

## Klasifikasi Kanker Payudara pada Citra Ultrasound Menggunakan Fitur Koefisien Discrete Cosine Transform (DCT)

Heru Arwoko

Jurusan Informatika, Fakultas Teknik, Universitas Surabaya  
Jl. Raya Kalirungkut No.1 Tenggilis Mejoyo, Surabaya, Indonesia  
Telp. (031) 2981000 Kode Pos : 60292  
Email : heru\_a@staff.ubaya.ac.id

### Abstrak

*Gejala dini kanker payudara dapat diketahui dari hasil analisa citra medis USG. Pada kenyataannya, citra medis USG memiliki kualitas rendah, banyak noise, dan sangat heterogen. Hal ini menimbulkan kesulitan dalam tahap penentuan klasifikasi kanker secara mesin. Berbagai metode telah dilakukan para peneliti untuk mengklasifikasikan citra medis USG. Metode yang umum dipakai berbasis fitur geometri, yaitu penentuan radius, luas, keliling, dan tekstur dari daerah citra yang diperhatikan sebagai lokasi kanker. Pada penelitian ini digunakan ekstraksi fitur koefisien DCT dari citra secara langsung, tanpa melalui proses analisis geometri. Transformasi DCT dari sebuah citra merupakan penyajian citra pada domain frekuensi, dimana fitur dari citra terpusat pada daerah frekuensi rendah. Pada penelitian ini dilakukan pemotongan koefisien DCT pada daerah frekuensi rendah untuk mendapatkan fitur utama. Hasil akurasi yang didapatkan mencapai 84% pada pengambilan 225 koefisien DCT dari dataset citra ultrasound BUSI yang terdiri dari 210 kanker jinak dan 210 kanker ganas. Proses klasifikasi efektif karena fitur diambil secara global tanpa analisis geometri manual. Hasil penelitian ini diharapkan bermanfaat pada bidang medis dan kecerdasan buatan untuk identifikasi citra medis USG.*

**Kata Kunci:** Citra Medis, USG, Kanker Payudara, Klasifikasi, DCT

### Abstract

*Early symptoms of breast cancer can be identified from the analysis of medical ultrasound images. In fact, ultrasound images are low quality, a lot of noisy, and very heterogeneous. This causes difficulties in the classification stage by machine. Various methods have been proposed by many researchers to classify ultrasound medical images. Generally based on geometric characteristics, such as determining the radius, area, circumference, and texture of the image that is suspected of being the location of the cancer. In this paper, feature extraction is taken from the DCT coefficient on the image directly, without going through the process of geometric analysis. The DCT transformation of an image is an image representation in the frequency domain, where the image characteristics are concentrated in the low frequency region. In this study, the DCT coefficient was cut in the low frequency region to obtain the main feature extraction. Accuracy results obtained reached 84% on the retrieval of 225 DCT coefficients from the BUSI ultrasound dataset consisting of 210 benign cancers and 210 malignant cancers. The process is effective because the features are fetched globally without manual geometry analysis. The results of this study are expected to be useful in the medical field and artificial intelligence to identify ultrasound medical images.*

**Keywords:** Medical Image, USG, Breast Cancer, Classification, DCT

## 1. Pendahuluan

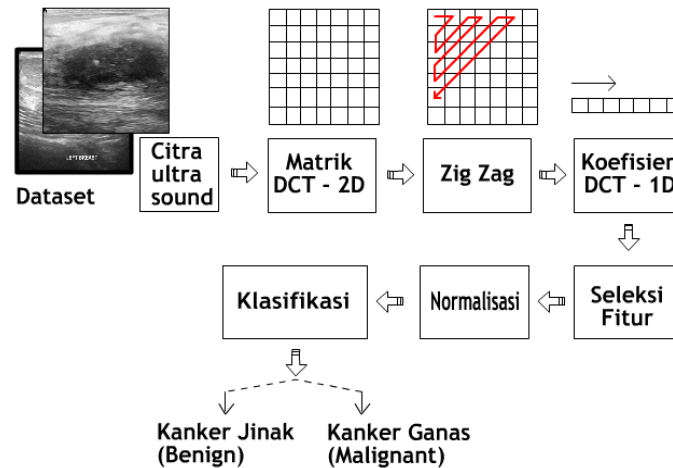
*Ultrasonografi* (USG) merupakan peralatan medis diagnostik yang sering digunakan dalam bidang medis. Keunggulan USG adalah murah, waktu akuisisi pendek, portable, aman, akurat, dan tidak menimbulkan radiasi berbahaya. Pada awal sejarah perkembangan studi penyakit kanker, peralatan medis USG berperan penting dalam pengukuran, identifikasi, dan penentuan stadium perkembangan kanker. Ketepatan hasil identifikasi secara manual pada analisis citra USG sangat bergantung pada pengamatan dokter atau operator yang menggunakannya. Tidak menutup kemungkinan terjadi kesalahan identifikasi secara manual. Analisa citra USG dalam sistem cerdas komputer merupakan langkah penting dalam diagnosis, klasifikasi, dan deteksi dini kanker. Menurut Walid Al-Dhabyani, et al. (2019), deteksi dini membantu dapat mengurangi jumlah kematian dini. Menurut Srwa Hasan Abdullah, et al. (2021), kanker payudara tetap menjadi salah satu penyakit teratas yang menyebabkan ribuan kematian pada wanita setiap tahun. Pada kenyataannya, proses identifikasi citra USG memiliki keterbatasan yaitu rendahnya kualitas citra dan banyaknya noise sehingga menimbulkan kesulitan dalam tahap analisa klasifikasinya. Pada penelitian Rajendra, et al. (2016) telah melakukan identifikasi citra USG penyakit liver menggunakan fitur koefisien DCT dari citra USG. Kemudian fitur koefisien DCT digunakan sebagai input menuju Locality Sensitive Discriminant Analysis (LSDA) untuk mengurangi banyaknya koefisien fitur.

Pada penelitian ini dilakukan ekstraksi fitur DCT secara langsung dengan pengambilan koefisien DCT pada daerah frekuensi rendah yang memiliki energi citra yang besar. Proses klasifikasi efektif karena fitur diambil secara global tanpa analisis geometri manual. Pada penelitian ini digunakan mesin klasifikasi, *Decision Tree*, *Random Forest*, *Support Vector Machine*, *Logistic Regression*. Menurut Breiman, et al (2005) *Decision Tree* (DT) merupakan mesin klasifikasi menggunakan *Tree* dari data pelatihan. Aturan pelatihan diperoleh dari *Tree* yang dibangun untuk mengklasifikasikan dua kelas. Semakin baik desain *Tree* yang dibangun semakin baik kinerjanya. Menurut Han, et.al (2005). *Random Forest* (RF) merupakan mesin klasifikasi ensemble yang memiliki kinerja sangat baik dan proses pelatihan yang cepat dan efisien. RF diawali pada sebuah nilai vektor random dengan distribusi yang sama pada semua *Tree* yang masing masing memiliki kedalaman *decision tree* yang sama. Kemudian dari kombinasi *Tree* yang terbentuk dipilih satu model yang mendapatkan akurasi maksimal. Menurut Tomar, et al. (2015), *Support Vector Machine* (SVM) merupakan klasifikasi yang bekerja dengan membangun sebuah hyperplane (N-1) dimensi. SVM mengklasifikasikan data ke dalam dua kelas dimana N mewakili jumlah fitur input. Hyperplane bertindak sebagai permukaan keputusan yang memisahkan dua kelas dengan margin maksimum. Vektor di dekat hyperplane di kedua sisi disebut *support vector*. Dataset pelatihan terdiri dari variabel target yang disebut label kelas dan variabel yang tersisa disebut variabel fitur. Model SVM yang dihasilkan memetakan fitur data uji pada ruang yang sama dengan fitur data latih. Menurut Breiman, et. Al (1984), *Logistic Regression* (LR) merupakan mesin klasifikasi untuk mendapatkan hubungan antara fitur dengan hasil label yang dihasilkan berdasarkan bobot probabilitas. LR membutuhkan suatu fungsi yang menghubungkan nilai output Y pada fungsi linier dengan nilai Y pada fungsi sigmoid.

## 2. Metodologi

Pada langkah awal citra USG dilakukan proses transformasi DCT untuk menghasilkan matrik DCT-2D. Nilai koefisien matrik DCT-2D pada sudut kiri atas merupakan daerah frekuensi rendah yang membawa sebagian besar fitur utama citra. Daerah frekuensi rendah memuat rata-rata seluruh matriks sedangkan koefisien frekuensi tinggi yang tersisa menggambarkan variasi intensitas di antara blok citra. Langkah berikutnya adalah melakukan pengambilan data koefisien secara zigzag pada matrik DCT-2D untuk

mendapatkan koefisien DCT-1D yang dimulai dari koefisien pada sudut kiri atas, seperti terlihat pada Gambar 1. Langkah ini merupakan proses pembentukan fitur citra yang hasilnya berupa koefisien DCT yang merupakan array 1D. Pengambilan data zigsag dilakukan karena fitur utama dari matrik DCT-2D berada pada daerah frekuensi rendah yaitu dimulai dari sudut kiri atas matrik. Fitur ekstraksi diambil dari koefisien DCT-1D dengan mengambil sejumlah koefisien terbatas untuk mengurangi kelebihan data (redundansi). Pada penelitian ini dilakukan eksperimen variasi pengambilan jumlah koefisien DCT agar didapatkan hasil akurasi yang maksimum.



Gambar 1. Diagram Alur Proses Klasifikasi

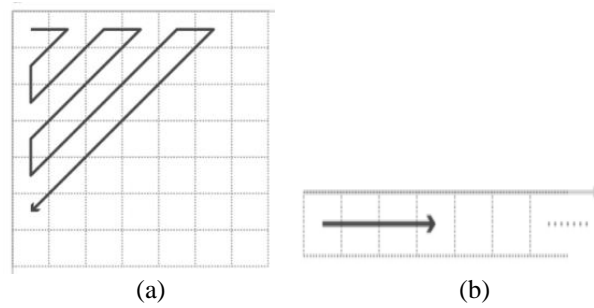
Langkah berikutnya adalah melakukan normalisasi data fitur, yaitu mengatur skala data-data fitur untuk diproses ke dalam mesin klasifikasi. Hasil klasifikasi dibagi menjadi dua label kelas, yaitu kelas kanker jinak dan kelas kanker ganas.

## 2.1 Dataset

Sebelumnya para peneliti mengalami kesulitan melakukan penelitian ini karena belum *tersedia* dataset publik dari citra USG kanker payudara. Pada tahun 2020 Fahmi, et.al melakukan publikasi dataset BUSI (Breast Ultrasound Images) dalam jurnal internasional. Dataset BUSI dikumpulkan dari *Baheya Hospital for Early Detection and Treatment of Women's Cancer*, Cairo (Mesir). Dataset BUSI diperoleh dari pasien wanita berusia antara 25 dan 75 tahun pada tahun 2018. Ukuran citra *grayscale* USG BUSI bervariasi rata-rata 500×500 piksel. Pada penelitian ini, dataset mengacu pada dataset (BUSI) diambil sebanyak 420 citra yang terdiri 210 kanker jinak dan 210 kanker ganas. Citra USG yang diteliti dipisahkan menjadi dua kelas, yaitu jinak dan ganas. Dataset dipisahkan pada rasio 10%, artinya sebanyak 42 citra data uji dan sisanya 396 citra data latih.

## 2.2 Transformasi Discrete Cosine Transform 2D

Pada *penelitian* ini akan dibentuk fitur koefisien DCT yang signifikan membentuk vektor fitur citra dengan cara pengambilan secara zigzag dari matrik DCT-2D. Pada Gambar 2(a) ditunjukkan cara mendapatkan fitur koefisien DCT secara zigzag. Hasil fitur koefisien DCT ditunjukkan pada Gambar 2(b) berupa array 1D. Pada matriks DCT-2D, koefisien sudut kiri atas membawa sebagian besar energi citra dan menggambarkan rata-rata seluruh matriks. Koefisien yang tersisa menggambarkan variasi intensitas di antara blok citra. Proses pengambilan secara zigzag dimulai dari posisi data di kiri atas matrik DCT-2D. Hasilnya berupa fitur koefisien DCT yang dipakai sebagai input menuju mesin klasifikasi untuk analisa klasifikasi jenis kanker.



Gambar 2. Pengambilan Koefisien DCT secara Zigzag

Transformasi DCT adalah metode yang umum digunakan untuk kompresi citra, yaitu mengubah citra dari domain spasial (posisi) menjadi domain frekuensi. Secara matematik, koefisien DCT dari citra berukuran  $M \times N$  yang memiliki fungsi spasial  $f(x,y)$  dirumuskan dalam persamaan berikut:

$$F(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos\left(\frac{(2x+1)u\pi}{2M}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right)$$

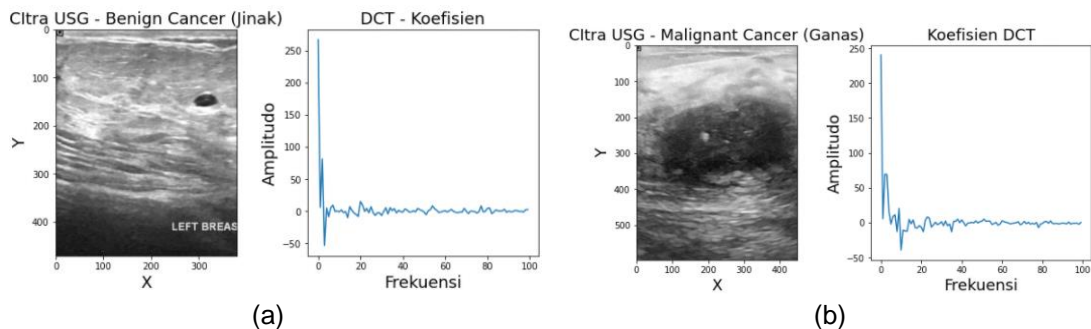
$$u = 0, 1, 2, \dots, M-1, \quad v = 0, 1, 2, \dots, N-1 \quad (1)$$

$$\text{Dimana } \alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & 1 \leq u \leq M-1 \end{cases} \quad \alpha(v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & 1 \leq v \leq N-1 \end{cases} \quad (2)$$

$F(u,v)$  adalah fungsi hasil transformasi DCT berupa matrik DCT-2D berupa domain frekuensi dari citra. Dalam hal ini  $\alpha(u=0)$  adalah nilai rata-rata citra dan koefisien DCT bersifat *ortogonal* satu sama lain. Fitur DCT pada citra USG merupakan fitur proyeksi seluruh citra yang mencerminkan perubahan nilai intensitas citra.

### 2.3 Transformasi Discrete Cosine Transform 2D

Koefisien DCT-1D diperoleh dari hasil matrik transformasi DCT-2D berukuran  $500 \times 500$  piksel. Pada penelitian ini, telah diekstrak sebanyak 250000 koefisien 1D-DCT dari matriks DCT-2D melalui pengambilan secara zigzag. Kemudian dilakukan ekstraksi fitur pada daerah frekuensi rendah, karena memiliki energi yang paling tinggi, dengan melakukan pemotongan koefisien fitur yang tidak penting untuk mengurangi kelebihan data.



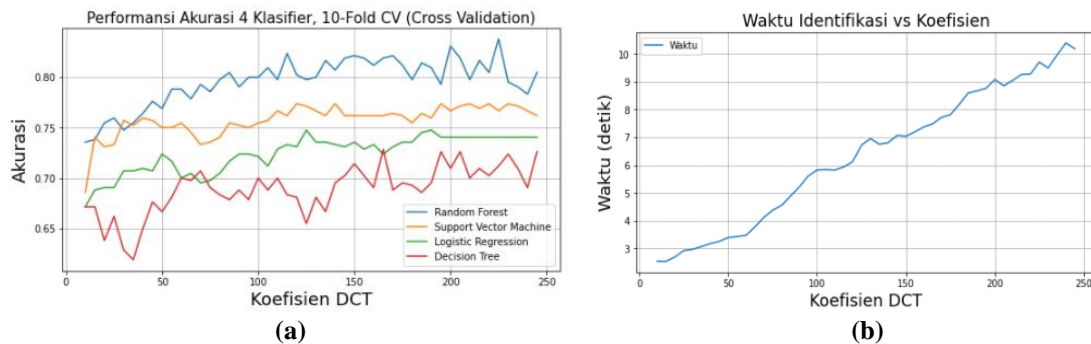
Gambar 3. Ekstraksi Fitur DCT Citra USG Kanker Payudara Jinak dan Ganas

Pada Gambar 3(a) ditunjukkan grafik hasil ekstraksi fitur pada citra USG kanker jinak. Pada frekuensi rendah memiliki amplitudo atau energi citra yang tinggi. Amplitudo berkurang dengan bertambahnya frekuensi. Pada Gambar 3(b) ditunjukkan grafik hasil ekstraksi fitur pada citra USG kanker ganas. Terlihat pada citra USG kanker ganas bentuk lonjakan amplitudo lebih fluktuatif perubahannya. Hal ini berarti bahwa citra USG kanker ganas memiliki perubahan energi yang

lebih kompleks. Pada penelitian ini, fitur ekstraksi DCT-1D dijadikan sebagai data fitur untuk diimplementasikan pada mesin klasifikasi yang dipilih yaitu *Random Forest*, *Support Vector Machine*, *Logistic Regression*, dan *Decision Tree* pada kinerja yang sama yaitu 10 Fold-Cross Validation (CV).

### 3. Hasil dan Pembahasan

Setiap citra USG berukuran 500×500 piksel akan menghasilkan 250000 total fitur koefisien DCT. Kemudian dipilih fitur utama yang berada pada daerah frekuensi rendah. Fitur direduksi hanya menjadi 300 koefisien yang dimodelkan sebagai ciri utama dari sebuah citra USG. Pada *penelitian* ini dilakukan variasi jumlah koefisien dan dianalisa hasil akurasi yang dicapai dan dianalisa juga waktu pemrosesan klasifikasi. Pada Gambar 4(a) menunjukkan grafik perbandingan performansi akurasi dari berbagai model mesin klasifikasi berkaitan dengan jumlah fitur koefisien DCT.



Gambar 4. Hasil Pengujian Akurasi

Terlihat bahwa mesin klasifikasi RF memiliki performansi akurasi yang tertinggi. Akurasi tertinggi sebesar 0.84 tercatat pada pengambilan 225 koefisien. Perubahan pengambilan jumlah koefisien *menimbulkan* fluktuasi nilai akurasi. Penambahan jumlah koefisien tidak signifikan terhadap kenaikan akurasi, namun akurasi maksimum terjadi pada pengambilan jumlah koefisien yang optimal. Penambahan jumlah koefisien signifikan terhadap proses waktu klasifikasi. Makin banyak jumlah koefisien makin lama waktu identifikasi yang dibutuhkan selama proses klasifikasi, seperti terlihat pada Gambar 4(b)

Tabel 1. Hasil Akurasi Pada Berbagai Koefisien DCT

Koefisien DCT	Metode Klasifikasi			
	LR	RF	DT	SVM
200	0.74	0.83	0.71	0.77
205	0.74	0.82	0.73	0.77
210	0.74	0.80	0.70	0.77
215	0.74	0.82	0.71	0.77
220	0.74	0.80	0.70	0.77
225	0.74	0.84	0.71	0.77
230	0.74	0.80	0.72	0.77
235	0.74	0.79	0.71	0.77

Pada Tabel 1 terlihat bahwa pengambilan jumlah koefisien sebanyak 225 koefisien tercatat memiliki akurasi yang maksimal sebesar 0.84. Terlihat bahwa mesin klasifikasi RF memiliki performansi yang lebih unggul dibandingkan yang lain.

### 4. Kesimpulan dan Saran

Pada *klasifikasi* secara otomatis kanker payudara jinak atau ganas dari citra USG dapat dilakukan menggunakan fitur koefisien DCT. Fitur koefisien DCT terbukti signifikan sebagai ciri dari citra. Hal ini dibuktikan dari hasil akurasi sebesar 0.84 atau 84% pada

pengambilan 225 koefisien DCT. Pemotongan koefisien DCT pada daerah frekuensi rendah terbukti sebagai fitur utama. Proses klasifikasi efektif karena fitur diambil secara global tanpa analisis geometri manual. Hasil penelitian ini diharapkan bermanfaat pada bidang medis dan kecerdasan buatan untuk identifikasi citra medis USG. Saran pengembangan lebih lanjut, penelitian ini perlu dikombinasikan dengan fitur wavelet dari citra kemudian diteliti hasil performansi akurasi.

### Daftar Pustaka

- Rajendra Acharya, Hamido Fujita, Vidya Sudarshan, 2016, An integrated index for identification of fatty liver disease using radon transform and discrete cosine transform features in ultrasound images, *Journal Elsevier*, Vol. 31 pp. 43–53.
- Srwa Hasan Abdullah, Ali Makki Sagheer, Hadi Veisi, 2021, Breast Cancer Classification Using Machine Learning Techniques: A Review, *Turkish Journal of Computer and Mathematics Education* Vol. 12
- Walid Al-Dhabyani, Aly Fahmy, 2019, Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images, (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 5, 2019
- Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, *CRC press*, 1984.
- Yang, S. Yang, S. Li, R. Zhang, F. Liu, L. Jiao, Coupled compressed sensing inspired sparse spatial-spectral LSSVM for hyperspectral image classification, *Knowl. Based Syst.* 79 (2015) 80–89.
- Tomar, S. Agarwal, A comparison on multi-class classification methods based on least squares twin support vector machine, *Knowl. Based Syst.* 81 (2015) 131–147.
- Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Morgan Kaufmann, Waltham, MA, 2005.
- Jeong, R. Jayaraman, Support vector-based algorithms with weighted dynamic time warping kernel function for time series classification, *Knowl. Based Syst.* 75 (2015) 184–191.