



Deteksi Nodul Paru pada Citra CT dengan Klasifikasi *Pseudo Nearest Neighbour Rule*

I Nyoman Surya Jaya ^{1*}, Kadek Yota Ernanda Aryanto ², dan Dewa Gede Hendra Divayana ³

1. Universitas Pendidikan Ganesha; jayaslx777@gmail.com
 2. Universitas Pendidikan Ganesha; yota.ernanda@undiksha.ac.id
 3. Universitas Pendidikan Ganesha; hendra.divayana@undiksha.ac.id
- * Korespondensi: jayaslx777@gmail.com

Sitasi: Jaya, I. N. S.; Aryanto, K. Y. E.; Divayana, D. G. H. (2024). Deteksi Nodul Paru pada Citra CT Dengan Klasifikasi *Pseudo Nearest Neighbour Rule*. JTIM: Jurnal Teknologi Informasi Dan Multimedia, 5(4), 312-321.

<https://doi.org/10.35746/jtim.v5i4.463>

Diterima: 15-12-2023

Direvisi: 09-01-2024

Disetujui: 15-01-2024



Copyright: © 2024 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Abstract: This research aims to obtain the classification performance of the Pseudo Nearest Neighbor Rule (PNNR) algorithm in detecting lung nodules in CT scan images. The PNNR classification algorithm is used to reduce the influence of noise or outliers in the classification process so that false positives (prediction of an object that is not a nodule as a nodule) can be reduced. The data set used is 200 patient data obtained from the public data of The Lung Image Database Consortium and Infectious Disease Research Institute (LIDC/IDRI) where 4 fold Cross Validation will be carried out. The preprocessing stage is carried out by segmenting the otsu image, where from the segmentation results the two largest blobs are then searched for to determine the area of the lung to be analyzed. Next, the feature extraction process from the candidate nodules (white pixels / foreground) is obtained from the Otsu segmentation process again. The results of this second segmentation contain information from the candidate nodules to then calculate the value of the shape features of the candidate nodules such as area, eccentricity, equivalent diameter, major axis length, minor axis length and perimeter which produces feature set values as the basis for training data and data test for the classification process in PNNR. The results of the classification proposed in this research, namely using the PNNR classification method, obtained an Accuracy value of 93.72%, which is included in the excellent classification level or the Accuracy level is very good but with a lower level of sensitivity or recognition of true positives, namely 72%. In further research, classification optimization can be carried out by selecting the feature set used

Keywords: *Pulmonary Nodule Detection, Classification, Pseudo Nearest Neighbour Rule*

Abstrak: Penelitian ini bertujuan untuk memperoleh performansi klasifikasi algoritma *Pseudo Nearest Neighbour Rule (PNNR)* dalam mendeteksi nodul paru pada citra *CT scan*. Algoritma klasifikasi PNNR digunakan untuk mengurangi pengaruh derau atau pencilan dalam proses klasifikasi sehingga *false positive* (prediksi suatu objek bukan nodul sebagai nodul) dapat dikurangi. Data set yang digunakan adalah 200 data pasien yang diperoleh dari data publik *The Lung Image Database Consortium and Infectious Disease Research Institute (LIDC/IDRI)* di mana akan dilakukan *4 fold Cross Validation*. Untuk tahapan *preprocessing* dilakukan dengan segmentasi citra *otsu* di mana dari hasil segmentasi tersebut kemudian dicari dua blob terbesar untuk menentukan area paru yang akan dianalisis. Selanjutnya untuk proses ekstraksi fitur dari calon nodul (bagian piksel putih / *foreground*) diperoleh dari proses segmentasi *otsu* lagi. Hasil dari segmentasi yang kedua ini mengandung informasi dari kandidat nodul untuk selanjutnya dihitung nilai dari fitur bentuk dari kandidat nodul tersebut seperti *area, eccentricity, equivalent diameter, major axis length, minor axis length dan perimeter* yang menghasilkan nilai *feature set* sebagai dasar data latih dan data uji untuk proses klasifikasi pada PNNR. Hasil klasifikasi yang diusulkan dalam penelitian ini yaitu dengan menggunakan metode klasifikasi PNNR mendapatkan nilai akurasi sebesar 93,72% termasuk ke

dalam tingkat nilai *excellent classification* atau tingkat akurasi yang sangat baik namun dengan tingkat *sensitivity* atau pengenalan *true positive* yang lebih rendah yaitu sebesar 72%. Pada Penelitian selanjutnya dapat dilakukan optimasi klasifikasi dengan pemilihan *feature set* yang digunakan.

Kata kunci: Deteksi Nodul Paru, Klasifikasi, Pseudo Nearest Neighbour Rule

1. Pendahuluan

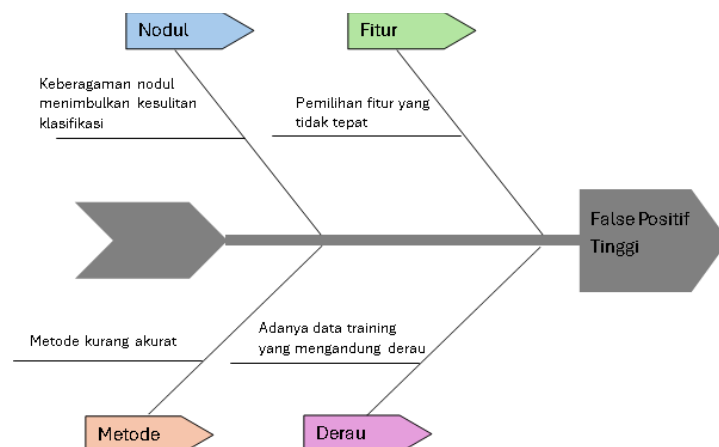
Nodul paru (*pulmonary nodule*) merupakan gumpalan yang terdapat di paru-paru berukuran ± 3 cm [1]. Nodul Paru digunakan sebagai penanda atau gejala awal untuk penyakit paru-paru tertentu seperti Bronkitis, *Cystic fibrosis*, *Emphysema*, *Pneumonia*, Tuberkulosis, edema paru atau bahkan kanker paru-paru [2]. Oleh karena itu deteksi nodul paru sangat penting dilakukan sebagai gejala awal suatu penyakit.

Banyak penelitian yang telah dilakukan dikhususkan untuk mengembangkan skema *computer aided detection* (CAD) untuk mendeteksi nodul paru. Salah satu penelitian tentang deteksi nodul adalah dilakukan oleh [3] di mana dalam penelitian ini didapat *sensitivity* 95,9% tetapi dengan FP yang tinggi yaitu 38,8 *FP/Case*. [4] juga melakukan penelitian di mana dalam penelitian ini didapat sensitivitas yang lebih kecil yaitu 84,6 dengan FP yang lebih tinggi yaitu 47 *FP/Case*.

Keragaman nodul paru menimbulkan kesulitan untuk skema *computer-aided diagnostic* (CAD) untuk mendeteksi nodul paru pada citra *computed tomography* (CT) scan gambar, khususnya di penelitian skrining CT skala besar. Oleh karena itu [5] mengusulkan skema CAD baru berdasarkan metode hibrida untuk mengatasi tantangan deteksi pada keberagaman nodul paru. Dalam penelitiannya didapatkan *sensitivity* sebesar 85,2% dan FP sebesar 3,13.

Pemilihan metode klasifikasi dalam deteksi nodul juga penting dilakukan. Salah satu metode yang digunakan yaitu dengan *Deep Learning*. *Deep Learning* memiliki kemampuan yang sangat baik dalam visi komputer. Salah satunya adalah pada kasus klasifikasi objek pada citra. Metode *Deep Learning* yang saat ini memiliki hasil paling signifikan dalam pengenalan citra adalah *Convolutional Neural Network* (CNN) [6]. Pemilihan metode CNN sebagai deteksi nodul dilakukan oleh [7] di mana dalam penelitian ini didapat *sensitivity* yang masih rendah yaitu 80% dengan *FP/case* yang tinggi sebesar 15,28. Penelitian dengan menggunakan variasi dari CNN juga dilakukan oleh [8], [9], [10] di mana didapat *sensitivity* $\pm 90\%$ dan ± 4 FP per *case*. Secara keseluruhan, beberapa sistem menunjukkan potensial tetapi masih diperlukan perbaikan untuk mengatasi tantangan-tantangan tertentu seperti sensitivitas tinggi dengan laju *False positive* yang rendah.

Diagram *fishbone* dari permasalahan *false positive* yang tinggi dari deteksi nodul dapat dilihat pada Gambar 1.



Gambar 1. Fishbone Diagram False Positive dalam Deteksi Nodul

Dalam diagram diatas dapat dilihat ada beberapa hal yang dapat menyebabkan *false positive* yang tinggi dalam deteksi nodul seperti keberagaman nodul, pemilihan fitur yang tidak tepat, metode klasifikasi kurang akurat, dan adanya data *training* yang mengandung derau (*outlier*) sehingga menyebabkan kesalahan dalam klasifikasi. Berangkat dari uraian masalah dan rujukan penelitian sebelumnya, maka dalam penelitian ini diusulkan pendekatan baru dalam meningkatkan sensitivitas pendeteksian nodul. Algoritma yang digunakan yaitu dengan implementasi algoritma klasifikasi *Pseudo Nearest Neighbour Rule* (PNNR). Penelitian tentang perbandingan teknik klasifikasi dilakukan oleh Yong Zeng [11]. Dalam penelitiannya dilakukan perbandingan metode PNNR dengan K-NN, *distance weighted* K-NN, dan *local mean-based nonparametric classification* dalam hal akurasi klasifikasi pada pola yang tidak diketahui. Hasil pengujian mengkonfirmasi keabsahan peraturan klasifikasi baru ini bahkan dalam situasi praktis.

Algoritma klasifikasi PNNR digunakan untuk mengurangi pengaruh derau atau pencilan dalam proses klasifikasi sehingga *false positive* (prediksi suatu objek bukan nodul sebagai nodul) dapat dikurangi. Dengan berkurangnya *false positive* maka tingkat kesalahan atau kekeliruan klasifikasi juga dapat dikurangi. PNNR merupakan varian terbaru dari K-NN yang juga bekerja secara lokal. PNNR didesain untuk mengatasi kelemahan metode K-NN yang umumnya memberikan performansi rendah untuk data yang mengandung derau atau pencilan.

2. Bahan dan Metode

Penelitian ini bertujuan untuk merekayasa sebuah sistem untuk mendeteksi nodul paru dengan menggunakan algoritma *Pseudo Nearest Neighbour Rule*. Untuk mengukur tingkat kevalidan atau akurasi dari suatu dataset maka dilakukan dengan *k-fold Cross Validation*. Nilai *k* yang digunakan dalam *k-fold Cross Validation* adalah 4 sehingga menjadi *4 Fold Cross Validation*, yang artinya dataset akan dibagi kedalam empat kelompok data. Keempat kelompok data ini yang akan digunakan dalam iterasi dalam menghitung akurasi. Di mana setiap iterasi komposisi datanya adalah satu kelompok data akan menjadi data uji dan sisanya menjadi data *training*. Pada penelitian ini dilakukan dengan empat proses utama yaitu pengolahan data *training*, pengolahan data uji, klasifikasi dan pengujian sistem. Data *training* diperoleh dari citra yang telah dianotaskan koordinat nodul dan non nodulnya oleh radiologist. Area yang dianotaskan oleh radiologist tersebut kemudian dilakukan ekstraksi fitur sehingga diperoleh dataset *training*. Sedangkan data uji diperoleh dari ekstraksi fitur citra yang telah dibinerisasi dan thresholding otsu. Dataset *training* dan dataset uji tersebut kemudian dijadikan dasar dalam proses klasifikasi. Hasil dari rekayasa sistem klasifikasi tersebut kemudian dilakukan pengujian sistem dengan menghitung nilai ukuran *Accuracy*, *Error Rate*, *sensitivity* atau *true positive rate*, *Specificity*, dan *Precision*.

2.1. Pengolahan Data

Dalam penelitian ini digunakan data publik yang diperoleh dari *The Lung Image Database Consortium and Infectious Disease Research Institute (LIDC/IDRI)* melalui situs <https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI> yang terdiri dari kumpulan citra dari diagnosis dan CT Scan skrining thorac kanker paru.

Sistem ini dibangun dengan menggunakan sebanyak 200 file data pasien, di mana didalamnya diambil 200 obyek yang dinotaskan sebagai nodul dan bukan nodul oleh radiologist. Karena menggunakan 4 Fold Cross Validation maka tiap dataset (K) berjumlah 50 file. Untuk data training adalah sebanyak 150 file sedangkan untuk pengujian sistem maka diambil juga 50 file.

2.2. *Pseudo Nearest Neighbour Rule*

Pseudo nearest neighbour rule (PNNR) merupakan varian terbaru dari K-NN yang juga bekerja secara lokal. PNNR didesain untuk mengatasi kelemahan metode K-NN yang umumnya memberikan performansi yang rendah untuk data yang mengandung derau atau pencilan. Untuk memahami langkah-langkah PNNR secara terstruktur,

berikut ini diberikan algoritma PNNR yang diadopsi dari [11]. Misalkan tersedia N sampel yang memiliki label kelasnya, dan misalkan n_i adalah jumlah sampel yang termasuk dalam kelas ke- i , $n_i \geq k$, dan $i = 1, \dots, M$. Sedangkan k adalah jumlah sampel pada setiap kelas yang jaraknya paling dekat dengan sampel uji, yang digunakan untuk menghitung total jarak dalam proses klasifikasi. PNNR mengklasifikasikan pola uji sebagai kelas yang memiliki jarak total paling minimum (kelas r). Jika terdapat lebih dari satu kelas (dinyatakan sebagai kelas r sampai kelas $r + s$), maka PNNR akan memilih salah satu secara acak.

Pada algoritma PNNR, sejumlah k pola terdekat terhadap pola masukan diberi bobot berdasarkan peringkat jaraknya secara terurut naik. Bobot ketetanggaan untuk pola ke- j dirumuskan sebagai $u_j = \frac{1}{j}$. Dengan demikian pola terdekat ke-1 (yang paling dekat) diberi bobot 1, pola terdekat ke-2 diberi bobot $\frac{1}{2}$, dan bobot tersebut terus menurun untuk pola-pola yang lebih jauh sehingga pola terdekat ke- k diberi bobot terkecil $\frac{1}{k}$. Dengan cara ini pola-pola yang berada di peringkat atas (yang lebih dekat) memiliki pengaruh lebih besar dalam proses klasifikasi dibanding pola-pola yang berada di peringkat bawah (yang lebih jauh).

2.3. Crossvalidation

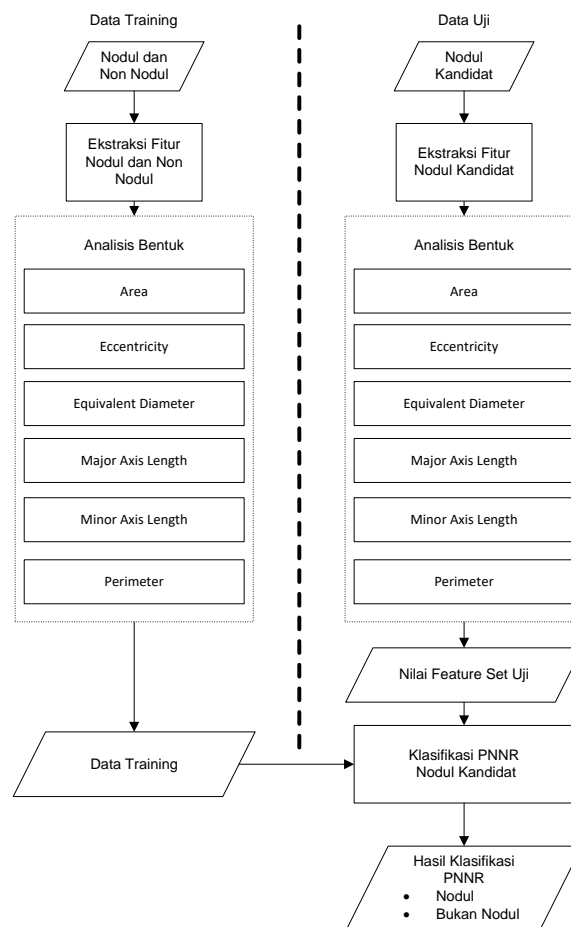
Validasi silang, atau dikenal juga sebagai estimasi rotasi, merupakan suatu metode untuk mengevaluasi sejauh mana hasil analisis statistik dapat menggeneralisasi kelompok data yang independen. Metode ini secara khusus digunakan untuk memprediksi model dan menilai seberapa tepat model prediktif tersebut saat diaplikasikan dalam situasi praktis [12]. Salah satu metode dalam validasi silang adalah k -fold cross-validation, di mana data dibagi menjadi K bagian set data dengan ukuran yang identik. Proses pelatihan dan pengujian dilakukan sebanyak K kali. Pada setiap iterasi, salah satu bagian dataset digunakan sebagai data pelatihan, sementara yang lainnya sebagai data pengujian [13]. Pada penelitian ini digunakan 4 -fold Cross Validation dalam mengevaluasi kinerja model. Pada 4 -fold Cross Validation data dibagi menjadi 4 -fold dengan ukuran sama, di mana 3 -fold akan digunakan sebagai data latih dan 1 -fold digunakan sebagai data uji. Skenario pengujian K -Folds Cross Validation yang akan dilakukan dapat dilihat pada Tabel 1 dibawah ini.

Tabel 1. Skema 4 Fold Cross Validation

Fold	K1	K2	K3	K4
1	Training	Training	Training	Uji
2	Training	Training	Uji	Training
3	Training	Uji	Training	Training
4	Uji	Training	Training	Training

2.4. Alur Klasifikasi

Pada tahap ini dilakukan proses klasifikasi untuk menentukan objek yang dianalisis merupakan nodul atau bukan. Nilai-nilai dari *feature set* yang diperoleh dari proses sebelumnya digunakan sebagai data *training* yang selanjutnya akan dijadikan dasar untuk melakukan proses klasifikasi. Proses klasifikasi menggunakan algoritma PNNR seperti ditunjukkan pada Gambar 2



Gambar 2. Tahapan Klasifikasi Nodul Paru menggunakan algoritma PNNR

3. Hasil

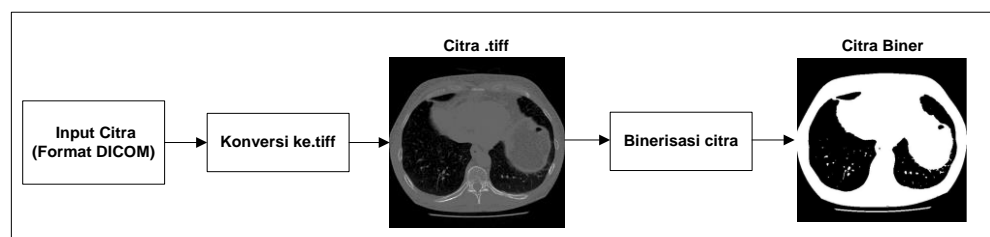
Untuk mengukur tingkat kevalidan atau akurasi dari suatu dataset maka dilakukan dengan *k-fold Cross Validation*. Nilai *k* yang digunakan dalam *k-fold Cross Validation* adalah 4 sehingga menjadi 4 *Fold Cross Validation*, yang artinya dataset dibagi kedalam empat kelompok data yaitu *K1, K2, K3, K4* dengan iterasi pengujian dilakukan sebanyak 4 kali di mana tiap dataset bergantian menjadi data latih dan data uji. Pada tahap ini nodul kandidat pada citra CT diklasifikasi sebagai nodul atau bukan. Model yang digunakan untuk klasifikasi ini adalah dengan PNNR.

3.1. Pre Processing Data Citra

Tahap *pre processing* dilakukan untuk mempersiapkan citra agar lebih mudah diolah pada proses selanjutnya. Tahap *pre-processing* terdiri dari tiga proses yaitu konversi citra ke .tiff dan binerisasi citra, Menentukan ROI Paru dan Segmentasi Otsu.

3.1.1. Konversi Citra Ke .Tiff Dan Binerisasi Citra

Tahap pertama yang dilakukan adalah dengan mengkonversi citra ke format .tiff dan binerisasi citra seperti ditunjukkan pada Gambar 3.

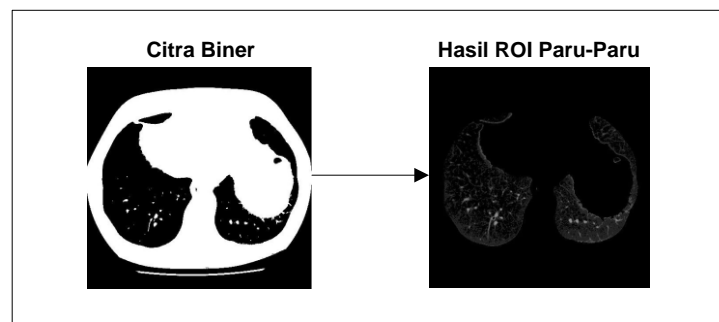


Gambar 3. Pre Processing Data Citra

Pada tahap pertama citra input berupa citra CT Scan Paru dengan format DICOM di konversi menjadi format .tiff untuk proses selanjutnya. Tahap kedua adalah binerisasi citra, di mana binerisasi citra ini dilakukan dengan menggunakan metode *thresholding* Otsu. *Thresholding* otsu dilakukan untuk memisahkan objek *foreground* dengan *background* di mana piksel dengan nilai dibawah ambang batas (*threshold*) ditransformasi menjadi hitam (*background*) dan piksel dengan nilai diatas ambang batas ditransformasi menjadi putih (*foreground*). Piksel putih menggambarkan objek (*foreground*) seperti nodul, pembuluh darah, jaringan paru-paru atau dinding toraks. Sedangkan piksel hitam (*background*) menggambarkan udara.

3.1.2. Menentukan ROI Paru

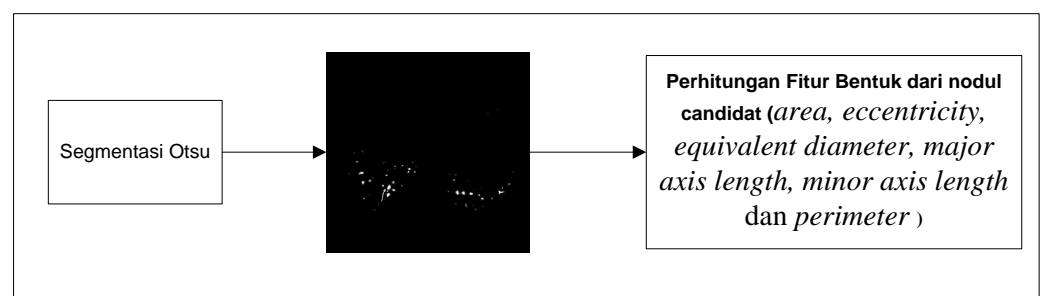
Pada tahap ini dilakukan penentuan terhadap *Region of Interest* (ROI) dari citra paru. ROI dilakukan dengan cara mencari kontur dari dua blob terbesar pada paru-paru. Istilah kontur identik dengan batas (*boundary*), di mana dalam hal ini adalah *boundary* terbesar dari objek pada citra toraks seperti ditunjukkan pada Gambar 4



Gambar 4 Menentukan ROI Paru

3.1.3. Segmentasi Otsu

Pada tahap ini dilakukan ekstraksi fitur dari calon nodul (bagian piksel putih / *foreground*) yang diperoleh dari proses segmentasi otsu. Fitur ini mengandung informasi dari calon nodul yang terdiri dari fitur bentuk dari calon nodul tersebut. Fitur bentuk adalah fitur yang diperoleh melalui bentuk objek di mana dalam penelitian ini meliputi enam atribut yaitu *area*, *eccentricity*, *equivalent diameter*, *major axis length*, *minor axis length* dan *perimeter* seperti ditunjukkan pada Gambar 5

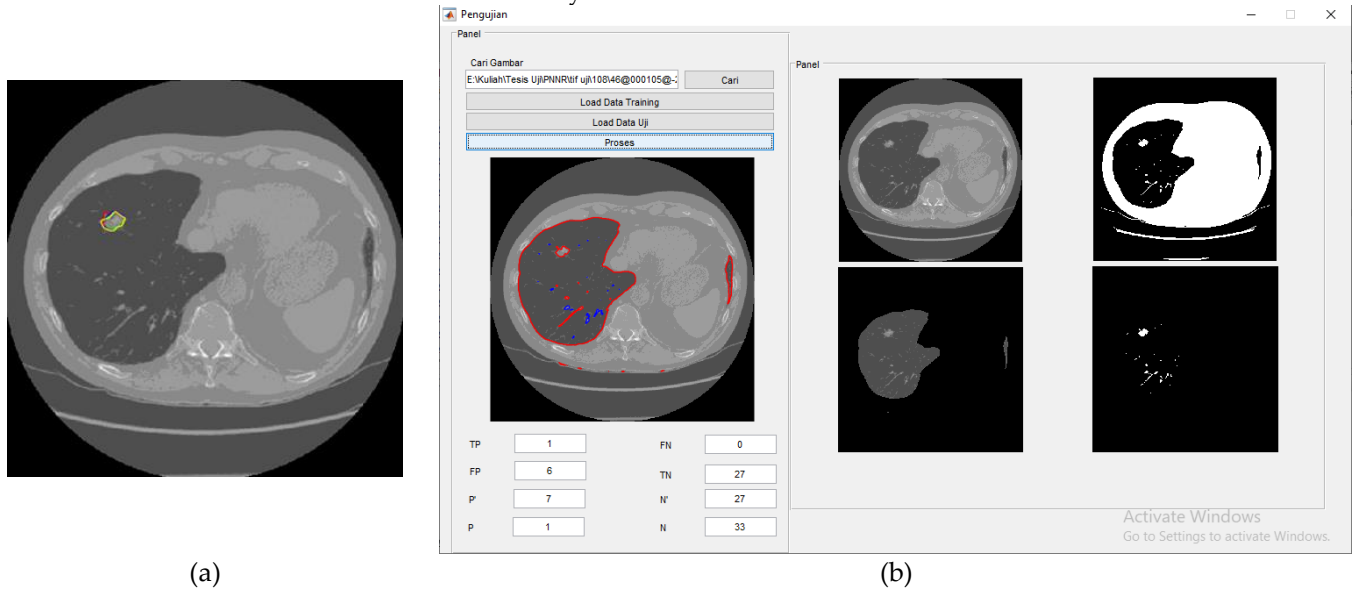


Gambar 5 Ekstraksi Fitur Nodul Kandidat

3.2. Antamuka Aplikasi

Sistem ini dibangun dengan menggunakan sebanyak 200 file data pasien, di mana didalamnya diambil 200 obyek yang dinotasikan sebagai nodul dan bukan nodul oleh radiologist. Karena menggunakan *4 Fold Cross Validation* maka tiap dataset (K) berjumlah 50 file. Untuk data training adalah sebanyak 150 file sedangkan untuk pengujian sistem maka diambil juga 50 file. Pada tahap ini dilakukan proses klasifikasi untuk menentukan

objek yang dianalisis merupakan nodul atau bukan. Nilai-nilai dari feature set yang diperoleh dari proses sebelumnya digunakan sebagai data training yang selanjutnya akan dijadikan dasar untuk melakukan proses klasifikasi. Hasil klasifikasi menggunakan algoritma PNNR seperti ditunjukkan pada Gambar 6 Untuk citra hasil klasifikasi PNNR (Classifier), objek yang ditandai warna merah menunjukkan nodul dan warna biru menunjukkan bukan nodul. Dari hasil validasi maka didapatkan jumlah tuple dari confusion matrix- nya berdasarkan notasi aktual dari radiolodist.



Gambar 6. Citra dengan Ploting Nodul Oleh Radiologist (a), Hasil Klasifikasi PNNR Citra LIDC-IDRI-0108 (b)

Dalam citra dari LIDC-IDRI terdapat informasi mengenai nodul dan bukan nodul yang dinotasikan oleh radiologist berupa koordinat tepi dari objek tersebut dan disimpan dalam file berformat xml. Informasi dari radiologist tersebutlah yang dijadikan acuan untuk validasi hasil klasifikasi. Validasi hasil klasifikasi ini dilakukan untuk mengetahui apakah hasil klasifikasi dari model yang digunakan untuk mengidentifikasi kandidat nodul sebenarnya adalah nodul atau bukan. Hal ini dilakukan dengan membandingkan nilai centroid kandidat nodul hasil klasifikasi dengan nilai centroid nodul yang telah dianotasi oleh radiologis. Kandidat nodul yang memiliki nilai centroid mendekati unit anotasi dinyatakan sebagai benar adalah nodul. Hasil dari validasi ini akan menghitung berapa jumlah dari true positive, true negative, false positive dan false negative.

3.3. Perbandingan dan Analisis Hasil Klasifikasi PNNR

Pada pengujian 4 Fold Cross Validation digunakan 150 data latih dan 50 data uji. Proses Klasifikasi PNNR dilakukan pada ke 50 data citra uji nilai k =1 sampai dengan nilai k = 10. Berdasarkan keseluruhan tahapan klasifikasi data uji yang telah dilakukan dalam penelitian ini , diperoleh hasil pengenalan objek yang dapat dilihat pada Tabel 2 berikut ini.

Tabel 2. Perbandingan Hasil Pengenalan Objek

Fold	k	FN	TN	TP	FP	P''	P	N''	N
Fold 1	5	17	5487	33	318	351	50	5504	5805
Fold 2	2	9	5502	41	303	344	50	5511	5805
Fold 3	6	20	4513	30	212	242	50	4533	4725
Fold 4	2	17	2976	33	288	321	50	2993	3264

Pada Tabel 2 dapat dilihat bahwa pada pengujian Klasifikasi PNNR fold ke 1 diperoleh hasil terbaik pada $k = 5$ dengan TP sebanyak 33, pada fold ke 2 diperoleh hasil terbaik pada $k = 2$ dengan TP sebanyak 41, pada fold ke 3 diperoleh hasil terbaik pada $k = 6$ dengan TP sebanyak 30 dan pada fold ke 4 diperoleh hasil terbaik pada $k = 2$ dengan TP sebanyak 33.

Untuk ukuran evaluasi klasifikasi PNNR didapat berdasarkan perhitungan sebagai berikut [14] :

a. Accuracy

$$\frac{TP + TN}{P + N} \times 100 \% \tag{1}$$

b. Error Rate

$$\frac{FP + FN}{P + N} \times 100 \% \tag{2}$$

c. Recall atau sensitivity atau true positive rate

$$\frac{TP}{P} \times 100 \% \tag{3}$$

d. Specificity atau true negative rate

$$\frac{TN}{N} \times 100 \% \tag{4}$$

e. Precision

$$\frac{TP}{TP + FP} \times 100 \% \tag{5}$$

Perbandingan hasil ukuran evaluasi dari model klasifikasi PNNR dapat ditunjukkan seperti pada Tabel 3 berikut.

Tabel 3. Perbandingan Hasil Ukuran Evaluasi

Fold	Accuracy (%)	Error Rate (%)	Sensitivity (%)	Specificity (%)	Precision (%)
Fold 1	94,27839453	5,721605465	80	94,40137812	10,95890411
Fold 2	94,67122118	5,328778822	82	94,78036176	11,91860465
Fold 3	95,14136126	4,858638743	60	95,51322751	12,39669421
Fold 4	90,7966204	9,203379602	66	91,17647059	10,28037383
Rata-Rata	93,72189934	6,278100658	72	93,9678595	11,3886442

Berdasarkan perbandingan hasil pengenalan objek klasifikasi data uji dengan algoritma PNNR seperti yang ditunjukkan pada Tabel 2 maka dapat dilihat bahwa model klasifikasi PNNR pada fold ke-3 berhasil mendapatkan false positive yang lebih kecil dibandingkan dengan fold ke 1,2 dan 4 yaitu sebesar 212 (4,24 FP/Case) sedangkan pada fold ke 1 sebesar 318 (6,36 FP/ Case), fold ke 2 sebesar 303 (6,06 FP/ Case) dan fold ke 4 sebesar 288 (5,76 FP/ Case). Tetapi algoritma PNNR pada fold ke 3 mendapatkan true positive yang lebih kecil dibandingkan dengan fold ke 1,2 dan 4 yaitu sebanyak 30 true positive sedangkan pada fold ke 1 sebesar 33, fold ke 2 sebesar 41 dan fold ke 4 sebesar 33 true positive. Hal ini menunjukkan bahwa algoritma PNNR pada fold ke 3 lebih baik dalam mengatasi false positive namun dengan tingkat sensitivity atau pengenalan true positive yang lebih rendah seperti ditunjukkan pada Tabel 3. Pada Tabel 3 dapat dilihat nilai sensitivity algoritma PNNR pada fold ke 1 sebesar 80 %, pada fold ke 2 sebesar 82 %, pada fold ke 3 sebesar 60 % dan pada fold ke 4 sebesar 66 %.

Berdasarkan pada kurva Receiver Operating Characteristics (ROC) sebagai alat ukur performansi akurasi dari permasalahan klasifikasi yang memiliki tingkat nilai pengukuran yaitu 0,50 – 0,60 disebut failure, 0,60 – 0,70 disebut poor classification, 0,70 – 0,80 disebut fair classification, 0,80 – 0,90 disebut good classification, dan 0,90 – 1,0 disebut excellent classification [15], maka pada penelitian ini dapat dikatakan nilai akurasi sebesar 0,9372 atau jika dipersentasekan menjadi 93,72% termasuk kedalam tingkat nilai excellent classification atau tingkat akurasinya sangat baik.

4. Pembahasan

Hasil klasifikasi yang diusulkan dalam penelitian ini yaitu dengan menggunakan metode klasifikasi PNNR pada *fold* ke 3 lebih baik dalam mengatasi *false positive* namun dengan tingkat *sensitivity* atau pengenalan *true positive* yang lebih rendah dari *fold* ke 1,2 dan 4 yang digunakan sebagai pembandingan. Hal ini dapat dipengaruhi oleh data *training* dan data uji yang digunakan. Data *training* dan data uji memiliki nilai *feature set* yang berbeda untuk setiap nodul. Sehingga bila digunakan data *training* dan data uji yang berbeda maka didapatkan hasil yang berbeda pula.

Pemilihan fitur pada klasifikasi PNNR sangat berpengaruh terhadap hasil klasifikasi. Dimana pada penelitian ini nilai fitur *area* rata-rata lebih besar dari nilai fitur yang lain seperti *eccentricity*, *equivalent diameter*, *major axis length*, *minor axis length* dan *perimeter*. Jadi jika nilai fitur *area* dihilangkan maka akan didapatkan hasil yang berbeda.

Hal ini dapat dibuktikan dengan melakukan penelitian algoritma PNNR dengan menggunakan data pada *Cross Validation* 3 dengan *k* sebanyak 6 dan menggunakan hanya lima fitur bentuk yaitu *eccentricity*, *equivalent diameter*, *major axis length*, *minor axis length* dan *perimeter* dari sebelumnya enam fitur bentuk dari nodul yang meliputi *area*, *eccentricity*, *equivalent diameter*, *major axis length*, *minor axis length* dan *perimeter*.

Berdasarkan hasil klasifikasi PNNR yang telah diujikan maka didapatkan total jumlah tuple seperti yang ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan Hasil Pengenalan Objek PNNR

Algoritma	Fold	k	FN	TN	TP	FP	P''	P	N''	N
PNNR (I) (6 Feature)	Fold 3	6	20	4513	30	212	242	50	4533	4725
PNNR (II) (5 Feature)	Fold 3	6	18	4449	32	276	308	50	4467	4725

Berdasarkan hasil perbandingan seperti yang ditunjukkan pada Tabel 4 didapatkan bahwa algoritma PNNR (II) menghasilkan *true positive* yang lebih besar yaitu 32 *true positive* dibandingkan dengan algoritma PNNR (I) menghasilkan *true positive* lebih sedikit yaitu 30 *true positive*. Tetapi algoritma PNNR (I) menghasilkan *false positive* yang lebih sedikit yaitu sebanyak 212 (4,24 FP/Case) dibandingkan dengan algoritma PNNR (II) menghasilkan *false positive* yang lebih banyak yaitu sebanyak 276 (5,52 FP/Case).

5. Kesimpulan

Hasil klasifikasi yang diusulkan dalam penelitian ini yaitu dengan menggunakan metode klasifikasi PNNR mendapatkan nilai akurasi sebesar 0,9372 atau jika dipersentasekan menjadi 93,72% termasuk kedalam tingkat nilai *excellent classification* atau tingkat akurasi yang sangat baik namun dengan tingkat *sensitivity* atau pengenalan *true positive* yang lebih rendah yaitu sebesar 72 %.

Untuk pengembangan penelitian juga dapat dilakukan penambahan data Citra CT scan paru agar tidak terbatas pada data publik saja, melainkan dapat juga menggunakan data rumah sakit pada suatu daerah tertentu. Karena mungkin saja data pada rumah sakit pada daerah tertentu memiliki ciri atau karakteristik tersendiri di mana tergantung pada gaya hidup dan kondisi lingkungan tertentu yang mempengaruhi kesehatan paru-paru masyarakatnya. Dalam penelitian ini menggunakan hanya fitur bentuk seperti *area*, *eccentricity*, *equivalent diameter*, *major axis length*, *minor axis length* dan *perimeter* sehingga perlu dikaji lebih lanjut dengan penggunaan selain fitur bentuk atau kombinasinya.

Ucapan Terima Kasih: Terima kasih kami ucapkan kepada seluruh teman serta dosen yang telah membantu dalam proses pengerjaan jurnal ini.

Referensi

- [1] N. Jawas, "Segmentasi Bagian Paru-paru di Citra 2D CT-Scan," E-Proceedings KNS&I STIKOM Bali, hal. 803–808, 2017.
- [2] I. W. B. Sentana, N. Jawas, dan S. A. Asri, "Automatic Solitary Lung Nodule Detection in Computed Tomography Images Slices," JPhCS, vol. 953, no. 1, hal. 12068, 2017.
- [3] S. Saien, A. H. Pilevar, dan H. A. Moghaddam, "Refinement of lung nodule candidates based on local geometric shape analysis and Laplacian of Gaussian kernels," Comput. Biol. Med., vol. 54, hal. 188–198, 2014.
- [4] T. Yanagihara dan H. Takizawa, "Pulmonary nodule detection from x-ray ct images based on region shape analysis and appearance-based clustering," Algorithms, vol. 8, no. 2, hal. 209–223, 2015.
- [5] L. Lu, Y. Tan, L. H. Schwartz, dan B. Zhao, "Hybrid detection of lung nodules on CT scan images," Med. Phys., vol. 42, no. 9, hal. 5042–5054, 2015.
- [6] I. Putra, "Klasifikasi citra menggunakan convolutional neural network (CNN) pada caltech 101." Institut Teknologi Sepuluh Nopember, 2016.
- [7] S. Hamidian, B. Sahiner, N. Petrick, dan A. Pezeshk, "3D convolutional neural network for automatic detection of lung nodules in chest CT," in Medical Imaging 2017: Computer-Aided Diagnosis, International Society for Optics and Photonics, 2017, hal. 1013409.
- [8] L. Fu, J. Ma, Y. Ren, Y. S. Han, dan J. Zhao, "Automatic detection of lung nodules: false positive reduction using convolution neural networks and handcrafted features," in Medical Imaging 2017: Computer-Aided Diagnosis, International Society for Optics and Photonics, 2017, hal. 101340A.
- [9] A. Teramoto, H. Fujita, O. Yamamuro, dan T. Tamaki, "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique," Med. Phys., vol. 43, no. 6Part1, hal. 2821–2827, 2016.
- [10] P.-P. Ypsilantis dan G. Montana, "Recurrent convolutional networks for pulmonary nodule detection in CT imaging," arXiv Prepr. arXiv1609.09143, 2016.
- [11] Y. Zeng, Y. Yang, dan L. Zhao, "Pseudo nearest neighbor rule for pattern classification," Expert Syst. Appl., vol. 36, no. 2, hal. 3587–3595, 2009.
- [12] H. Azis, P. Purnawansyah, F. Fattah, dan I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," Ilk. J. Ilm., vol. 12, no. 2, hal. 81–86, 2020.
- [13] L. Mardiana, D. Kusnandar, dan N. Satyahadewi, "Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak," Bimaster Bul. Ilm. Mat. Stat. dan Ter., vol. 11, no. 1, 2022.
- [14] D. Suyanto, "Data Mining untuk klasifikasi dan klasterisasi data," Bandung Inform. Bandung, 2017.
- [15] T. Rosandy, "Perbandingan Metode Naive Bayes Classifier Dengan Metode Decision Tree (C4. 5) Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus: KSPPS/BMT Al-Fadhila)," J. Teknol. Inf. Magister, vol. 2, no. 01, hal. 52–62, 2016.