untuk berbagai jenis penyakit kanker. Menurut Ini

melaporkan, jumlah kasus baru dan kematian akibat kanker payudara masing-masing adalah 2.088.849 dan 626.679. Dalam hal jumlah kasus baru, kanker payudara menempati urutan kedua, sedangkan dalam hal jumlah kematian, kanker payudara adalah penyebab kematian kelima di seluruh dunia untuk kedua jenis kelamin, sedangkan untuk wanita kanker payudara adalah penyebab utama kematian dan telah jumlah kasus baru tertinggi dibandingkan dengan jenis kanker lainnya [8]. Gambar 2 dan 3 menunjukkan jumlah kasus baru dan kematian untuk jenis kanker selama tahun 2018, masing-masing. Di Amerika Serikat, jumlah kasus baru kanker payudara pada tahun 2019 adalah 271.270, sedangkan jumlah kematiannya adalah 42.260. Hal ini mencerminkan pentingnya mengembangkan metode yang efisien yang mendukung deteksi dini penyakit berbahaya ini.

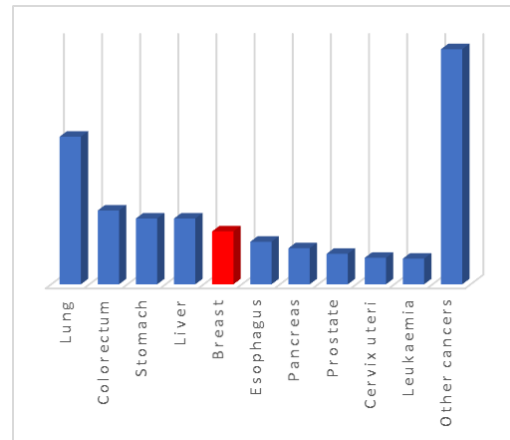


Gambar 2: Grafik menyajikan persentase kasus baru untuk sepuluh kanker umum di seluruh dunia pada tahun 2018 [8].

Analisis mikroskopis dari biopsi adalah salah satu metode terpenting yang digunakan untuk diagnosis kanker payudara. Namun, metode ini membutuhkan ahli patologi untuk melakukan analisis khusus. Analisis ini mahal dan memakan waktu dan sering mengarah pada hasil yang tidak sesuai kesepakatan [10]. Oleh karena itu, kebutuhan akan computer-aided diagnosis (CAD), seperti algoritma klasifikasi, lebih besar.

Diagnosis berbantuan komputer (CAD) membantu para ahli medis dalam diagnosis dini dan dengan demikian meningkatkan tingkat pemulihan. CAD menggunakan pendekatan pembelajaran mesin untuk memprediksi apakah tumor adalah kanker ganas atau jinak [11]. Ini adalah sistem berbasis komputer yang mendukung spesialis dalam mengambil keputusan dengan cepat [12, 13]. Pembaca yang tertarik menggunakan teknik komputer untuk mengekstrak informasi

dari berbagai sumber (gambar, teks, dll.) dirujuk ke [14,15]



Gambar 3: Grafik menyajikan persentase kasus kematian untuk sepuluh kanker umum di seluruh dunia pada tahun 2018 [8].

Perlu disebutkan bahwa teknik pembelajaran mesin memastikan kegunaannya dalam menemukan dan menentukan pola kumpulan data citra medis yang besar dan dengan demikian mengarah pada klasifikasi dan pemilihan citra yang berhasil. Namun, memilih pendekatan pembelajaran mesin yang memadai, kumpulan data tepercaya, metode pemilihan fitur yang efisien, dan pendekatan validasi yang akurat sangat berharga ketika mempelajari waktu bertahan hidup dari kanker payudara.

Makalah ini mengusulkan pendekatan pembelajaran mesin untuk deteksi dini kanker payudara. Pendekatan ini menggunakan algoritma klasifikasi yang memadai dan kumpulan data yang umum setelah menerapkan metode pemilihan fitur yang efisien yang mengarah pada pencapaian akurasi klasifikasi yang tinggi. Pemilihan algoritma klasifikasi kami dilakukan melalui perbandingan praktis yang didasarkan pada eksperimen nyata, termasuk menerapkan metode pemilihan fitur kami dengan metode pelatihan, pengujian, dan evaluasi yang sesuai. Eksperimen ini diimplementasikan pada tujuh pengklasifikasi yang sering digunakan dalam deteksi kanker payudara dan digunakan untuk mencapai akurasi klasifikasi yang kompeten.

Makalah ini memberikan tinjauan komprehensif tentang penyebaran kanker payudara dan metode kecerdasan buatan yang digunakan untuk mendeteksi penyakit tersebut, berkonsentrasi pada beberapa pendekatan pengklasifikasi yang umum digunakan di bidang ini. Inti dari makalah ini adalah untuk memperkenalkan pengklasifikasi yang tepat yang mampu mencapai tingkat akurasi tinggi untuk deteksi dini

penyakit ini. Makalah ini disusun sebagai berikut: bagian pertama memberikan pengenalan singkat tentang kanker payudara dan penyebarannya ke seluruh dunia. Ini menyoroti beberapa metode diagnostik berbantuan komputer dan kontribusi efektifnya terhadap proses diagnosis penyakit ini. Bagian selanjutnya memberikan tinjauan komprehensif tentang metode relatif untuk mendeteksi kanker payudara, pencapaian dan keterbatasannya, dengan lebih fokus pada tujuh pengklasifikasi umum yang memiliki reputasi baik untuk mendeteksi kanker payudara. Bagian 3 menunjukkan metode yang kami usulkan dan Bagian 4 menunjukkan rincian pekerjaan eksperimental kami, hasilnya dan metode yang diterapkan untuk memvalidasi hasil. Selain itu, bagian ini memberikan diskusi menyeluruh berdasarkan hasil eksperimen kami. Akhirnya, makalah ini disimpulkan di bagian 5, sementara bagian 6 merangkum pekerjaan kami di masa depan.

## 2. PEKERJAAN TERKAIT

Penggunaan pengklasifikasi pembelajaran mesin untuk mendukung proses diagnostik meningkat pesat di bidang medis. Faktanya, evaluasi data pasien dan keputusan ahli adalah pengaruh paling penting dalam proses diagnostik. Namun, sistem pakar dan pendekatan kecerdasan buatan untuk klasifikasi juga sangat mendukung para ahli. Pengklasifikasi pembelajaran mesin dapat membantu untuk memperkecil itu mungkin kesalahan menyebabkan oleh ahli yang belum berpengalaman, dan juga memberikan data medis yang detail dalam waktu singkat untuk diperiksa.

Faktanya, banyak masalah dengan menggunakan pendekatan pembelajaran mesin terkait dengan kurangnya validasi yang tepat dan efisien. Memang benar bahwa pendekatan pembelajaran mesin meningkatkan akurasi prediksi kelangsungan hidup. Namun, memilih pendekatan validasi yang memadai adalah nilai yang besar dalam mempelajari waktu bertahan hidup kanker payudara [16].

Gao dkk. [17] membandingkan metode tradisional vs. metode berbasis pembelajaran mesin yang digunakan dalam pencitraan payudara (mamografi). Tujuan mereka adalah untuk menekankan keterbatasan metode tradisional dan untuk menyoroti solusi potensial dari pendekatan deteksi berbantuan komputer (CAD). Mereka menyimpulkan bahwa pengembangan CAD mengalami pergerakan pandangan dunia berdasarkan kemajuan daya komputasi yang tak ada habisnya dan munculnya pendekatan pembelajaran mendalam yang cepat. Oleh karena itu, kami menyaksikan saat ini area yang menjanjikan untuk mengembangkan algoritme berbasis pembelajaran mesin yang memainkan peran penting dan efektif dalam meningkatkan klinis

sistem perawatan. Dalam konteks yang sama, Araújo et al. [4] menambahkan bahwa pendekatan pembelajaran mendalam merupakan alternatif yang efisien untuk pendekatan klasifikasi konvensional, karena mereka mampu mengatasi hambatan dari pendekatan berbasis fitur.

Abdar dkk. [2] menunjukkan bahwa sebagian besar pekerjaan penelitian pada analisis citra histopatologi kanker payudara dilakukan dengan menggunakan dataset kecil. Selain itu, kendala utama untuk mengembangkan metode analisis citra histopatologi baru adalah tidak adanya kumpulan data publik, besar, dan beranotasi [18]. Faktanya, anotasi adalah kunci penting untuk mengembangkan dan memvalidasi pendekatan berbasis pembelajaran mesin apa pun [2].

Sejumlah penelitian telah menyoroti kebutuhan untuk menggabungkan banyak pengklasifikasi bersama-sama, daripada menggunakan pengklasifikasi tunggal. Tren ini telah menyebabkan pendekatan klasifikasi yang efisien yang mampu mencapai akurasi klasifikasi yang dapat diterima. Pembaca yang tertarik dirujuk ke [19-21].

### 2.1 Mengapa pengklasifikasi ini?

Dalam subbagian ini, makalah ini memperkenalkan beberapa karya yang relevan pada tujuh pengklasifikasi umum yang telah sering digunakan dalam diagnosis kanker payudara. Pengklasifikasi ini mencerminkan kinerja tinggi dalam hal akurasi klasifikasi dan telah memberikan kontribusi yang baik untuk bidang ini. Apa yang telah kami nyatakan di sini di subbagian ini akan sepenuhnya mendukung gagasan kami untuk memilih pengklasifikasi ini sebagai pengklasifikasi yang paling umum. Jadi, pemilihan pengklasifikasi terbaik kami (melalui skenario eksperimental nyata) akan didasarkan pada asumsi yang solid, akurat, dan konkrit.

Baru-baru ini, Hosni dkk. [22] melaporkan bahwa pengklasifikasi yang paling sering digunakan untuk membangun pengklasifikasi ansambel adalah mesin vektor pendukung, jaringan saraf tiruan, dan pohon. Penulis dalam [23] juga menyatakan bahwa semakin banyak dilaporkan bahwa pengklasifikasi SVM memiliki kemampuan diagnosis yang akurat dan unggul [23].

Mendasarkan pekerjaan mereka pada dataset WBCD (Wisconsin Breast Cancer Database), Medjahed et al. menerapkan pengklasifikasi tetangga K-terdekat menggunakan aturan klasifikasi dan tipe jarak yang berbeda. Oleh karena itu, mereka memperoleh hasil bahwa dua jenis jarak, Euclidean dan Manhattan, lebih efektif dalam hal akurasi dan kinerja klasifikasi dibandingkan dengan jenis jarak lain yang diperiksa dalam penelitian mereka. Ini

dua jenis jarak memakan waktu, tetapi mereka masih memberikan hasil yang lebih baik [7].

Meenakshi dkk. [24] melakukan perbandingan antara dua jenis pengklasifikasi jaringan saraf, radial basis function (RBF) dan multilayer perceptron (MLP). Mereka melaporkan bahwa MLP mengungguli RBF dengan menghasilkan hasil yang lebih akurat dan spesifik. Selain itu, Daniele Soria dkk. [25] menguraikan bahwa MLP adalah pengklasifikasi yang paling efektif dalam penelitian mereka, yang dilakukan untuk membandingkan tiga pengklasifikasi yang berbeda: pengklasifikasi pohon C4.5, MLP dan naïve Bayes (NB). Katija dkk. merekomendasikan penggunaan MLP untuk memprediksi kelangsungan hidup kanker payudara. Hal ini didasarkan pada klaim mereka bahwa MLP membawa akurasi dan kinerja yang lebih baik [26].

Pengklasifikasi naïve Bayes adalah pengklasifikasi sederhana yang dibangun di atas teorema Bayes dengan asumsi independen. Ini mengasumsikan bahwa (i) atribut prediktif independen bersyarat dan (ii) nilai atribut numerik terdistribusi secara normal dalam setiap kelas [25]. Dibandingkan dengan pengklasifikasi MLP, Daniele et al. [25] menyatakan bahwa MLP adalah pengklasifikasi yang sangat efektif dengan kemampuan interpretasi yang buruk, sedangkan NB mencapai kinerja dan kemampuan interpretasi yang baik.

Moyana dkk. [27] melakukan studi perbandingan pada jaringan saraf tiruan dan pengklasifikasi regresi logistik. Mereka menyatakan bahwa pengklasifikasi ini adalah dua yang paling sering digunakan dalam estimasi risiko klinis. Kedua pendekatan tersebut digunakan untuk memperkirakan risiko kanker payudara dan menunjukkan kinerja yang serupa. Namun, kedua pendekatan memiliki kekuatan dan keterbatasan, yang harus dipertimbangkan. JST berguna ketika ada hubungan yang kompleks dan interaksi implisit dalam data, sedangkan regresi logistik direkomendasikan ketika kesimpulan statistik dari output perlu ditarik. Kedua pendekatan berpotensi membantu dokter dalam memahami faktor risiko kanker, estimasi risiko dan diagnosis [27]. Moyana dkk. dengan jelas menyatakan bahwa kedua pendekatan tersebut dapat digunakan secara saling melengkapi untuk membantu dalam pengambilan keputusan.

Random Forest adalah teknik pembelajaran mesin ensemble yang berjalan secara efisien pada database besar. Ini terdiri dari kumpulan pengklasifikasi terstruktur pohon [28].

Nguyen dkk. [29] mengembangkan model pengklasifikasi berdasarkan pengklasifikasi hutan acak dan teknik seleksi fitur untuk membantu dalam diagnosis kanker payudara. Model mereka, yang diterapkan pada dua kumpulan data yang berbeda, telah memperoleh akurasi klasifikasi yang tinggi dan dianggap sebagai hasil yang sangat menjanjikan di bidang ini. Hasil yang menjanjikan ini memotivasi kami untuk memasukkan pengklasifikasi hutan acak sebagai salah satu pengklasifikasi yang diteliti dalam makalah ini.

Sebuah studi baru-baru ini melakukan percobaan dengan menerapkan empat metode seleksi yang berbeda dan empat pengklasifikasi umum. Percobaan ini dilakukan pada empat dataset yang berbeda. Saat menerapkan seleksi fitur, hasilnya menunjukkan bahwa pendekatan jaringan saraf tiruan mengungguli pendekatan lain (Naïve Bayes, Support Vector Machines dan Decision Trees) dengan mencapai peningkatan yang signifikan dalam akurasi klasifikasi kanker payudara, sementara tidak ada peningkatan akurasi saat menerapkan bagian fitur dengan pengklasifikasi lainnya [11].

Oleh karena itu, dan berdasarkan studi-studi yang relevan di atas, telah dipilih tujuh pengklasifikasi yang telah berhasil memastikan efisiensinya dalam mencapai akurasi klasifikasi yang baik untuk diteliti dalam makalah ini. Ketujuh pengklasifikasi tersebut adalah ANN, SVM, Random Forest, Logistic Regression, NB, KNN dan DT.

### 3. METODE YANG DIUSULKAN

Makalah ini telah memilih tujuh pengklasifikasi paling populer yang digunakan dalam deteksi kanker payudara. Pengklasifikasi ini telah dilatih dan diuji menggunakan kumpulan data populer dan tepercaya setelah menerapkan metode pemilihan fitur yang efisien untuk kumpulan data ini (memilih fitur penting yang memiliki pengaruh paling besar akan sangat meningkatkan deteksi ketepatan). Kemudian pengklasifikasi yang mengungguli lain akan dipertimbangkan dan diusulkan sebagai pendekatan pengklasifikasi untuk makalah ini. Pengklasifikasi yang dipilih adalah: ANN, SVM, Random Forest, Logistic Regression, NB, KNN dan DT. Pembelaan untuk memilih pengklasifikasi ini dengan jelas dinyatakan di bagian sebelumnya. Setiap pengklasifikasi dibangun secara terpisah, dan karenanya hasil yang diperoleh dicatat dan dilaporkan.

#### 3.1 Kumpulan Data

Makalah ini menggunakan dataset di seluruh dunia dan umum untuk pelatihan dan pengujian semua pengklasifikasi yang digunakan dalam pekerjaan ini. Ini menggunakan dataset Kanker Payudara Wisconsin dari Mesin UCI

Learning Repository, yang memiliki 699 instance numerik dan 30 fitur. Dataset ini telah disempurnakan dengan menggunakan metode pemilihan fitur yang membantu mengurangi jumlah fitur dengan mempertimbangkan hanya fitur yang memiliki pengaruh tinggi dan mengabaikan fitur yang redundan dan tidak berbobot. Subbagian di bawah ini memberikan rincian lebih lanjut dalam hal ini. Dalam tulisan ini, dataset telah dipartisi menjadi tiga set untuk tujuan pelatihan dan pengujian, sebagai berikut: (i) set 1: 70% untuk pelatihan dan 30% untuk pengujian (ii) set 2: 80% untuk pelatihan dan 20% untuk pengujian (iii) set 3: 90% untuk pelatihan dan 10% untuk pengujian. Setiap classifier dibangun, dilatih dan diuji dengan menerapkan set pelatihan-pengujian ini. Menerapkan tiga set ukuran pelatihan yang berbeda ini dimaksudkan untuk mencerminkan hasil yang akurat dan andal.

	A	B	C	D	E
1	texture_worst	radius_worst	perimeter_worst	perimeter_mean	radius_
2	17.33	25.38	184.6		122.8
3	23.41	24.99	158.8		132.9
4	25.53	23.57	152.5		130
5	26.5	14.91	98.87		77.58
6	16.67	22.54	152.2		135.1
7	23.75	15.47	103.4		82.57
8	27.66	22.88	153.2		119.6
9	28.14	17.06	110.6		90.2

Gambar 4: Bagian dari kumpulan data yang direvisi dengan sepuluh fitur.

### 3.2 Pilihan Fitur:

Seleksi fitur adalah bagian penting dalam membangun pendekatan klasifikasi yang efektif. Penurunan jumlah fitur yang memadai pasti dapat meningkatkan prediktabilitas pengklasifikasi dan menghasilkan sistem klasifikasi yang kurang intensif secara komputasi. Di bidang klasifikasi citra medis, membatasi jumlah fitur mengurangi biaya dan waktu diagnostik. Dalam literatur yang relevan, ada banyak metode dan algoritma yang digunakan untuk pemilihan fitur, untuk mendukung dan meningkatkan akurasi klasifikasi dan mengurangi biaya sistem. Dalam pendekatan kami, algoritma pemilihan fitur relief digunakan. Algoritma ini dirancang untuk sistem klasifikasi biner. Untuk setiap fitur, algoritma relief menghitung skor fitur. Demikian, fitur diberi peringkat untuk memilih fitur teratas yang memiliki dampak paling efektif pada proses klasifikasi. Berdasarkan algoritma pemilihan fitur di atas, hanya sepuluh fitur yang dipilih. Gambar 4 menunjukkan bagian dari dataset yang disempurnakan dengan fitur yang dipilih.

### 3.3 Metode Validasi:

Dalam penelitian ini, teknik validasi random sampling diterapkan pada setiap percobaan. Teknik ini diulang 20 kali untuk mencapai hasil yang andal dan realistis. Kinerja pengklasifikasi dievaluasi menggunakan matriks konfusi, yang digunakan untuk mengevaluasi kesalahan klasifikasi, positif palsu dan negatif palsu.

Akurasi dihitung dengan menggunakan persamaan yang ditunjukkan pada (1).

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

Di mana:

True positives (TP): mengacu pada gambar positif yang diklasifikasikan dengan benar.

Negatif sejati (TN): mengacu pada gambar negatif yang diklasifikasikan dengan benar.

Positif palsu (FP): mengacu pada gambar negatif yang diklasifikasikan secara tidak benar.

Negatif palsu (FN): mengacu pada gambar positif yang diklasifikasikan secara tidak benar [30, 32].

Akhirnya, tujuh pengklasifikasi diberi peringkat sesuai dengan pencapaian akurasi mereka.

### 3.4 Alat Implementasi Pengklasifikasi

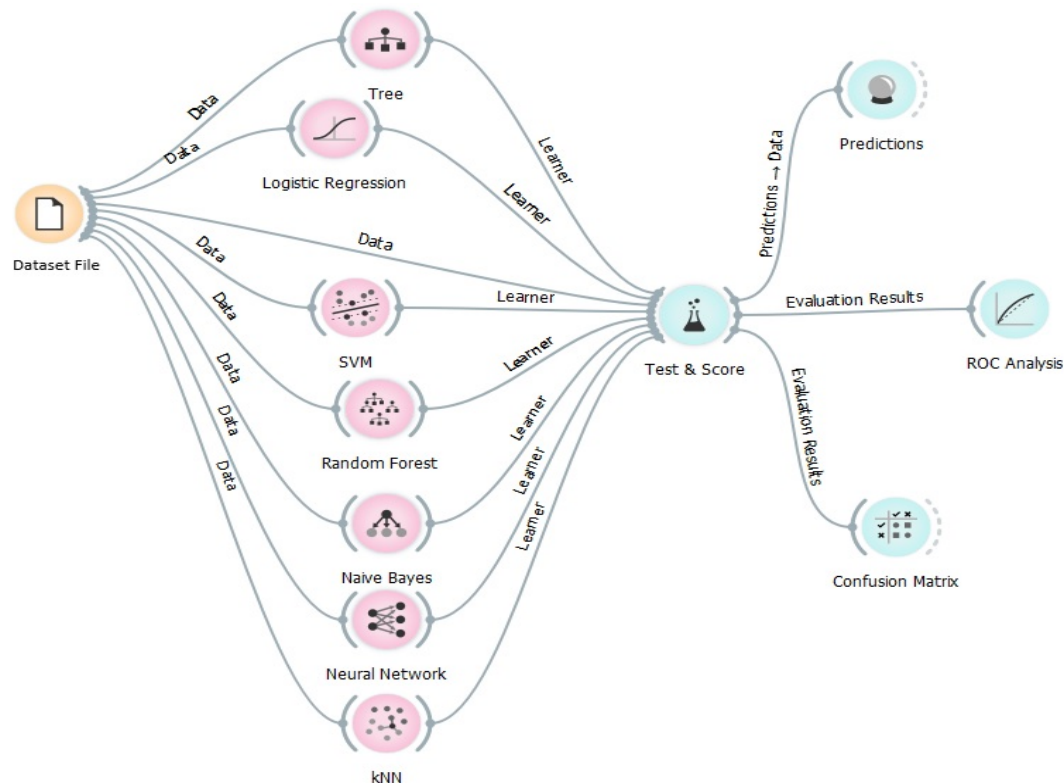
Dalam penelitian ini, semua percobaan telah dilakukan dengan menggunakan software data mining Orange. Ini adalah open source dan alat pembelajaran mesin yang sangat berguna untuk menerapkan algoritma klasifikasi yang berbeda dan untuk memvisualisasikan data Anda secara efisien. Gambar 5 menunjukkan salah satu skenario percobaan kami menggunakan perangkat lunak Orange.

### 3.5 Faktor Keberhasilan

Dalam makalah ini, ada lebih dari satu faktor yang sangat meningkatkan akurasi pengklasifikasi atau metode validasi yang diterapkan dan dengan demikian mendukung temuan kami. Faktor-faktor ini meliputi:

- Memilih pengklasifikasi yang paling umum dan populer untuk ditempatkan dalam penyelidikan. Menggunakan
- kumpulan data internasional terpercaya dari sumber terpercaya.

- Menerapkan metode pemilihan fitur yang memadai untuk memilih hanya fitur dengan pengaruh terbesar, yang meningkatkan akurasi dan mengurangi waktu pemrosesan. Menerapkan set yang berbeda dari pelatihan-pengujian.
- Menggunakan teknik validasi random sampling.
- Teknik pengambilan sampel acak ini diulang 20 kali untuk setiap percobaan untuk mencapai hasil yang andal dan realistis.



Gambar 5: Implementasi Classifier menggunakan software Orange

#### 4. HASIL DAN PEMBAHASAN EKSPERIMENTAL

Dalam makalah ini, tujuh pengklasifikasi umum untuk kanker payudara telah dipilih untuk memilih salah satu yang mencapai kinerja tertinggi dalam hal akurasi klasifikasi dibandingkan dengan yang lain, setelah menerapkan berbagai faktor dan teknik untuk memperkaya akurasi klasifikasi. Pemilihan pengklasifikasi ini didasarkan pada reputasi baik mereka dalam domain klasifikasi, dan terutama dalam diagnosis kanker payudara.

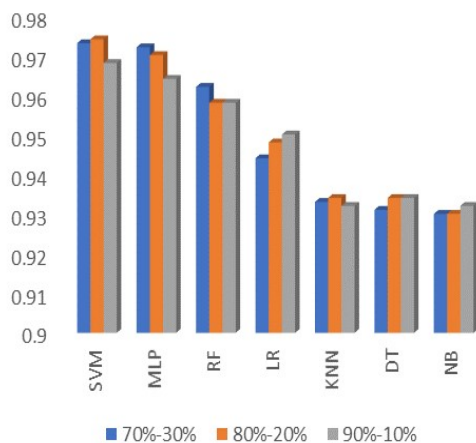
Sesuai metode kami yang disebutkan di bagian sebelumnya - yang mencakup akuisisi dan prapemrosesan data, pemilihan fitur, pengklasifikasi bangunan, pelatihan pengklasifikasi, pengujian dan validasi pengklasifikasi - semua eksperimen telah diimplementasikan sesuai dengan itu. Tabel 1 menunjukkan hasil akurasi untuk tujuh pengklasifikasi selama tiga set pelatihan yang berbeda (70%:30%, 80%:20% dan 90%:10%). Jelas terbukti bahwa SVM telah mengungguli pengklasifikasi lain dengan mencapai hasil akurasi tertinggi: 97,3%, 97,4% dan 96,8% selama tiga percobaan yang berbeda dari set pelatihan-pengujian, 70% - 30%, 80% - 20%, dan 90%-10%, masing-masing.



Tabel 1 Perbandingan Akurasi Tujuh Pengklasifikasi

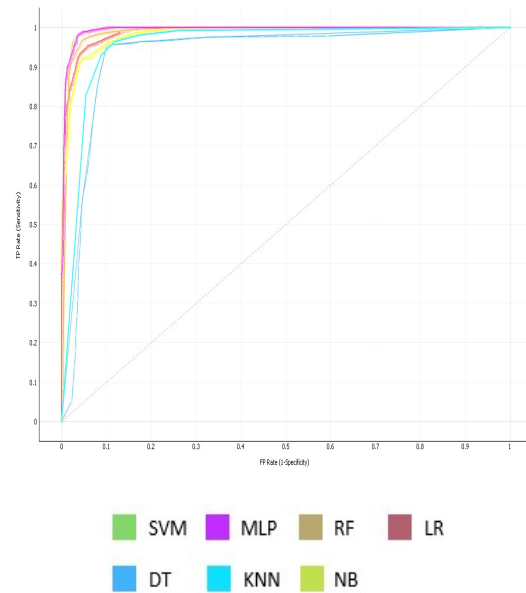
Penggolong	Ukuran pelatihan 70%		Ukuran pelatihan 80%		Ukuran pelatihan 90%	
	Ketepatan					
SVM	0,973		SVM	0,974	SVM	0,968
MLP	0,972		MLP	0,970	MLP	0,964
RF	0,962		RF	0,958	RF	0,958
LR	0,944		LR	0,948	LR	0,950
KNN	0,933		KNN	0,934	DT	0,934
DT	0,931		DT	0,934	KNN	0,932
catatan	0,930		catatan	0,930	catatan	0,932

Sebuah presentasi grafis dari temuan kami juga diberikan pada Gambar. 6, yang secara jelas mencerminkan persentase akurasi dari pengklasifikasi diselidiki yang dicapai di bawah tiga tolak ukur yang berbeda dari pelatihan-tes.

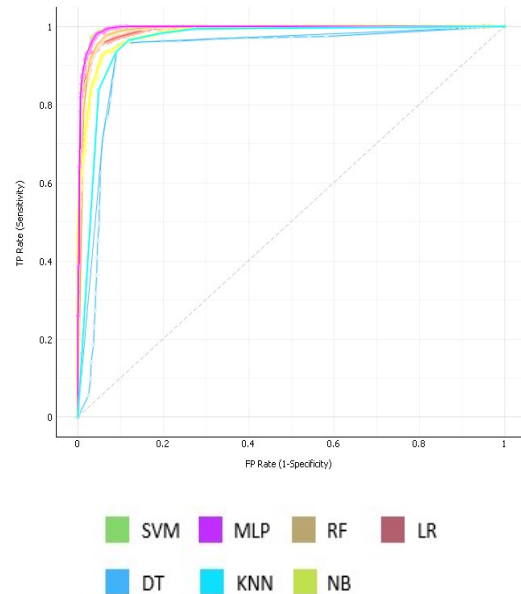


Gambar 6: Representasi grafis dari akurasi klasifikasi yang dicapai oleh tujuh pengklasifikasi di bawah yang berbeda set pelatihan-pengujian

Gambar lainnya - Gambar 7, 8 dan 9 - menggambarkan kurva Receiver Operating Characteristic (ROC) untuk tujuh pengklasifikasi yang diselidiki di bawah tiga set pengujian pelatihan yang berbeda: 70% - 30%, 80% -20%, dan 90% -10%, masing-masing.

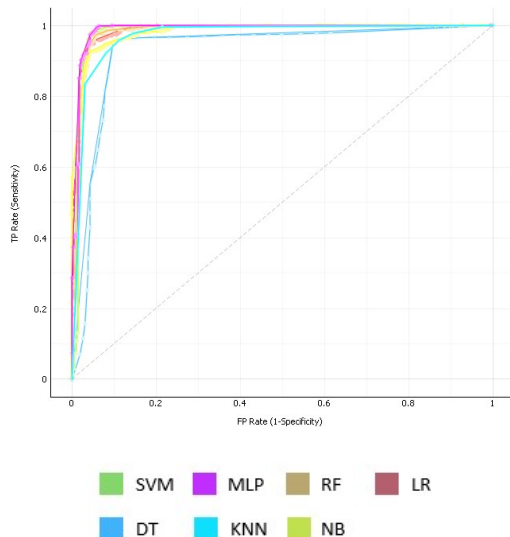


Gambar 7: Kurva ROC: 70% -30% pengujian pelatihan skenario



Gambar 8: Kurva ROC: 80% -20% pelatihan-pengujian skenario

Angka-angka ini mencerminkan bahwa, dalam hal spesifisitas dan sensitivitas, SVM adalah yang terbaik dibandingkan dengan pengklasifikasi lainnya.



Gambar 9: Kurva ROC: 90% -10% pengujian pelatihan skenario

Berdasarkan temuan ini, SVM akan disarankan sebagai pengklasifikasi terbaik untuk deteksi dini penyakit semacam ini. Namun, lebih banyak pekerjaan yang mempertimbangkan kumpulan data yang berbeda, metode pemilihan fitur yang berbeda, parameter pengujian pelatihan yang berbeda, dll. diperlukan untuk mendukung temuan ini.

## 5. KESIMPULAN

Setelah tinjauan menyeluruh dan eksperimen efisien yang didasarkan pada penerapan metode pemilihan fitur yang memadai, makalah ini telah memperkenalkan pendekatan untuk memprediksi kanker payudara menggunakan SVM. Pendekatan yang diusulkan telah berhasil divalidasi untuk mencapai akurasi tinggi, mencapai 97,4%. Ini mengungguli enam pengklasifikasi umum lainnya yang memiliki reputasi baik di bidang ini.

Juga, makalah ini menyajikan studi perbandingan dan penilaian pengklasifikasi paling umum teratas dalam hal akurasi klasifikasi mereka dan melaporkan pencapaian mereka. Kekuatan pendekatan yang diusulkan bergantung pada tiga fase peningkatan sebagai berikut:

Tahap 1: Menerapkan metode seleksi fitur yang memadai adalah cara paling efektif untuk meningkatkan akurasi klasifikasi. Menerapkan metode ini

mengurangi jumlah fitur kumpulan data. Mengurangi jumlah fitur – dengan memilih fitur berbobot dan mengabaikan fitur yang memiliki bobot ringan – memperkaya kinerja pengklasifikasi dalam hal akurasi klasifikasi dan juga mengurangi overhead dengan mengurangi biaya dan waktu komputasi. Cara paling efektif untuk meningkatkan akurasi klasifikasi adalah dengan memilih fitur klasifikasi yang tepat.

Tahap 2: Pemilihan pengklasifikasi teratas yang tepat, yang didasarkan pada reputasi baik mereka di bidang ini dan pada akurasi klasifikasi yang mereka capai, yang dihitung melalui eksperimen yang dilakukan menggunakan versi yang disempurnakan dari kumpulan data yang andal dan tepercaya. Dataset tanpa fitur yang tidak relevan, berisik, dan berlebihan.

Tahap 3: Penggunaan teknik pengambilan sampel acak yang diulang 20 kali untuk mencapai hasil yang andal dan realistis.

Akhirnya, kita dapat menyimpulkan bahwa penelitian ini berkontribusi dengan memperkenalkan pendekatan pengklasifikasi yang efektif berdasarkan mesin vektor dukungan (SVM) dengan metode pemilihan fitur yang memadai yang hanya mempertimbangkan fitur dengan pengaruh tinggi dan mengabaikan yang lain. Pengurangan fitur yang tidak relevan, berisik, dan berlebihan berpotensi meningkatkan akurasi model klasifikasi kami.

Selain itu, makalah ini telah menyajikan perbandingan eksperimental dari tujuh pengklasifikasi teratas yang merupakan kontribusi penting lainnya untuk pekerjaan ini.

## 6. PEKERJAAN MASA DEPAN

Pekerjaan kami di masa mendatang akan mencakup skenario implementasi yang berbeda di bawah kumpulan data yang berbeda menggunakan SVM.

## PENGAKUAN

Karya ini merupakan bagian dari proyek kami No. 2019-348- DSR yang didukung oleh Dekan Riset Ilmiah IAU. Oleh karena itu, penulis sangat menghargai dorongan dan dukungan mereka.

## REFERENSI:

- [1] N. Kumari dan K. Verma, "Sebuah Survei Pada Berbagai Pendekatan Pembelajaran Mesin yang Digunakan Untuk Deteksi Kanker Payudara," *International Journal of Advanced Research in Computer Science*, vol. 10, tidak. 3, hal. 76, 2019.



- [2] FA Spanhol, LS Oliveira, C. Petitjean, dan L. Heutte, "Sebuah dataset untuk klasifikasi citra histopatologi kanker payudara," IEEE Transactions on Biomedical Engineering, vol. 63, tidak. 7, hlm. 1455-1462, 2015.
- [3] AK Dubey, U. Gupta, dan S. Jain, "A Survey on Breast Cancer Scenario and Prediction Strategy," 2015: Springer, hlm. 367-375.
- [4] T. Araújo et al., "Klasifikasi gambar histologi kanker payudara menggunakan jaringan saraf convolutional," PloS satu, vol. 12, tidak. 6, hal. e0177544, 2017.
- [5] RA Smith, V. Cokkinides, dan HJ Eyre, "Pedoman American Cancer Society untuk deteksi dini kanker, 2004," CA: jurnal kanker untuk dokter, vol. 54, tidak. 1, hlm. 41-52, 2004.
- [6] S. Khan, N. Islam, Z. Jan, IU Din, dan JJP C. Rodrigues, "Sebuah kerangka kerja berbasis pembelajaran mendalam baru untuk deteksi dan klasifikasi kanker payudara menggunakan pembelajaran transfer," Surat Pengenalan Pola, vol. 125, hlm. 1-6, 2019.
- [7] SA Medjahed, TA Saadi, dan A. Benyettou, "Diagnosis kanker payudara dengan menggunakan k-nearest neighbor dengan jarak dan aturan klasifikasi yang berbeda," International Journal of Computer Applications, vol. 62, tidak. 1, 2013.
- [8] F. Bray, J. Ferlay, I. Soerjomataram, RL Siegel, LA Torre, dan A. Jemal, "Statistik kanker global 2018: GLOBOCAN perkiraan kejadian dan kematian di seluruh dunia untuk 36 kanker di 185 negara," CA: a jurnal kanker untuk dokter, vol. 68, tidak. 6, hlm. 394-424, 2018.
- [9] MD Bacolod et al., "Prediksi biomarker berbasis darah dan desain selanjutnya dari uji bisulfite PCR-LDR-qPCR untuk kanker payudara deteksi," kanker BMC, vol. 20, no. 1, hal. 85, 2020.
- [10] G. Aresta et al., "Bach: Tantangan besar pada gambar histologi kanker payudara," Analisis gambar medis, 2019.
- [11] N. Lindqvist and T. Price, "Evaluation of Feature Selection Methods for Machine Learning Classification of Breast Cancer," ed, 2018.
- [12] Q. Li dan RM Nishikawa, dengan bantuan komputer deteksi dan diagnosis dalam pencitraan medis. Taylor & Francis, 2015.
- [13] K. Doi, "Diagnosis berbantuan komputer dalam pencitraan medis: tinjauan sejarah, status saat ini dan potensi masa depan," Pencitraan dan grafik medis terkomputerisasi, vol. 31, tidak. 4-5, hlm. 198-211, 2007.
- [14] M. Abdelmagid, A. Ahmed, dan M. Himmat, "Metode Ekstraksi Informasi dan Teknik Ekstraksi dalam Isi Dokumen Kimia: Survei," ARPN Journal of Engineering and Applied Sciences, vol. 10, tidak. 3, 2015.
- [15] M. Abdelmagid et al., "Survei tentang Ekstraksi Informasi dari Literatur Senyawa Kimia: Teknik dan Tantangan," Jurnal Teknologi Informasi Teoretis dan Terapan, vol. 67, tidak. 2, 2014.
- [16] I. Mihaylov, M. Nisheva, dan D. Vassilev, "Penerapan Model Pembelajaran Mesin untuk Prognosis Kelangsungan Hidup dalam Studi Kanker Payudara," Informasi, vol. 10, tidak. 3, hal. 93, 2019.
- [17] Y. Gao, KJ Geras, AA Lewin, dan L. Moy, "New Frontiers: An Update on Computer-Aided Diagnosis for Breast Imaging in the Age of Artificial Intelligence," American Journal of Roentgenology, vol. 212, tidak. 2, hlm. 300-307, 2019.
- [18] M. Veta, JPW Pluim, PJ Van Diest, dan M. A. Viergever, "Analisis gambar histopatologi kanker payudara: Tinjauan," IEEE Transactions on Biomedical Engineering, vol. 61, tidak. 5, hal. 1400-1411, 2014.
- [19] M. Abdar et al., "Teknik ansambel bersarang baru untuk diagnosis otomatis kanker payudara," Pattern Recognition Letters, vol. 132, hlm. 123-131, 2020.
- [20] D. Saraswathi dan E. Srinivasan, "Sebuah pendekatan ansambel untuk mendiagnosis kanker payudara menggunakan pengklasifikasi jaringan saraf relaksasi bernilai kompleks sepenuhnya," International Journal of Biomedical Engineering and Technology, vol. 15, tidak. 3, hlm. 243-260, 2014.
- [21] S. Lee, M. Amgad, M. Masoud, R. 2020. Subramanian, D. Gutman, dan L. Cooper, "Sebuah Pembelajaran Aktif Berbasis Ensemble untuk Klasifikasi Kanker Payudara," 2019: IEEE, hlm. 2549-2553.
- [22] M. Hosni, I. Abnane, A. Idri, JMC de Gea, dan JLF Alemán, "Reviewing Ensemble Classification Methods in Breast Cancer," Computer methods and programs in biomedicine, 2019.
- [23] MF Akay, "Mendukung mesin vektor dikombinasikan dengan pemilihan fitur untuk diagnosis kanker payudara," Sistem pakar dengan aplikasi, vol. 36, no. 2, hlm. 3240-3247, 2009.
- [24] C. Meenakshi, M. Govindarajan dan AM Sameeullah, "Deteksi Kanker Payudara Menggunakan

- Pengklasifikasi MLP dan RBF," IMS Manthan, 5 (1), hlm. 35 – 38, 2010.
- [25] D. Soria, JM Garibaldi, E. Biganzoli, dan I. O. Ellis, "Perbandingan tiga metode berbeda untuk klasifikasi data kanker payudara," 2008: IEEE, hlm. 619-624.
- [26] A. Kathija, SS Nisha, dan MM Sathik, "Klasifikasi data kanker payudara menggunakan pendekatan jaringan saraf dari algoritma MLP," International Journal of Trend dalam Penelitian dan Pengembangan, vol. 4, tidak. 3, hlm. 275-279, 2017.
- [27] T. Ayer, J. Chhatwal, O. Alagoz, CE Kahn Jr, RW Woods, dan ES Burnside, "Perbandingan regresi logistik dan model jaringan saraf tiruan dalam estimasi risiko kanker payudara," Radiografi, vol. 30, tidak. 1, hlm. 13-22, 2010.
- [28] K. Vy dan PK Sinha, "Pengklasifikasi hutan acak: survei dan arah penelitian masa depan," Int. J. Adv. Hitung, vol. 36, tidak. 1, hal. 1144-1153, 2013.
- [29] C. Nguyen, Y. Wang, dan HN Nguyen, "Pengklasifikasi hutan acak dikombinasikan dengan seleksi fitur untuk diagnosis dan prognostik kanker payudara," Journal of Biomedical Science and Engineering, vol. 6, tidak. 05, hal. 551, 2013.
- [30] MA Elsadig dan YA Fadlalla, "Panjang Paket Saluran Terselubung: Skema Deteksi," 2018: IEEE, hlm. 1-7.
- [31] MA Elsadig, "Menyelesaikan Saluran Terselubung Panjang Paket Jaringan," Diss. Universitas Sains & Teknologi Sudan, 2018.
- [32] A. Ahmed, OM Barukab dan MA Elsadig, "Metode Multi-Classifer Heterogen berdasarkan Voting Tertimbang untuk Deteksi Kanker Payudara," International Journal of Advances in Science Engineering and Technology, vol. 7, tidak. 4, 2019.