

## Klasifikasi tumor payudara jinak dan ganas pada citra ultrasonografi (USG) berdasarkan karakteristik tekstur menggunakan metode random forest

Aida Rahmawati\*, Ice Uliya Sari, Heni Sumarti

Jurusan Fisika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang

\*email: 2108026003@student.walisongo.ac.id

**Abstrak** – Penyakit kanker payudara terus meningkat setiap tahun di Indonesia. Saat ini, kanker payudara tidak hanya menyerang pada usia lanjut, tetapi juga pada usia muda. Beberapa penelitian menunjukkan bahwa klasifikasi kanker payudara dapat dilakukan dengan menggunakan ultrasonografi. Tujuan penelitian mengeksplorasi penggunaan metode Random Forest untuk mengklasifikasikan kanker payudara pada citra ultrasonografi (USG), berdasarkan karakteristik tekstur. Hasil uji data pelatihan diperoleh akurasi sebesar 100% dan uji coba dengan variasi folds diperoleh akurasi antara 51% sampai 54%. Meskipun belum optimal, penggunaan metode Random Forest menunjukkan potensi sebagai algoritma klasifikasi, untuk diagnosis tumor payudara. Faktor seperti pemilihan fitur, kualitas data set, dan parameter model menjadi pertimbangan penting untuk penelitian lanjut guna mendukung diagnosis yang lebih akurat. oleh tenaga kesehatan.

**Kata Kunci:** ultrasonografi, tumor payudara, karakteristik tekstur, random forest.

**Abstract** – The incidence of breast cancer continues to increase every year in Indonesia. Currently, breast cancer affects not only older individuals but also younger ones. Several studies indicate that breast cancer classification can be performed using ultrasonography. This research aims to explore the use of the Random Forest method for classifying breast cancer in ultrasound (USG) images based on texture characteristics. The training data test results showed an accuracy of 100%, and trials with varying folds achieved an accuracy between 51% and 54%. Although not yet optimal, the use of the Random Forest method demonstrates potential as a classification algorithm for breast tumor diagnosis. Factors such as feature selection, data set quality, and model parameters are important considerations for further research to support more accurate diagnosis by healthcare professionals.

**Key words:** ultrasonography, breast cancer, texture characteristic, random forest.

## PENDAHULUAN

Kanker adalah kondisi penyakit yang dicirikan oleh pertumbuhan sel yang tidak terkendali dan berpotensi menyerang jaringan tubuh. Dalam satu tahun terakhir, Indonesia mengalami 348.809 kasus baru kanker. Terdiri dari 58.000 kasus kanker payudara, 32.000 kasus kanker leher rahim, dan 30.000 kasus kanker usus besar, dengan total kematian akibat kanker diangka 207.000 [1]. Terdapat 5 jenis kanker di dunia yang paling besar yaitu; kanker hati (*liver cancer*), kanker payudara (*breast cancer*), kanker lambung (*stomach cancer*), kanker usus besar (*colorectal cancer*), dan kanker paru-paru (*lung cancer*). Hasil survei yang dilakukan oleh Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa antara 8-9% wanita menderita kanker payudara.

Kanker merupakan salah satu penyebab kematian paling umum. hingga Sampai saat ini penderita kanker mencapai 8,2 juta orang

meninggal karena kanker [2]. Tingkat kesembuhan dan angka kematian meningkat ketika tumor payudara jinak dan ganas ditemukan dan diklasifikasikan dengan benar. Salah satu metode diagnostik yang paling umum untuk mengevaluasi kelainan pada payudara adalah citra ultrasonografi (USG).

Pencitraan gelombang suara ultrasonik atau yang biasa disebut USG adalah metode pencitraan medis yang menggunakan gelombang suara berfrekuensi ultrasonik untuk menggambarkan struktur jaringan dan organ tubuh. Beberapa kelebihan dari USG mencakup ekonomis, portabilitas yang baik, aman, akurat, aman serta tanpa mengeluarkan radiasi berbahaya. Identifikasi hasil melalui interpretasi manual gambar USG bergantung pada diagnosis dokter serta operator. Ada potensi kesalahan yang mungkin terjadi dalam proses manual ini. Sebagai langkah signifikan dalam diagnosis, klasifikasi, dan deteksi dini kanker, penting

untuk dilakukan analisis gambar USG menggunakan komputer [3].

Ultrasonografi menggunakan frekuensi suara cukup tinggi yaitu berkisar antar 1-15MHz [4]. Pemeriksaan USG dapat memberikan hasil optimal dalam berbagai pemeriksaan organ tubuh, memberikan sejumlah keunggulan baik bagi pasien maupun para sonografer atau radiolog [5]. Diagnosis tumor dapat dilakukan dengan pendekatan triple test, yaitu melalui pemeriksaan klinis, mamografi, dan sitologi aspirasi. Karena tidak semua wilayah memiliki fasilitas mamografi, ultrasonografi (USG) dapat digunakan sebagai alternatif [6]. Kelebihan USG dibandingkan dengan metode pencitraan diagnostik lainnya tidak melibatkan penggunaan radiasi pengion, memiliki biaya yang relatif terjangkau, bersifat non-invasif, serta memberikan kemudahan dan kecepatan dalam pelaksanaannya [7].

Saat ini hasil pemeriksaan USG memiliki tingkat akurasi diagnosis yang rendah karena seringkali terdapat perbedaan antar dokter radiologi mengenai hasilnya. Situasi ini diperkirakan bahwa meningkatnya kompleksitas kemungkinan terjadi pada citra USG karena citra tersebut rentan terhadap gangguan speckle yang dapat menurunkan kualitas visual. Gangguan speckle pada citra USG timbul akibat interferensi antara gelombang yang dipantulkan dan gelombang yang tiba di probe transduser, sehingga citra memiliki pola tekstur berbintik yang kasar [8]. Selain itu, citra ultrasonografi juga mengandung gangguan semacam "salt and pepper," yang muncul sebagai penanda pada gambar [9]. Hasil citra dari ultrasonografi dijelaskan berdasarkan sistem kategori BIRADS (Breast Imaging, Reporting & Data System). BIRADS merupakan suatu sistem yang mengklasifikasikan hasil pemeriksaan *screening* (pencitraan untuk mendeteksi kanker payudara) ke dalam kategori yang telah ditentukan secara jelas. [10]. Guna mengatasi permasalahan tersebut dapat diatasi menggunakan metode Radom Forest

Random Forest, salah satu algoritma pembelajaran mesin, dapat menangani variabilitas dan kompleksitas dalam data gambar. Kita dapat membedakan tumor payudara jinak dan ganas dengan melihat karakteristik tekstur pada gambar USG [11]. Random Forest adalah suatu koleksi dari decision trees yang dibuat dari sampel yang dipilih secara acak, namun dengan aturan pembagian simpul yang berbeda. Dalam model

ini, setiap pohon menggunakan subset fitur dan berusaha menemukan ambang batas terbaik untuk membagi data. Akibatnya, banyak pohon dilatih dengan cara yang lebih lemah, dan masing-masing pohon menghasilkan prediksi yang berbeda [12].

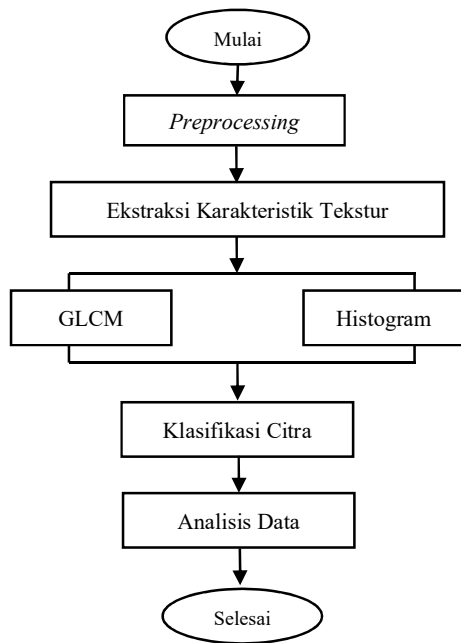
Peluang baru untuk meningkatkan akurasi klasifikasi tumor payudara telah muncul berkat kemajuan teknologi. Analisis citra berbasis karakteristik tekstur, yang menggunakan algoritma pembelajaran mesin dan teknik pengolahan gambar merupakan salah satu metode yang banyak digunakan. Dalam penelitian ini, metode Random Forest akan digunakan untuk mengklasifikasikan tumor payudara jinak dan ganas pada gambar USG berdasarkan karakteristik tekstur.

Banyak dilakukan penelitian terkait tumor payudara, salah satu penelitian dilakukan oleh Arwoko [13] mengenai klasifikasi tumor payudara jinak dan ganas menggunakan fitur koefisien *Discrete Cosine Transform* (DCT). Tingkat akurasi yang dihasilkan dengan fitur koefisien DCT hingga 84%, dan penelitian sebelumnya oleh Nanhndini dkk. [14] mengenai klasifikasi kanker kulit dengan metode Random Forest berhasil mencapai tingkat akurasi sebesar 97,3%. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi dan bantuan yang diberikan kepada tenaga kesehatan saat membaca citra USG, yang berfokus pada kasus tumor payudara yang jinak dan ganas, dengan mempertimbangkan karakteristik tekstur.

## METODE PENELITIAN

Pada penelitian ini database yang digunakan bersumber dari *Kaggle* dengan data yang diambil sebanyak 162 citra terdiri dari 81 kanker jinak dan 81 kanker ganas. Ukuran citra USG pada penelitian ini 256×256 pixel.

## Teknik Pengolahan Data



**Gambar 1.** Prosedur penelitian

Setelah tahapan pengumpulan data, akan dilakukan langkah-langkah pengolahan data seperti; flowchart pada Gambar 1 yang terdiri dari preprocessing, berupa konversi citra dari skala RGB (*Red, Green, Blue*) ke abu-abu. Kemudian Ekstraksi karakteristik tekstur menggunakan Histogram dan Gray Level Co-occurrence Matrix (GLCM).

### Ekstraksi Karakteristik Tekstur

Ekstraksi ciri adalah prosedur untuk mendapatkan suatu atribut yang dapat digunakan untuk menggambarkan gambar itu sendiri. Suatu objek dapat diidentifikasi berdasarkan kategorinya, yang merupakan suatu karakteristik. Ekstraksi tekstur adalah perulangan elemen pada citra tekstur dasar yang biasa disebut "texel". Tekstur dibuat secara alami sehingga bersifat acak, sedangkan tekstur pada citra buatan pola eksklusif atau berulang secara periodik biasanya memiliki penempatan probabilitas periodik, semi periodik, atau acak. Histogram dan GLCM digunakan untuk mengekstraksi ciri.

### Histogram

Histogram adalah representasi grafis dari distribusi frekuensi nilai-nilai piksel dalam citra. Ini memberikan gambaran tentang seberapa sering intensitas piksel tertentu muncul dalam citra. Histogram orde pertama tidak

memperhitungkan hubungan spasial antara piksel, hanya fokus pada frekuensi kemunculan nilai piksel. Histogram dapat dihitung dengan mengelompokkan nilai piksel ke dalam "bins" atau interval nilai, dan menghitung jumlah piksel dalam setiap bin. Dengan histogram, kita dapat melihat sebaran intensitas piksel secara keseluruhan[15].

Output dari histogram mengindikasikan skala abu-abu (*grayscale*). Beberapa metode statistik yang dimanfaatkan untuk mengekstrak ciri-ciri tekstur melibatkan: [16]

1) Rata-rata (*Mean*) merupakan nilai tengah dari intensitas citra.

$$\mu = \sum_{n=0}^N f_n P(f_n) \quad (1)$$

dengan  $\mu$  mewakili nilai rata-rata,  $f_n$  adalah tingkat kecerahan citra,  $p(f_n)$  menyatakan distribusi histogram dari tingkat kecerahan  $n$ , dan  $N$  adalah tingkat kecerahan maksimum.

2) Standar deviasi adalah indikator dari tingkat variasi kontras rata-rata.

$$\sigma = \sqrt{\mu_2} = \sqrt{\sigma^2} \quad (2)$$

dengan  $\sigma$  adalah standar deviasi, dan  $z$  adalah probabilitas kemunculan nilai keabuan.

3) Varian merupakan variasi elemen suatu citra.

$$\alpha_2 = \sum_{n=0}^N (f_n - \mu)^2 P(f_n) \quad (3)$$

dengan  $\alpha_2$  adalah skewness, dan  $n$  adalah nomor piksel dari 0 sampai  $N$ .

4) *Skewness* merupakan kemiringan relatif dari kurva histogram citra.

$$\alpha_3 = \frac{1}{\alpha_2} \sum_{n=0}^N (f_n - \mu)^3 P(f_n) \quad (4)$$

dengan  $\alpha_3$  adalah *skewness*.

5) Kurtosis merupakan keruncingan relatif dari kurva histogram citra.

$$\alpha_4 = \frac{1}{\alpha_2^2} \sum_{n=0}^N (f_n - \mu)^4 P(f_n) - 3 \quad (5)$$

dengan  $\alpha_4$  adalah kurtosis

6) Entropi adalah representasi dari tingkat ketidakberaturan sebuah citra.

$$H = - \sum_{n=0}^N P(f_n)^2 \log P(f_n) \quad (6)$$

dengan  $f_n$  merupakan intensitas keabuan,  $\mu$  adalah *mean*, dan  $P(f_n)$  merupakan nilai histogram.

### GLCM (Gray Level Co-Occurrence Matrix)

Haralick mengusulkan GLCM pada tahun 1973 dengan 28 atribut untuk mengrepresentasikan pola spasial. GLCM memanfaatkan perhitungan orde kedua, yang berarti bahwa dalam mengukur tekstur, keterkaitan antara dua piksel citra asli diperhitungkan. Sebagai perbandingan, metode pengukuran tekstur pada orde pertama melibatkan perhitungan statistik berdasarkan nilai piksel citra asli saja, seperti varians, dan tidak menunjukkan hubungan ketetanggaan piksel [17].

Beberapa atribut yang dimanfaatkan menurut GLCM melibatkan [16]:

1) Energi, mengukur jumlah elemen kuadrat dalam GLCM, memiliki nilai 1 untuk gambar konstan. Kemudian sudut kedua dan keseragaman merupakan istilah lain dari energi.

$$\sum_{i,j=0}^{N-1} P(i,j)^2 \quad (7)$$

Jumlah skala abu-abu dinyatakan N-1 yaitu 256, dan matriks GLCM yang telah ternormalisasi dinyatakan dengan  $P(i,j)$ .

2) *Correlation* digunakan untuk menilai sejauh mana ketergantungan linear dari tingkat keabuan yang sedang diukur. Setiap diukur, terdapat nilai korelasi yang mencerminkan hubungan positif antara Tingkat keabuan. Hasil korelasi mengukur sejauh mana piksel berkorelasi dengan piksel tetangganya dalam keseluruhan gambar.

$$\sum_{i=0}^{g-1} \sum_{j=0}^{g-1} \frac{\{i \times j\} \times P(i,j) - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \quad (8)$$

dengan  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$ ,  $\sigma_y$ , dengan berturut-turut merupakan *mean* dan standar deviasi dari matriks  $P_x$  dan  $P_y$ .

3) Homogenitas mengukur bagian tertentu didasarkan dengan pola pengelompokkan berdasarkan perbedaan tingkat abu-abunya (*grayscale*).

$$\sum_{i=0}^{g-1} \sum_{j=0}^{g-1} \frac{P(i,j)}{1 + |i - j|} \quad (9)$$

Matriks GLCM yang telah ternormalisasi dinyatakan dengan  $P(i,j)$ .

4) Ukuran kontras intensitas antara satu piksel dengan piksel lainnya dalam seluruh gambar disebut sebagai *Contrast*. Nilainya menjadi nol ketika gambar bersifat konstan.

$$\sum_{i=0}^{g-1} |i - j|^2 P(i,j) \quad (10)$$

### Random Forest

Random Forest adalah suatu teknik dalam pembelajaran mesin yang memanfaatkan sejumlah besar pohon keputusan sebagai dasar untuk proses klasifikasi yang disatukan. Algoritma ini merupakan hasil penggabungan dari metode Random Sub Spaces dan Bagging [18]. Ada tiga aspek utama yang terlibat dalam implementasi algoritma Random Forest:

1. Membangun pohon prediksi dengan menggunakan bootstrap sampling.
2. Setiap pohon keputusan melakukan prediksi dengan memanfaatkan prediktor yang dipilih secara acak.
3. Dalam tahap klasifikasi, hasil prediksi dari setiap pohon keputusan digabungkan menggunakan metode *majority vote*.

Pada penelitian ini, klasifikasi data menggunakan algoritma Random Forest dengan skenario uji dataset yang menggunakan pembagian persentase (*percentage split*). Pembagian ini melibatkan pengelompokan total data menjadi dua bagian, yaitu data latih dan data uji. Dalam penelitian ini, seluruh jumlah data digunakan (100% dari total data). Persentase split yang diuji mencakup 5%, 10%, 15%, 20%, dan 25%, tujuannya adalah untuk mendapatkan hasil yang objektif dari berbagai persentase split yang berbeda.

Waikato Environment for Knowledge Analysis (WEKA) merupakan perangkat lunak (*software*) pembelajaran mesin yang terkenal, yang dikembangkan di University of Waikato, Selandia Baru. Weka merupakan platform yang mencakup visualisasi, algoritma analisis data, serta pemodelan prediktif, dilengkapi dengan antarmuka grafis yang memudahkan pengguna untuk mengakses fungsionalitas tersebut dengan mudah. Dalam penelitian ini menggunakan WEKA karena mendukung untuk beberapa pengembangan data agar lebih spesifik, mendukung processing data, clustering, klasifikasi serta visualisasi data [19].

### Teknik Analisis Data

Olaniyi dkk. [20] mengusulkan konsep terkait indeks evaluasi kinerja dalam konteks metode klasifikasi, dengan memanfaatkan perhitungan matriks konfusi. Matriks konfusi dirinci sebagai berikut:

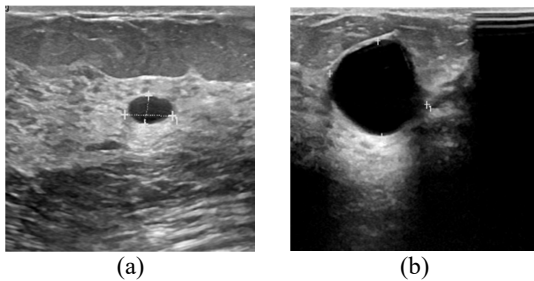
$$\text{Sensitivitas} = \frac{TP}{TP+FN} \times 100\% \quad (11)$$

$$\text{Spesifisitas} = \frac{TN}{FP+TN} \times 100\% \quad (12)$$

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (13)$$

True Negative (TN), False Positive (FP), False Negative (FN), True Positif (TP), Indeks pengukuran yang merupakan hasil dari pengelompokan data dengan menggunakan perangkat lunak WEKA untuk menghasilkan matriks kebingungan. TP menunjukkan data yang benar-benar positif dalam sistem klasifikasi, sedangkan TN adalah data yang benar-benar negatif dalam sistem klasifikasi [21].

## HASIL DAN PEMBAHASAN



**Gambar 2.** Citra USG Tumor (a) Jinak dan (b) Ganas dari Dataset Kaggle

Data penelitian yang digunakan adalah citra USG bersumber dari database *Kaggle*. Gambar 2 (a) dan 2(b) menunjukkan citra hasil USG pada tumor jinak dan ganas setelah *processing*. Perbedaan terlihat dalam konsistensi ukuran dan warna citra setelah diproses, diubah menjadi skala abu-abu (*grayscale*) untuk langkah selanjutnya.

**Tabel 1.** Nilai Rata-rata Karakteristik Tekstur Histogram dan GLCM

No	Atribut	Rata-rata Jinak	Rata-rata Ganas
1	Mean	88,509 ± 21,9736	80,2508 ± 20,1578
2	Standar Deviasi	51,5097±8,5647	52,6224 ± 10,0629
3	Varian	2726,1145±896,861	2869,9329 ± 1034,5162
4	Entropi	10,8694 ± 0,1419	10,8256 ±

			0,1487	
5	Skewness	0,5135 ± 0.3487	0,5731 ± 0,3623	
6	Kurtosis	-0,3151 ± 0,6738	-0,2535 ± 0,932	
7	Energi	0,0362 ± 0,0515	0,0424 ± 0,0667	
8	Contrast	76,5236 ± 38,5236	61,2999 ± 31,1563	
9	Correlation	0,9844 ± 0,0098	0,9882 ± 0,0064	
10	Homogenitas	0,2284 ± 0,0604	0,2403 ± 0,0648	

Ekstraksi ciri tekstur dilakukan dengan menggunakan 10 atribut, di mana 6 diantaranya didasarkan pada histogram dan 4 berasal dari GLCM. Hasil ekstraksi histogram mencakup nilai *mean*, variasi, standar deviasi, kurtosis, skewness, dan entropi. Sementara itu, atribut GLCM melibatkan korelasi, energi, homogenitas dan kontras. Tabel 1 menunjukkan nilai rata-rata dari ciri tekstur, dengan perbedaan yang signifikan terlihat pada atribut standar deviasi, varian, entropi, dan kurtosis. Sebaliknya, perbedaan yang lebih kecil terlihat pada mean, skewness, energi, kontras, korelasi, dan homogenitas. Perbedaan nilai rata-rata antara citra USG tumor jinak dan tumor ganas mencerminkan variasi dalam karakteristik tekstur. Rata-rata dari ciri tekstur ditambahkan dengan nilai dari standar deviasi untuk mengevaluasi sebaran data. Standar deviasi bermanfaat untuk menilai tingkat kontras citra, di mana semakin rendah kontras citra akan diikuti oleh standar deviasi yang lebih rendah, dan sebaliknya. Untuk setiap nilai atribut yang diperoleh, atribut dengan sebaran data seragam termasuk skewness, kontras, korelasi, energi, dan homogenitas. Sementara itu, atribut dengan sebaran data acak termasuk mean, varian, entropi, standar deviasi, dan kurtosis.

**Tabel 2.** Hasil analisis data

No	Persentase Split	TP	FP	FN	TN	Accuracy %	Precision %	Recall %
1	Training	81	0	0	81	100%	100%	100%
2	5	46	40	35	41	54%	53%	57%
3	10	44	39	37	42	53%	53%	54%
4	15	41	39	40	42	51%	51%	51%
5	20	45	41	36	40	52%	52%	56%
6	25	45	43	36	38	51%	51%	56%

Berdasarkan hasil yang ditunjukkan pada Tabel 2 memberikan hasil yang bervariasi tergantung

pada persentase pembagian data latih dan uji. Pada tahap *training* saat menunjukkan nilai 100% artinya data yang diolah berhasil dan dapat mengelompok dengan baik, Meskipun pada tahap pelatihan (*training*) didapatkan hasil yang sangat baik dengan akurasi, presisi, dan recall mencapai 100%, namun pada uji data dengan persentase split yang berbeda, terjadi fluktuasi hasil klasifikasi.

Pada variasi percentage split 5% dan 10%, meskipun akurasi masih berada pada kisaran 53-54%, presisi dan recall menunjukkan angka yang kurang konsisten, dengan nilai tertinggi pada recall sebesar 57%. Pada percentage split 15%, hasil klasifikasi menunjukkan penurunan dengan akurasi, presisi, dan recall sekitar 51%. Sementara pada percentage split 20% dan 25%, terlihat sedikit peningkatan kembali dengan recall mencapai 56%, namun masih mempertahankan tingkat akurasi yang relatif rendah.

Random Forest dapat memberikan hasil yang sangat baik pada tahap pelatihan, namun pada uji data nyata dengan variasi pembagian persentase, performanya menjadi lebih tidak stabil. Oleh karena itu, perlu dilakukan perbaikan selanjutnya terkait faktor-faktor yang memengaruhi fluktuasi hasil klasifikasi, serta eksplorasi metode atau parameter yang dapat meningkatkan konsistensi dan akurasi hasil klasifikasi pada kasus kanker payudara. Meskipun demikian, penelitian ini tetap memberikan kontribusi positif dengan menggambarkan tantangan dan potensi pengembangan lebih lanjut dalam penggunaan teknologi Random Forest untuk mendukung diagnosis kanker payudara melalui citra ultrasonografi [12].

Pada penelitian ini metode klasifikasi menggunakan Random Forest menghasilkan nilai *accuracy* tertinggi sebesar 54%, nilai *precision* tertinggi sebesar 53%, dan nilai *recall* tertinggi sebesar 57%. Hasil ini jika dibandingkan dengan penelitian sebelumnya yang dilakukan oleh Julia dkk. [22] mengenai klasifikasi tumor payudara dengan metode Naive Bayes dengan nilai akurasi sebesar 80%. menunjukkan bahwa akurasi yang dihasilkan lebih rendah. Hal ini disebabkan karena pada penelitian sebelumnya, tahap preprocessing gambar tidak melalui tahap cropping, filtering, dan penggunaan data yang terbatas. Selain itu, citra yang digunakan berbeda, pada penelitian ini menggunakan citra USG sedangkan pada penelitian sebelumnya menggunakan cita

mamografi. Kelebihan pada penelitian ini bisa digunakan untuk tools tambahan diagnosa tumor payudara.

## KESIMPULAN

Telah dilakukan klasifikasi jenis tumor payudara jinak dan ganas berbasis karakteristik tekstur dengan 81 citra tumor jinak dan 81 citra tumor ganas yang merupakan citra USG. Penelitian ini menggunakan metode Random Forest dengan 2 sesi yaitu *training* dan *testing*. Pada tahap *training* dihasilkan *accuracy* sebesar 100%, *precision* 100%, dan *recall* 100%. Selanjutnya pada tahap *testing* dilakukan 5 variasi persentase split dengan nilai tertinggi pada persentase split 5, dihasilkan *accuracy* sebesar 54%, *precision* 53%, dan *recall* 57%. Penelitian ini perlu dikembangkan kembali untuk memberikan kontribusi dalam bidang medis terutama dalam diagnosa tumor payudara dan menjadi instrumen tambahan dalam bidang medis.

## DAFTAR PUSTAKA

- [1] Globocan, "Cancer Today," Int. Agency Res. Cancer. <http://gco.iarc.fr/today/>, 2018.
- [2] B. Asliminarti, "Literatur Review Faktor-Faktor Yang Mempengaruhi Tingkat Pengetahuan," 2020.
- [3] Akhmadi, R. Nia Pratiwi, D. Rusita, and Nurhasuna, "Peran Radiografer Dalam Pemeriksaan Ultrasonografi (Usg)," *JRI (Jurnal Radiogr. Indones.)*, vol. 4, no. 2, pp. 79–81, 2021, doi: 10.55451/jri.v4i2.92.
- [4] M. Papadakis, S. J. McPhee, M. W. Rabow, and K. R. McQuaid, *Current Medical Diagnosis & Treatment*. 2007.
- [5] M. Khan, *Pass Ultrasound Physics Exam*. Blue Cube Venture, LLC, 2014.
- [6] E. Luthfiana AS, A. Mayub, and N. Nirwana, "The Ultrasonography Examination of Fetal Heart Beats as the Application of Mechanic Waves Principle," *Bencoolen J. Sci. Educ. Technol.*, vol. 1, no. 2, pp. 29–37, 2020, doi: 10.33369/bjset.1.2.29-37.
- [7] J. Bates, *Abdominal Ultrasound*. 2004. doi: 10.1002/9781119860792.ch37.

- [8] A. Nugroho, "Klaasifikasi Nodul Tiroid Berbasis Ciri Tekstur pada Citra Ultrasonografi," *Tesis. Yogyakarta Progr. Pascasarj. Univ. Gadjah Mada Yogyakarta*, 2015.
- [9] H. K. N. Yusufiyah, "Klasifikasi Nodul Payudara Berbasis Ciri Bentuk pada Citra Ultrasonografi," *Tesis. Yogyakarta Progr. Pascasarj. Univ. Gadjah Mada Yogyakarta*, 2015.
- [10] D. A. Khusna, "Klasifikasi Lesi Citra Ultrasonografi Payudara Berdasarkan Karakteristik Tepi," *Tesis. Yogyakarta Progr. Pascasarj. Univ. Gadjah Mada Yogyakarta*, 2016.
- [11] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [12] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*, vol. 9781107057. 2013. doi: 10.1017/CBO9781107298019.
- [13] H. Arwoko, "Klasifikasi Kanker Payudara pada Citra Ultrasound Menggunakan Fitur Koefisien Discrete Cosine Transform (DCT)," *Pros. HUBISINTEK*, vol. 2, no. 1, p. 451, 2022.
- [14] M. S. Nandhini, M. A. Sofiyan, S. Kumar, and A. Afridi, "Skin Cancer Classification using Random Forest," *Int. J. Manag. Humanit.*, vol. 4, no. 3, pp. 39–42, 2019, doi: 10.35940/ijmh.c0434.114319.
- [15] K. S. and I. D. R. M. Haralick, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. SMC-3, no., pp. 610–621, 1973.
- [16] K. Harismah, D. Vitasari, A. Fuadi, and T. Setyawan, "Diversifikasi Minyak Sereh dan Minyak Cengkeh Sebagai Bahan Repellent Pembuatan Tinta Anti Nyamuk," *4th Univesity Res. Coloquium*, pp. 526–530, 2016, [Online]. Available: <https://publikasiilmiah.ums.ac.id/xmlui/handle/11617/10986>
- [17] A. Kadir, A., & Susanto, "Teori dan Aplikasi Pengolahan Citra.," (*Dewibertha Hardjono, Ed.*). Yogyakarta: Andi.No Title, 2013.
- [18] & Sari, C. A., Sari, W. S., Nilawati, F. E., Doheir, M. and C. A. Sari, "Analysis of Childbirth Methods Using Random Forest Algorithms.," 2022, doi: <https://doi.org/10.12487/AMRI.v1i1.xxxx.x>.
- [19] R. Arora and S. Suman, "Comparative Analysis of Classification Algorithms on Different Datasets using WEKA," *Int. J. Comput. Appl.*, vol. 54, no. 13, pp. 21–25, 2012, doi: 10.5120/8626-2492.
- [20] E. O. Olaniyi, A. A. Adekunle, T. Odekuoye, and A. Khashman, "Automatic system for grading banana using GLCM texture feature extraction and neural network arbitrations," *J. Food Process Eng.*, vol. 40, no. 6, 2017, doi: 10.1111/jfpe.12575.
- [21] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA workbench," *Data Min.*, pp. 553–571, 2017, doi 10.1016/b978-0-12-804291-5.00024-6, 2017.
- [22] B. Julia, H. Sumarti, and H. H. Kusuma, "Klasifikasi Tumor Payudara Berbasis Ciri Tekstur pada Citra Mammografi Menggunakan Metode Naive Bayes," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 7, no. 2, pp. 165–176, 2022, [Online]. Available: <https://doi.org/10.26760/mindjournal.v7i2.165-176>