

KLASIFIKASI KANKER PARU PARU MENGUNAKAN NAÏVE BAYES DENGAN VARIASI FILTER DAN EKSTRAKSI CIRI *GRAY LEVEL CO- OCCURRENCE MATRIX (GLCM)*

Mohtar Yuniarto^{1*}, Soeparmi¹, Cari¹, Fuad Anwar¹, Delta Nur Septianingsih¹,
Tonang Dwi Ardyanto², Resta Farits Pradana²

¹Program Studi Fisika, Fakultas MIPA, Universitas Sebelas Maret, Surakarta

²Rumah Sakit Universitas Sebelas Maret, Surakarta

*mohtar_yuniarto@staff.uns.ac.id

Received 14-07-2021, Revised 07-09-2021, Accepted 29-09-2021, Published 01-10-2021

ABSTRACT

Lung cancer classification has been successfully carried out from 120 CT Scan image data. In the research, the preposition process begins with filtering variations using a low pass filter, median filter, and high pass filter. The segmentation used is Otsu Thresholding which then the texture will be extracted using the Gray Level Co-occurrence Matrix (GLCM) feature with variations in the angle direction. The results of the GLCM extraction are used as a database that will become a dataset for image classification using naive Bayes classification. The results of the study with 12 variations, the best variation results were the median filter with the GLCM angle of 0^0 indicating the highest level of accuracy of 88.33%.

Keyword : Lung cancer; otsu thresholding; gray level co-occurrence matrix (GLCM); naïve bayes.

ABSTRAK

Telah berhasil dilakukan klasifikasi kanker paru-paru dari 120 data citra CT Scan. Pada penelitian, proses preposisi dimulai dengan variasi filtering menggunakan *low pass filter*, *median filter*, dan *high pass filter*. Segmentasi yang digunakan yaitu *Otsu Thresholding* yang kemudian teksturnya akan diekstraksi menggunakan fitur *Gray Level Co-occurrence Matrix (GLCM)* dengan variasi arah sudut. Hasil dari ekstraksi GLCM dijadikan database yang akan menjadi dataset untuk pengklasifikasian citra menggunakan klasifikasi naïve bayes. Hasil dari penelitian dengan 12 buah variasi diperoleh hasil variasi terbaik adalah median filter dengan arah sudut GLCM 0^0 menunjukkan tingkat akurasi yang paling tinggi sebesar 88,33 %.

Kata kunci: kanker paru; *otsu thresholding*; *gray level co-occurrence matrix (GLCM)*; naïve bayes.

PENDAHULUAN

Pengolahan citra digital dapat didefinisikan sebagai akuisi dan pemrosesan informasi visual oleh komputer^[1]. Pengolahan citra digital memiliki spektrum aplikasi yang luas, salah satunya yaitu dalam bidang biomedis (*Biomedical*). Dalam bidang kedokteran pada

tahun 1970-an ketika ditemukan tomografi terkomputerisasi (*Computerized Tomography/CT*), pengolahan citra digital mengalami kemajuan yang sangat penting dalam bidang medis^[2]. Seiring dengan perkembangan pengolahan citra digital yang semakin maju, dalam bidang medis dapat digunakan untuk menganalisis seperti gambar rontgen^[3] *Computed Tomography* (CT), *Positron Emission Tomography* (PET), *X-Ray Tomography*, *Ultrasound Transmission Tomography*, *Single Photon Emission Computed Tomography* (SPECT), dan sebagainya^[4]. Oleh karena meningkatnya teknologi dalam bidang medis, pencitraan medis semakin banyak digunakan dalam diagnosis klinis dan pengobatan penyakit^[5].

Kanker merupakan penyakit yang terjadi karena perubahan sel yang menyebabkan pertumbuhan dan pembelahan sel tak terkendali. Sel kanker dapat menyebabkan tumor, merusak sistem kekebalan tubuh dan menyebabkan kelainan lain yang mencegah tubuh berfungsi dengan benar^[6]. Menurut *World Health Organization* (WHO), kanker paru-paru merupakan salah satu penyebab kematian yang paling banyak di antara kematian akibat kanker lainnya, baik pada pria ataupun wanita disemua umur. Menurut data Globocan (2018), presentase kasus kanker paru-paru yang terjadi di dunia sebanyak 11,6% dengan angka kematian mencapai 18,4%, sedangkan kasus kanker paru-paru di Indonesia mencapai 8,6% atau 30.023 kasus dengan angka kematian sebesar 12,6% atau 26.095 kasus kematian akibat kanker paru-paru.

Salah satu cara yang digunakan untuk mendeteksi kanker paru-paru ialah melalui pemeriksaan sinar-X (*Rontgen*). Penggunaan sinar-X dalam pemeriksaan paru-paru merupakan teknik yang paling sering digunakan. Hasil proses dari sinar-X memberikan citra yang berbeda antara paru-paru yang sehat dan tidak sehat, seperti paru-paru normal ataupun kanker paru^[7]. Namun, biasanya citra medis memiliki *noise* yang dapat menurunkan kualitas dari citra yang dihasilkan. *Noise* seharusnya dihilangkan atau direduksi agar informasi dalam citra dapat diperoleh dengan jelas, akan tetapi hingga saat ini *noise* selalu ada dalam setiap pengambilan citra^[8]. Oleh karena itu, pada tahap pra-pemrosesan dimulai dengan perbaikan citra yang bertujuan untuk memperbarui dari informasi yang terkandung dalam citra serta menemukan masukan yang tepat untuk berbagai jenis metode pemrosesan citra terkomputerisasi. *Filtering* mempunyai peran penting dalam pemrosesan citra dan visi komputer. Dalam hal menghilangkan *noise*, *low pass filter* bekerja sebagai *filter* linier dimana filter gaussian bekerja dengan baik di daerah yang halus. Akan tetapi, secara signifikan membuat blur deteksi tepi pada struktur citra^[9]. Untuk meningkatkan resolusi frekuensi, didekomposisi ulang dengan menguraikan keluaran dari *low pass filtering*^[10].

Teknologi segmentasi citra merupakan salah satu masalah mendasar di bidang pengolahan citra dan komputer. Segmentasi citra adalah teknologi yang membagi citra asli menjadi beberapa area dan mengekstrak area efektif sesuai dengan karakteristik citra yang berbeda, seperti tekstur, tingkat abu-abu dan tingkat kecerahannya^[5]. Segmentasi citra memiliki peran yang sangat penting dalam analisis citra. Masalah penting dalam proses segmentasi adalah menemukan tepinya. Citra medis yang tersegmentasi dengan baik akan didapatkan informasi mengenai batasan-batasan area dengan jelas yang akan sangat membantu tenaga medis secara objektif dan akurat dalam menganalisa, mendiagnosa penyakit, melakukan perencanaan pengobatan dan tindakan selanjutnya^[11]. Dalam penelitian yang dilakukan oleh Rulaningtyas dan Ain (2020) mengenai segmentasi citra CT Scan berdasarkan nilai Houndsfield Unit (HU) menggunakan metode otsu thresholding untuk mendapatkan hasil segmentasi jantung. Dikembangkan metode Otsu Thresholding berdasarkan nilai threshold

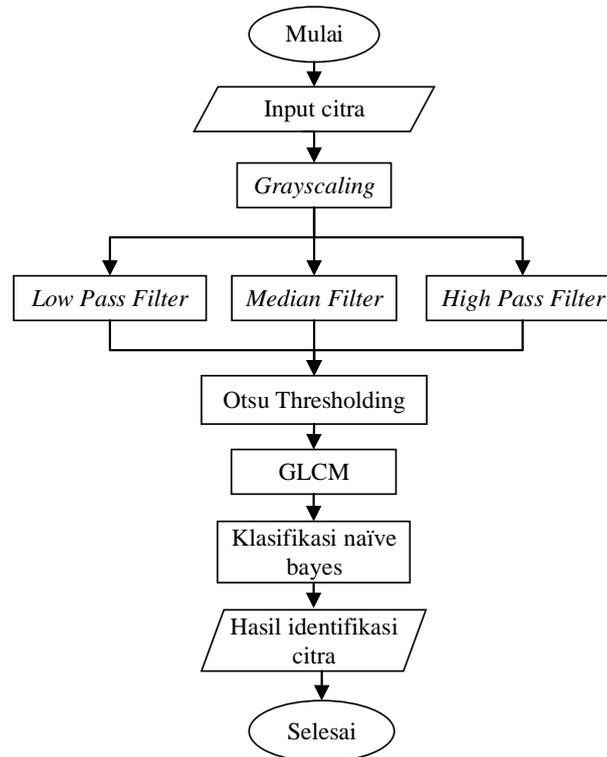
yang digabungkan dengan nilai HU yang akan menjadi input untuk proses segmentasi. Hasil yang diperoleh berdasarkan perhitungan akurasi tertinggi sebesar 77,43%.

Grey Level Co-Occurrence Matrix (GLCM) adalah prosedur aritmatika yang secara umum dapat menghilangkan artefak secara efektif. Selain itu, juga dapat membedakan tekstur gambar dengan jelas. GLCM dapat menentukan frekuensi *pixel* khususnya dalam area yang berbeda. Dalam penelitian mengenai validitas dan efisiensi fuzzy-GLCM dalam identifikasi dari *lung carcinoma* menggunakan citra bronkoskopi membuktikan bahwa metode tersebut menunjukkan 98% dalam hal akurasi, spesifitas, sensitivitas dan presisi jika dibandingkan dengan metode lain^[12]. Kemudian untuk pengkalsifikasian citra medis yang mengidentifikasi normal atau abnormal objek digunakan naïve bayes. Naïve bayes adalah klasifikasi berdasarkan statistik dan probabilitas sederhana yang banyak digunakan untuk metode pembelajaran identifikasi citra medis saat ini. Kelebihan lain dari naïve bayes adalah perhitungannya dapat dilakukan secara paralel dan terdistribusi. Dalam penerapannya pada klasifikasi citra juga mendapatkan hasil yang relatif sangat baik^[13].

Penelitian sebelumnya mengenai deteksi keberadaan nodul paru menggunakan teknik segmentasi Otsu Thresholding dan ekstraksi ciri yang disebut *Local Ternary Co-occurrence Pattern* (LTCoP) serta klasifikasi Support Vector Machine (SVM) diperoleh hasil bahwa fitur LTCoP mendeteksi paru nodul dengan akurasi 91,5%^[14]. Terdapat penelitian yang dilakukan oleh Sofian dan Laluma mengenai klasifikasi hasil citra MRI otak untuk memprediksi jenis tumor otak dengan metode *image threshold* dan GLCM menggunakan algoritma K-NN (Nearest Neighbor) Classifier berbasis web^[15]. Dilakukan penelitian data *training* sebanyak 20 data dan data testing sebanyak 10 data dan diperoleh tingkat akurasi sebesar 83,33%. Penelitian lain mengenai klasifikasi mengenai gambar dermoskopi menggunakan teknik naïve bayes dan *decision tree* menunjukkan tingkat akurasi naïve bayes mempunyai akurasi yang lebih tinggi daripada teknik *decision tree*^[16]. Hal tersebut menunjukkan klasifikasi naïve bayes menunjukkan akurasi diagnostik yang sangat baik. Pada penelitian ini dilakukan dengan mengkombinasikan antara segmentasi otsu thresholding, ekstraksi fitur GLCM dan klasifikasi naïve bayes yang telah dilakukan oleh penelitian-penelitian sebelumnya. Metode yang digunakan diharapkan dapat menghasilkan diagnosa citra kanker paru dan citra paru normal yang memiliki tingkat akurasi lebih baik.

METODE

Diagram alir tahapan pelaksanaan penelitian untuk klasifikasi citra kanker paru dan paru normal ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir penelitian.

Metode yang diusulkan mengklasifikasikan citra berdasarkan prosedur berikut. Akuisisi data dilakukan untuk citra berformat .jpg dan .png. kemudian citra hasil akuisisi data diolah dan dilakukan tahapan *image pre-processing* untuk meningkatkan kualitas citra dan menghilangkan noise dengan melakukan variasi *low pass filtering*, *median filtering*, dan *high pass filtering*. Tahapan berikutnya yaitu segmentasi dengan melakukan *otsu thresholding* dan ekstraksi ciri fitur untuk mencari karakteristik yang dimiliki oleh citra. Kemudian database yang diperoleh dari proses ekstraksi fitur diklasifikasikan menggunakan naïve bayes untuk menentukan citra paru normal atau citra kanker paru.

Dataset

Data citra yang digunakan dalam penelitian adalah data citra CT Scan dengan format .jpg dan .png yang diunduh dari <http://nbia.cancerimagingarchive.net/nbia-search/>. Data yang digunakan berupa gabungan citra paru normal dan citra kanker paru sejumlah 120 data dengan rincian 60 data citra kanker paru dan 60 data citra paru normal.

Image pre-processing

Tahapan ini bertujuan untuk menghasilkan citra yang lebih baik sebelum diproses pada tahapan selanjutnya. *Grayscale* yaitu tahapan untuk mengubah tipe citra dari RGB menjadi grayscale yaitu penyeragaman warna keabuan pada citra yang akan diproses untuk mempermudah proses pengolahan citra. Citra *grayscale* adalah matrik dengan setiap piksel yang dipresentasikan sebagai level diskrit dari 0 hingga 255. Nilai minimum mewakili warna yang paling gelap, sedangkan nilai tertinggi mewakili warna yang paling terang^[17]. Setelah dilakukan proses *grayscale*, menghilangkan *noise* merupakan langkah pertama dalam meningkatkan kualitas citra. Proses selanjutnya dilakukan dengan proses *filtering*, yaitu proses *low pass filter*, *median filter*, dan *high pass filter*. Pada proses *low pass filter*, proses ini memfilter sinyal frekuensi rendah dan melemahkan sinyal dengan frekuensi tinggi daripada frekuensi ambang^[18].

$$H(u, v) = \begin{cases} 1 & \text{if } D(u, v) \leq D_0 \\ 0 & \text{if } D(u, v) > D_0 \end{cases} \quad (1)$$

Dimana $D(u, v)$: jarak dari titik (u, v) menuju titik tengah dari frekuensi.

Median filter salah satu teknik yang efektif untuk mengurangi *noise* pada gambar tanpa menghilangkan tepinya^[19]. Setiap piksel gambar keluaran pada *median filter* diatur sesuai dengan nilai rata-rata piksel dari gambar masukan. *Median filter* lebih sensitif terhadap nilai rata-rata dan kurang sensitif terhadap nilai ekstrim piksel yang membantu dalam pengurangan *noise*^[20]. Nilai median dihitung dengan memilih nilai tengah untuk mengubah tingkat intensitas piksel (x, y) . Persamaan untuk *median filter* yaitu.

$$y[m, n] = \text{median}\{x[t, f]\}, \quad (t, f) \in \omega \quad (2)$$

Dimana ω merupakan piksel tetangga, yang ditetapkan pada lokasi $[m, n]$ ^[21].

High pass filter mempertahankan frekuensi tinggi dan banyak digunakan dalam penajaman citra. Keterbatasan utama dari *high pass filter* adalah menghilangkan sebagian latar belakang. *High pass filter* membuat tepi dan batas citra menjadi lebih cerah. Beberapa *noise* yang juga diamati dalam histogram dari citra *high pass filter* membuat citra lebih jelas dengan tepi yang lebih tajam^[22].

Segmentasi

Setelah dilakukan proses perbaikan citra, selanjutnya dilakukan segmentasi citra. Proses segmentasi memudahkan untuk menganalisis lebih lanjut dan mengenali informasi yang terdapat dalam citra. Metode *thresholding* adalah teknik segmentasi citra yang paling sederhana. Teknik ini melibatkan nilai ambang batas yang digunakan untuk mengubah citra skala abu-abu menjadi citra biner^[23]. Pada penelitian ini digunakan *otsu thresholding*. Konsep *otsu thresholding* adalah mengelompokkan citra biner secara otomatis berdasarkan bentuk histogramnya, dengan asumsi citra tersebut mengandung dua kelas dasar histogram bimodal (*foreground* dan *background*)^[24]. *Otsu thresholding* membuat gambar biner dari yang tingkat abu-abu dengan mengubah semua piksel di bawah ambang batas menjadi nol dan semua piksel di atas ambang batas menjadi satu. Proses tersebut dapat ditulis secara matematis^[25].

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{if } f(x, y) < T \end{cases} \quad (3)$$

Dimana $g(x, y)$ merupakan hasil piksel *threshold* dan $f(x, y)$ merupakan nilai piksel citra input pada baris x dan kolom y . T adalah nilai *threshold* sebagai batas nilai keabuan untuk mengkonversi nilai piksel $f(x, y)$ menjadi $g(x, y)$ ^[25].

Ekstraksi Fitur

Proses ekstraksi fitur berperan penting dalam pengolahan citra digital. Akurasi klasifikasi tergantung pada kualitas nilai ekstraksi fitur yang dihasilkan. *Gray Level Co-occurrence Matrix* (GLCM) digunakan untuk evaluasi fitur gambar yang berkaitan dengan statistik orde kedua. Langkah-langkah dalam melakukan GLCM yaitu^[26]:

1. Pembacaan input citra yang merupakan hasil dari proses *otsu thresholding*.
2. Nilai *gray-level* ditentukan dengan nilai tertinggi yakni 256 yang digunakan untuk membangun matrik framework.

3. Menentukan arah dan jarak piksel referensi dengan piksel tetangga. Arah yang digunakan yakni 0° , 45° , 90° , dan 135° dengan jarak piksel yang digunakan yakni 1.
4. Menghitung jumlah nilai kookurensi berdasarkan arah dan jarak yang telah ditentukan.
5. Membagi nilai kookurensi matrik simetris dengan jumlah keseluruhan nilai kookurensi untuk diperoleh matrik normal
6. Menghitung 11 fitur statistik, yaitu *conctrast*, *correlation*, *energy*, *homogeneity*, *mean*, *standart deviation*, *entropy*, *variance*, *smoothness*, *kurtosis*, *skewness*.
7. Fitur statistik dihitung untuk masing-masing matrik kookurensi pada arah yang sudah ditentukan.

Bagian citra yang digunakan dalam perhitungan matriks kookurensi adalah bagian kanker yang digambarkan pada citra input yang telah melalui proses *pre-processing*. Teknik yang digunakan dalam penelitian ini menggunakan 11 fitur GLCM yaitu:

1. Kontras (*conctrast*), merupakan jumlah ragam intensitas keabuan dalam citra^[27].

$$\sum_{i_1} \sum_{i_2} (i_1 - i_2)^2 p(i_1, i_2) \quad (4)$$

2. Korelasi (*correlation*), merupakan keterkaitan linier pada derajat keabuan citra^[28].

$$\sum_i \sum_j \frac{(i - \mu_i)(j - \mu_j) p_{(ij)}}{\sigma_i \sigma_j} \quad (5)$$

3. Energi (*energy*), merupakan nilai konsentrasi pada pasangan intensitas matriks kookurensi^[27].

$$\sum_{i_1} \sum_{i_2} p^2(i_1, i_2) \quad (6)$$

4. Homogenitas (*homogeneity*), merupakan ukuran nilai kesamaan variasi dari intensitas citra^[28].

$$\sum_{i_1} \sum_{i_2} \frac{p(i_1, i_2)}{1 + |i_1 - i_2|} \quad (7)$$

5. Rata-rata (*mean*), merupakan nilai rata-rata yang ada pada citra^[29].

$$\sum_i \sum_j (i, j) p(i, j) \quad (8)$$

6. Standar deviasi (*standart deviation*), merupakan ragam histogram dari suatu citra^[29].

$$\sqrt{\sum_{i=1}^L \sum_{j=1}^L (i - \mu)^2 p(i, j)} \quad (9)$$

7. Entropi (*entropy*), merupakan ketidakteraturan atas keabuan yang terdapat dalam citra^[28].

$$\sum_{i_1} \sum_{i_2} p(i_1, i_2) \log p(i_1, i_2) \quad (10)$$

8. *Variance*, merupakan variasi dari elemen matriks kookurensi dimana jika citra memiliki transisi derajat keabuan kecil akan memiliki variasi yang kecil^[27].

$$\sum_i \sum_j (i - \mu)^2 p(i, j) \quad (11)$$

9. Kehalusan (*smoothness*), merupakan tingkat kehalusan dari suatu intensitas citra^[29].

$$R = 1 - \frac{1}{1 + \sigma^2} \quad (12)$$

10. Kurtosis (*kurtosis*), merupakan tingkat keruncingan relatif kurva histogram yang ada pada suatu citra^[27].

$$\frac{1}{\sigma^4} \sum_i \sum_j (i - \mu)^4 p(i, j) - 3 \quad (13)$$

11. *Skewness*, merupakan tingkat kemiringan relatif kurva histogram suatu citra^[27].

$$\frac{1}{\sigma^3} \sum_i \sum_j (i - \mu)^3 p(i, j) \quad (14)$$

Dimana: i/j = nilai derajat keabuan; $p(i, j)$ = nilai histogram; L = nilai tertinggi dari intensitas; σ = standar deviasi; dan μ = mean

Klasifikasi

Tahap terakhir untuk mengklasifikasikan citra kanker dan citra normal, beberapa data di input sebagai data latih, kemudian informasi yang diperoleh dari proses training tersebut digunakan sebagai acuan untuk identifikasi kanker paru dengan menggunakan klasifikasi naïve bayes.

1. Tahap pelatihan naïve bayes

Langkah-langkah pelatihan data dengan naïve bayes yaitu sebagai berikut^[27]:

1. Hitung probabilitas (*prior*) tiap kelas yang ada.
2. Hitung rata-rata tiap fitur dan tiap kelas dengan persamaan

$$\mu = \frac{\sum n}{k} \quad (15)$$

Dimana: k = banyaknya data; $\sum n$ = jumlah nilai data.

3. Hitung nilai standar deviasi tiap fitur dan tiap kelas dengan persamaan

$$\sigma = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2 \quad (16)$$

Pada penelitian ini terdapat 100 data yang dijadikan data *training set* dan data tersebut berhasil diklasifikasikan dengan baik yang selanjutnya program tersebut akan dijadikan program untuk pengujian data *testing set*.

2. Tahap pengujian naïve bayes

Data dari set uji kemudian dilakukan pengujian dengan menggunakan metode naïve bayes. Proses ini merupakan proses pengujian dengan menghitung nilai peluang terbesar dimana akan mengklasifikasikan citra sesuai pada kelasnya. Langkah pengujian yang dilakukan menggunakan metode naïve bayes yaitu^[27]:

1. Menghitung peluang fitur terhadap kelas

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{1}{2}\left(\frac{X_i - \mu_{ij}}{\sigma_{ij}}\right)^2} \quad (17)$$

2. Menghitung peluang citra terhadap kelas

$$P(\text{fitur} | \text{kelas}) = \frac{1}{n} \sum_{i=1}^n P(X_i = x_i | Y = y_i) \quad (18)$$

Tingkat akurasi sistem mengenali *input* yang diberikan sehingga menghasilkan *output* yang benar secara matematis dituliskan sebagai berikut:

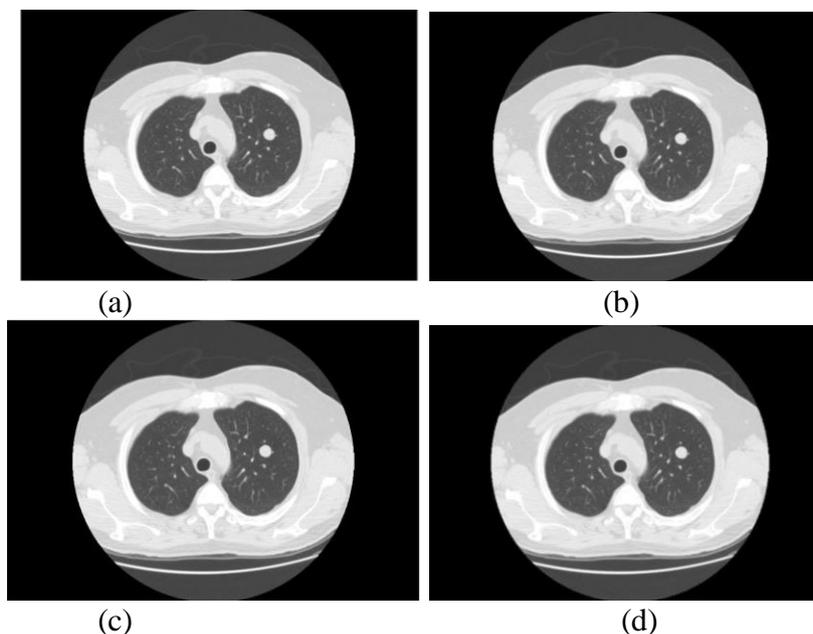
$$\text{akurasi} = \frac{\text{jumlah data yang benar}}{\text{jumlah data keseluruhan}} \times 100\% \quad (19)$$

Untuk nilai error yang merupakan tingkat kesalahan sistem dalam mengenali data *input* dirumuskan sebagai berikut:

$$\text{error} = \frac{\text{jumlah data yang salah}}{\text{jumlah data keseluruhan}} \times 100\% \quad (20)$$

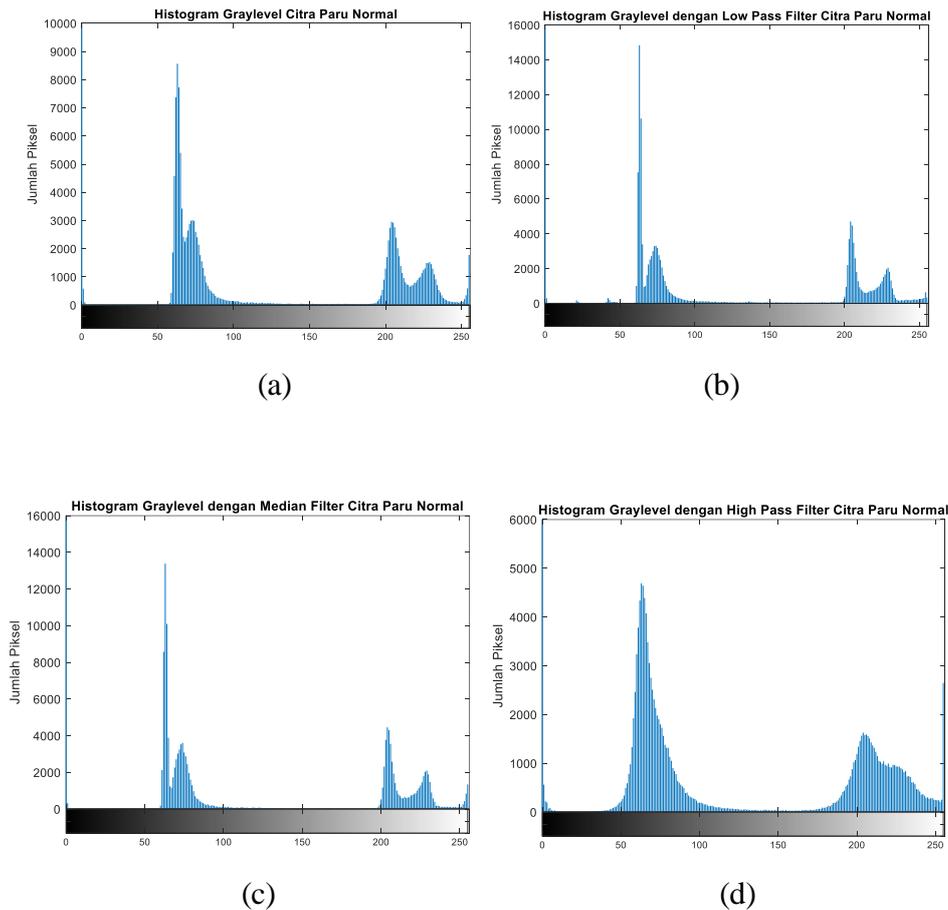
HASIL DAN PEMBAHASAN

Penelitian yang diusulkan telah menyelesaikan prosedur untuk mengklasifikasikan citra CT Scan paru normal dan kanker paru menggunakan program berbasis MATLAB 2018a melalui tahapan *filtering*, metode *otsu thresholding*, ekstraksi fitur GLCM dan klasifikasi naïve bayes. Pada penelitian ini menggunakan sebanyak 120 data citra yang terdiri dari 60 data citra kanker paru dan 60 data citra paru normal. Sampel citra berformat .jpg atau .png seperti pada gambar 2(a) yang diinputkan kemudian diubah ke *grayscale* yang akan diproses ke dalam *low pass filter*, *median filter*, dan *high pass filter* seperti yang ditunjukkan pada gambar 2(a), 2(b), dan 2(c).



Gambar 2. (a) Citra masukan, (b) hasil *low pass filtering*, (c) hasil *median filtering*, (d) hasil *high pass filtering*.

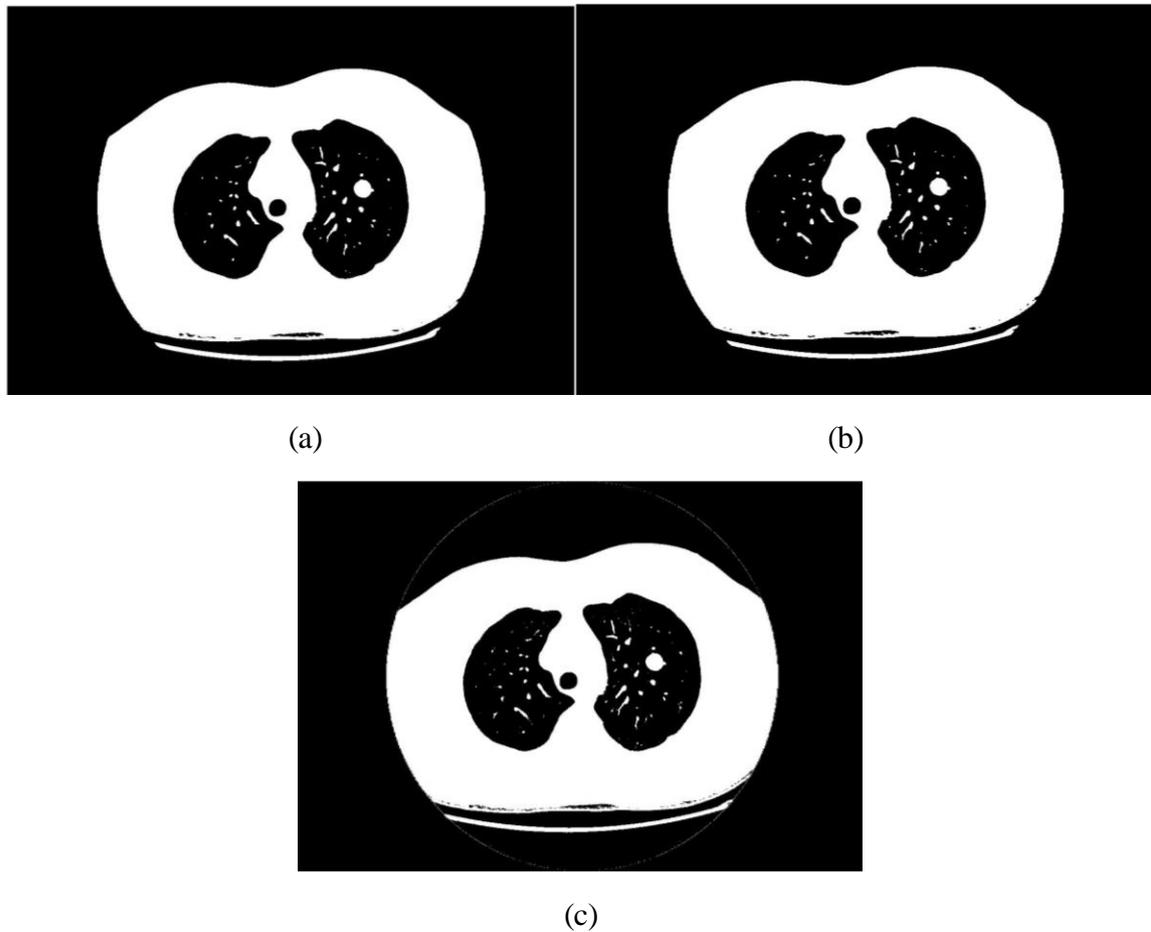
Hasil *filtering* dapat dilihat perbedaannya berdasarkan hasil histogram yang dihasilkan pada gambar 3.



Gambar 3. Histogram hasil proses (a) citra masukan, (b) hasil *low pass filtering*, (c) hasil *median filtering*, (d) hasil *high pass filtering*.

Gambar 3 menunjukkan hasil histogram hubungan antara jumlah piksel dan intensitas. Distribusi nilai intensitas menunjukkan berada pada *range* gelap dan terang secara bersamaan. Histogram hasil *low pass filtering* dan *median filtering* mempunyai nilai intensitas dan jumlah piksel yang hampir sama. Hal ini dikarenakan pada kedua *filter* mereduksi *noise* dengan membuat intensitas citra semakin merata dan menghaluskan citra. Hasil citra segmentasi *grayscale* dan *low pass filter* didapatkan citra yang lebih sedikit *noise*-nya dan mempunyai nilai akurasi rata-rata yang lebih tinggi^[28]. Pada citra MRI, proses pengolahan citra menggunakan *median filter* mampu menghilangkan *noise* berupa *salt and papper* pada citra^[29]. Pada proses *high pass filter*, citra yang dihasilkan lebih tajam daripada citra asli. Dari hal ini dapat diketahui bahwa pada proses *high pass filter* bagian piksel-piksel tepi ditampilkan lebih terang sedangkan piksel yang bukan daerah tepi dibuat lebih gelap sehingga membuat citra menjadi lebih tajam. Perbaiki citra dengan metode *high pass filter*, citra yang dihasilkan dalam penelitian terlihat terlalu tajam dari citra aslinya.

Pada proses segmentasi, digunakan otsu thresholding yang merupakan proses mengubah citra dari *grayscale* menjadi citra biner diperoleh hasil seperti yang ditunjukkan pada gambar 4.



Gambar 4. (a) Hasil *otsu thresholding* menggunakan *low pass filter*, (b) hasil *otsu thresholding* menggunakan *median filter*, (c) hasil *otsu thresholding* menggunakan *high pass filter*.

Dalam proses *otsu thresholding* terpisah dengan jelas bagian paru-paru dengan latar belakang dari gambar. Pada gambar 4 dapat dilihat bahwa tepi gambar terdeteksi dengan jelas setelah menggunakan metode *otsu thresholding* untuk memungkinkan pemisahan antara bagian paru dan bagian gambar lainnya. CT Scan paru terdiri dari piksel intensitas tinggi pada tubuh dan piksel intensitas rendah di paru-paru dan rongga sekitarnya yang dapat dipisahkan menggunakan proses *thresholding* dengan secara otomatis menghitung nilai ambang batas optimalnya^[30].

Setelah melalui tahap *pre-processing* maka informasi nilai ekstraksi fitur tekstur yang terdapat dalam citra akan diambil melalui metode *Gray Level Co-occurrence Matrix* (GLCM) dengan arah sudut 0° , 45° , 90° dan 135° . Hasil dari proses ekstraksi fitur ini nantinya akan menjadi database yang akan digunakan untuk *learning algorithm* pada proses klasifikasi citra.

Proses klasifikasi yang digunakan pada penelitian ini menggunakan klasifikasi naïve bayes. Hasil database yang diperoleh dari proses ekstraksi fitur GLCM dimasukkan ke dalam proses klasifikasi sebagai dataset. Dengan menggunakan metode naïve bayes, akan menghitung setiap peluang fitur yang ada sehingga memunculkan nilai baru untuk menghitung akurasi data. Dari proses klasifikasi yang dilakukan, diperoleh nilai akurasi yang ditunjukkan pada Tabel 1 berikut.

Tabel 1. Akurasi Hasil Pengujian

Arah sudut GLCM	Akurasi		
	Low	Medium	High
0°	87,50%	88,33%	85,83%
45°	87,50%	87,50%	85,00%
90°	85,83%	85,83%	85,83%
135°	87,50%	86,67%	85,00%

Tingkat akurasi yang diperoleh dari hasil klasifikasi dapat diketahui bahwa variasi median filter dengan arah sudut 0° memiliki tingkat akurasi yang paling tinggi dibandingkan dengan metode filtering lainnya dan arah sudut GLCM lainnya dengan tingkat akurasi sebesar 88,33%.

Tabel 2. Perbandingan Kinerja Sistem

Penelitian	Metode	Kinerja sistem
[31]	Median filter, GLCM, klasifikasi Instance based K-Nearest using Log and Gaussian weight Kernel (IbkLG)	Akurasi 86,6%
[32]	Median filter, Morfologi, GLCM, KNN	Akurasi 84,84% pada arah sudut 0°
[33]	GLCM, SVM	Akurasi 82,97% pada arah sudut 45°
Metode yang diusulkan	Median filter, Otsu thresholding, GLCM, Naïve Bayes	Akurasi 88,33% pada arah sudut 0°

Berdasarkan tabel 2 dapat dilihat bahwa metode yang diusulkan peneliti menunjukkan kinerja program yang lebih akurat. Hal tersebut menunjukkan bahwa program dapat membedakan antara citra kanker paru dan citra paru normal mempunyai keberhasilan yang lebih tinggi. Hal tersebut dapat dipengaruhi oleh banyaknya parameter ciri yang digunakan dalam proses ekstraksi tekstur, karena semakin banyak parameter ciri yang digunakan dapat meningkatkan hasil akurasi dari program.

KESIMPULAN

Pada penelitian ini untuk mengklasifikasikan citra kanker paru dan paru normal, pada tahap *preprocessing* dilakukan menggunakan variasi *low pass filter*, *median filter* dan *high pass filter* yang kemudian dilakukan proses segmentasi menggunakan *otsu thresholding*. Hal tersebut dapat menghasilkan citra biner yang mempunyai batas tepi yang jelas. Pada proses ekstraksi tekstur digunakan GLCM dengan arah sudut 0°, 45°, 90°, dan 135°. Berdasarkan klasifikasi yang dilakukan menggunakan metode naïve bayes, diperoleh nilai akurasi terbaik dari program menggunakan variasi *median filter* dengan segmentasi *otsu thresholding* serta ekstraksi fitur GLCM dengan arah sudut 0° dengan tingkat akurasi sebesar 88,33%, sehingga metode ini cukup baik digunakan dalam pengklasifikasian citra kanker paru dan paru normal hasil citra CT Scan.

DAFTAR PUSTAKA

- 1 Umbaugh, S.E., 2017. *Digital Image Processing and Analysis: Applications with MATLAB and CVIPtools*. Florida: CRC Press.
- 2 Putra, D. 2010. *Pengolahan Citra Digital*. Yogyakarta: C.V Andi Offset.
- 3 Nurhayati, O. D., & Susanto, A. 2008. Penerapan Metode Segmentasi Pada Analisis Citra Digital Head CT-Scan. *Disertasi. Universitas Gajah Mada, Yogyakarta*.
- 4 Rahmadewi, R. and Kurnia, R., 2016. Klasifikasi penyakit paru berdasarkan citra rontgen dengan metoda segmentasi sobel. *Jurnal Nasional Teknik Elektro*, 5(1), 7-12.
- 5 Gao, J., Wang, B., Wang, Z., Wang, Y. and Kong, F., 2020. A Wavelet Transform-Based Image Segmentation Method. *Optik*, 208, p.164123.
- 6 Nasser, I.M. and Abu-Naser, S.S., 2019. Lung cancer detection using artificial neural network. *International Journal of Engineering and Information Systems (IJEAIS)*, 3(3), 17-23.
- 7 Listyalina, L., 2017. Peningkatan Kualitas Citra Foto Rontgen Sebagai Media Deteksi Kanker Paru. *Respati*, 12(34).
- 8 Trisnawati, L. and Hakim, L., 2018. Segmentasi Citra Ct Scan Lung Menggunakan Deteksi Tepi Sobel Dan Metode Distance Regularized Level Set Evolution (Drlse). *Explore IT!: Jurnal Keilmuan dan Aplikasi Teknik Informatika*, 10(1), 1-13.
- 9 Matsumoto, M., 2010. Band-pass ε -filter for edge enhancement and noise removal. *IEICE transactions on information and systems*, 93(2), 367-375.
- 10 Wimmer, G., Tamaki, T., Tischendorf, J.J., Häfner, M., Yoshida, S., Tanaka, S. and Uhl, A., 2016. Directional wavelet based features for colonic polyp classification. *Medical image analysis*, 31, 16-36.
- 11 Kusuma, A.W. and Ellyana, R.L., 2018. Penerapan Citra Terkompresi Pada Segmentasi Citra Menggunakan Algoritma K-Means. *Jurnal Terapan Teknologi Informasi*, 2(1), 65-74.
- 12 Yamunadevi, M.M. and Ranjani, S.S., 2021. Efficient segmentation of the lung carcinoma by adaptive fuzzy-GLCM (AF-GLCM) with deep learning based classification. *Journal of Ambient Intelligence and Humanized Computing*, 12(5), 4715-4725.
- 13 Susanto, A., Dewantoro, Z.H., Sari, C.A., Setiadi, D.R.I.M., Rachmawanto, E.H. and Mulyono, I.U.W., 2020. Shallot Quality Classification using HSV Color Models and Size Identification based on Naive Bayes Classifier. In *Journal of Physics: Conference Series* (Vol. 1577, No. 1, p. 012020). IOP Publishing.
- 14 Bruntha, P.M., Pandian, S.I.A., Anitha, J., Mohan, P. and Dhanasekar, S., 2020. Local Ternary Co-occurrence Patterns based Lung Nodules Detection. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 489-492). IEEE.
- 15 Sofian, J. and Laluma, R.H., 2019. Klasifikasi Hasil Citra Mri Otak Untuk Memprediksi Jenis Tumor Otak dengan Metode Image Threshold Dan GLCM Menggunakan Algoritma K-NN (Nearest Neighbor) Classifier Berbasis Web. *Infotronik: Jurnal Teknologi Informasi dan Elektronika*, 4(2), 51-56.
- 16 Arasi, M.A., El-Horbaty, E.S.M. and El-Sayed, A., 2018. Classification of Dermoscopy Images Using Naïve Bayesian and Decision Tree Techniques. In *2018 1st Annual International Conference on Information and Sciences (AiCIS)* (7-12). IEEE.
- 17 Kalaivani, S., Chatterjee, P., Juyal, S. and Gupta, R., 2017. Lung cancer detection using digital image processing and artificial neural networks. In *2017 International*

- conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 2, 100-103). IEEE.
- 18 Makandar, A. and Halalli, B., 2015. Image enhancement techniques using highpass and lowpass filters. *International Journal of Computer Applications*, 109(14), 12-15.
 - 19 Sevani, A., Modi, H., Patel, S. and Patel, H., 2018. Implementation of image processing techniques for identifying different stages of lung cancer. *International Journal of Applied Engineering Research*, 13(8), 6493-6499.
 - 20 Pratap, G.P. and Chauhan, R.P., 2016. Detection of Lung cancer cells using image processing techniques. In *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)* (pp. 1-6). IEEE.
 - 21 Hussain, A. and Khunteta, A., 2020. Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (38-43). IEEE.
 - 22 Singh, S., Sharma, A. and Mittal, M., 2017. Performance Evaluation of High Pass, Low Pass and Median filter on Webcam Pictures.
 - 23 Ganesh, P.S., Kumar, T.S., Kumar, M. and Kumar, M.S.R., 2021. Brain Tumor Detection and Classification Using Image Processing Techniques. *Brain*, 4(3).
 - 24 Rulaningtyas, R. and Ain, K., 2021. CT scan image segmentation based on hounsfield unit values using Otsu thresholding method. In *Journal of Physics: Conference Series* (Vol. 1816, No. 1, p. 012080). IOP Publishing.
 - 25 Arifin, T., 2016. Analisa Perbandingan Metode Segmentasi Citra Pada Citra Mammogram. *Jurnal Informatika*, 3(2).
 - 26 Zotin, A., Hamad, Y., Simonov, K. and Kurako, M., 2019. Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural networks. *Procedia Computer Science*, 159, 1439-1448.
 - 27 Rasmi, R.P., 2020. Peningkatan Hasil Diagnosis Kanker Payudara Dari Hasil Citra Mammogram Menggunakan Metode Ekstraksi Ciri Dan Klasifikasi.
 - 28 Wijaya, R. S. D., Suksmono, A. B., and Mengko., T. L. 2021. Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 139-147.
 - 29 Hussain, A., and Khunteta, A., 2020. Semantic Segmentation of Brain Tumor from MRI Images and SVM Classification using GLCM Features. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 38-43). IEEE.
 - 30 Kanagaraj, G., and Kumar, P. S. 2020. Pulmonary tumor detection by virtue of GLCM. *Journal of Scientific & Industrial Research*, 79, 132-134
 - 31 Chauhan, S., More, A.,Uikey R., Malviya, P., and Moghe, A., 2017. Brain tumor detection and classification in MRI images using image and data mining. *International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, 2017, pp. 223-231.
 - 32 Afriyana, Y., Purnamasari, R., dan Patmasari, R. 2018. Deteksi Kelainan Tulang Belakang Berdasarkan Citra Medis Digital Dengan Menggunakan Gray Level Co-occurrence Matrix (GLCM) dan K-nearest Neighbor (KNN). *eProceedings of Engineering*, 5(3). 4675-4682.
 - 33 Augustin, I., Hidayat, B., dan Oscandar, F. 2018. Identifikasi Jenis Kelamin Berdasarkan Teraan Gigitan Berbasis Pengolahan Citra Digital Dengan Metode Gray Level Co-Occurrence Matrix (GLCM) dan Klasifikasi Support Vector Machine (SVM). *eProceedings of Engineering*, 5(3). 4835-4842.