Repeater : Publikasi Teknik Informatika dan Jaringan
Vol. 2 No. 3 Juli 2024
e-ISSN : 3046-7276, p-ISSN : 3046-7284, Hal.12- 25
DOI: https://doi.org/10.62951/repeater.v2i3.77

Implementasi Algoritma K-Means dan Knearest Neighbors (KNN) Untuk Identifikasi Penyakit Tuberkulosis Pada Paru-Paru

Rachmadhany Iman [1], Basuki Rahmat [2], Achmad Junaidi [3]
[1,2,3] Universitas Pembangunan Nasional Veteran Jawa Timur

Jl. Rungkut Madya No. 1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294
Korespondensi penulis: rachmadhanymanman@gmail.com

Abstract. *In Indonesia, tuberculosis is ranked third in terms of prevalence among countries with the highest tuberculosis burden. Radiological examination, such as X-rays or X-rays, is a method generally used to detect tuberculosis. Chest X-ray examination is one method used to detect tuberculosis. To achieve these goals, the research will combine two powerful data processing techniques. First, the K-Means algorithm will be used to group x-ray image data based on similar characteristics, making it easier to identify typical patterns from images infected with tuberculosis. The research results show the highest accuracy of 93% using data division with a ratio of 80 : 20 with parameter K = 1. These results show that the combined model of the two algorithms can be applied to identify tuberculosis in the lungs.*

Keywords: *Lungs, Tuberculosis, X-Ray Image, K-Means, KNN*

Abstrak. Di Indonesia, tuberkulosis menduduki peringkat ketiga dalam hal prevalensi di antara negara-negara dengan beban tuberkulosis tertinggi. Pemeriksaan radiologi, seperti foto sinar-X atau rontgen, adalah metode yang umumnya digunakan untuk mendeteksi tuberkulosis. Pemeriksaan sinar-X dada merupakan salah satu cara yang digunakan dalam mendeteksi tuberkulosis. Untuk mencapai tujuan tersebut, penelitian akan menggabungkan dua teknik pemrosesan data yang kuat. Pertama, algoritma K-Means akan digunakan untuk mengelompokkan data citra x-ray berdasarkan karakteristik yang serupa, sehingga memudahkan proses identifikasi pola khas dari citra yang terinfeksi tuberkulosis. Hasil penelitian menunjukkan akurasi tertinggi sebesar 93% menggunakan pembagian data dengan rasio 80 : 20 dengan parameter K = 1. Hasil tersebut menunjukkan model gabungan dari dua algoritma tersebut dapat diterapkan untuk identifikasi penyakit tuberkulosis pada paru-paru.

Kata kunci: Paru-paru, Tuberkulosis, Citra X-Ray, K-Means, KNN

LATAR BELAKANG

Kesehatan adalah aset berharga yang tidak bisa dibeli oleh siapa pun, menjadikannya sangat penting bagi setiap orang. Salah satu organ vital dalam tubuh manusia yang memiliki pengaruh besar terhadap kesehatan adalah paru-paru (Kusuma & Chairani 2015). Fungsi organ paru-paru ini yaitu sebagai alat yang bekerja untuk menampung atau memasok oksigen dan menyaring udara yang masuk ke dalam tubuh sehingga dapat mengeluarkan udara kotor sehingga keseluruhan tubuh manusia dapat menerima oksigen dan pada akhirnya seluruh organ tubuh manusia dapat berfungsi dengan baik (Rosmita Ritonga & Dedi Irawan 2023). Kesehatan akan terganggu apabila paru-paru terserang oleh penyakit sehingga paru-paru tidak dapat melakukan fungsinya dengan baik. Salah satu penyakit yang dapat menyerang paru-paru adalah tuberkulosis.

Tuberkulosis adalah sebuah penyakit menular yang disebabkan oleh bakteri Microbacterium tuberculosis. Tuberkulosis dapat menyerang berbagai bagian tubuh, tetapi organ yang paling sering terkena adalah paru-paru. Penularan Tuberkulosis

Received Mei 30, 2024; Accepted Juni 04, 2024; Published Juli 31, 2024
* Rachmadhany Iman, rachmadhanymanman@gmail.com

terutama terjadi melalui partikel udara yang dilepaskan saat seseorang yang terinfeksi batuk atau bersin (Abdullah, Rahmi, & Yunizar 2022). Menurut Global TB Report tahun 2022, tuberkulosis di Indonesia menduduki peringkat ketiga dalam hal prevalensi di antara negara-negara dengan beban Tuberkulosis tertinggi setelah India dan Cina, dengan total kasus tahunan mencapai 824 ribu dan menyebabkan sekitar 93 ribu kematian per tahun, atau setara dengan 11 kematian per jam. jumlah kasus Tuberkulosis terbanyak ditemukan pada kelompok usia produktif, terutama antara 25 hingga 34 tahun secara global. Namun, di Indonesia, kelompok usia dengan kasus tuberkulosis terbanyak adalah mereka yang berusia 45 hingga 54 tahun, menurut data terbaru dari Kementerian Kesehatan Republik Indonesia pada tahun 2023.

Pemeriksaan radiologi, seperti foto sinar-X atau rontgen, adalah metode yang umumnya digunakan untuk mendeteksi tuberkulosis, sering kali menghasilkan gambar yang menunjukkan perbedaan langsung pada kondisi paru-paru yang terinfeksi (Rahmadewi & Kurnia 2016). Citra *x-ray* ini memberikan gambaran tentang kondisi jantung, dada, paru-paru, dan saluran pernafasan. Salah satu ciri-ciri yang dapat dilihat dari citra *x-ray*, area dengan warna abu-abu terang seringkali menandakan infeksi oleh virus, bakteri, jamur, atau parasit lainnya. Ini dapat menjadi indikasi bagi dokter untuk mencurigai keberadaan penyakit tertentu pada pasien, seperti tuberkulosis (Maysanjaya 2020). Dalam konteks ini, kecerdasan buatan dan pembelajaran mesin dapat memberikan bantuan kepada dokter dalam mengidentifikasi tuberkulosis dengan cepat dan efektif. Di sektor kesehatan, teknologi pembelajaran mesin digunakan untuk memprediksi penyakit. Kecepatan dan akurasi prediksi penyakit menjadi kunci dalam penanganan penyakit oleh dokter radiologi yang berkualifikasi (Mustopa dkk., 2022).

Dengan kemajuan dalam teknologi medis, algoritma kecerdasan buatan seperti K-Means dan K-Nearest Neighbors (KNN) semakin sering digunakan untuk meningkatkan proses diagnostik dan mengidentifikasi penyakit secara lebih akurat. Dalam konteks ini, algoritma K-Means membantu dalam klasifikasi pola atau fitur dalam citra *x-ray*, sementara KNN digunakan untuk pendekatan klasifikasi yang lebih terfokus. Kombinasi kedua metode ini dapat meningkatkan akurasi dan kecepatan dalam mendiagnosis tuberkulosis dari gambar *x-ray* paru-paru. Penelitian ini bertujuan untuk berkontribusi pada pengembangan sistem diagnostik yang lebih canggih dan responsif. Diharapkan, hasil dari penelitian ini tidak hanya akan meningkatkan efektivitas diagnosis tuberkulisus dini tetapi juga membantu para profesional kesehatan dalam merencanakan

strategi pengobatan yang lebih efektif. Selain itu, penggunaan algoritma kecerdasan buatan dalam layanan kesehatan diharapkan dapat membuka jalan bagi pengembangan teknologi baru yang mendukung peningkatan layanan kesehatan.

Dari uraian latar belakang yang telah dijelaskan, penulis bertujuan menggunakan metode K-Means untuk klastering data, yang kemudian akan diklasifikasikan menggunakan KNN. Tujuannya adalah untuk membedakan antara citra rontgen normal dan yang menunjukkan tanda-tanda tuberkulosis. Oleh karena itu, penulis mengusulkan judul penelitian "Implementasi Algoritma K-Means dan K-Nearest Neighbor (KNN) Untuk Identifikasi Penyakit Tuberkulosis". Hasil utama yang diharapkan dari penelitian ini adalah menentukan tingkat akurasi dari penggunaan kedua algoritma tersebut dalam mengidentifikasi tuberkulosis. Penelitian ini diharapkan tidak hanya akan validasi metodologi yang diproposikan tetapi juga memperkaya literatur dengan temuan tentang efektivitas teknik-teknik ini dalam konteks medis, khususnya dalam diagnosa penyakit paru-paru seperti tuberkulosis.

KAJIAN TEORITIS

Tuberkulosis

Tuberkulosis adalah penyakit menular yang terutama disebabkan oleh bakteri Mycobacterium tuberculosis. Penyakit ini umumnya diawali saat bakteri tersebut masuk ke dalam tubuh melalui udara yang dihirup ke dalam paru-paru. Setelah berada di paru-paru, bakteri ini dapat menyebar ke berbagai bagian tubuh lainnya menggunakan beberapa jalur, termasuk sistem peredaran darah, sistem limfatik, saluran pernapasan seperti bronkus, atau melalui penyebaran langsung ke organ atau jaringan lain (Nurmalinda Noviansyah, Nur Eni Lestari, & Eka Rokhmiati 2021). Gejala utama yang sering muncul pada orang yang terinfeksi termasuk batuk persisten selama lebih dari dua minggu, batuk dengan dahak yang terkadang berdarah, sesak nafas, penurunan nafsu makan, kelemahan umum, malaise, demam berkepanjangan lebih dari satu bulan, dan keringat malam tanpa aktivitas fisik (Kementrian Kesehatan RI, 2016).

Pengolahan Citra

Pengolahan citra digital merupakan bidang ilmu yang berfokus pada berbagai aspek peningkatan dan modifikasi citra, seperti peningkatan kontras, pembaruan warna, pemulihan gambar dari degradasi, serta penyimpanan dan transmisi data yang efisien. Proses ini melibatkan pengubahan citra menjadi matriks numerik yang kemudian dapat

dimanipulasi untuk menghasilkan output yang lebih informatif dan mudah diinterpretasikan oleh manusia (Munantri, Sofyan, & Florestiyanto 2020)(Munantri et al. 2020) .

Machine Learning

Machine learning (ML) atau pembelajaran mesin adalah bidang studi yang memberikan kemampuan pada komputer untuk belajar suatu hal tanpa harus diatur secara spesifik. Pembelajaran mesin digunakan untuk mengajari mesin bagaimana cara mengelola data secara lebih efisien. Pembelajaran mesin bekerja dalam meinginterpretasikan informasi yang dihasilkan dari data yang besar dimana manusia pada umumnya tidak bisa langsung melakukan hal tersebut (Agung Mujiono dkk., 2024).

K-Means

K-Means adalah teknik pengelompokan data non-hierarki yang membagi data menjadi kelompok-kelompok berdasarkan kualitas atribut numerik. Algoritma ini menggabungkan *partitioning clustering* untuk memisahkan data ke dalam k sub-wilayah yang terpisah. K-Means sangat efektif dalam mengumpulkan data besar dan *outlier* dengan cepat. Dalam penggunaannya, setiap data harus termasuk ke dalam *cluster* tertentu dan dapat berpindah ke *cluster* lainnya pada tahapan berikutnya.

Penggunaan algoritma K-Means bergantung pada data yang tersedia dan tujuan yang ingin dicapai. Oleh karena itu, algoritma ini digunakan untuk membuat prinsip-prinsip sebagai berikut: jumlah *cluster* harus diinputkan, dan hanya atribut numerik yang digunakan. Algoritma K-Means mengambil sebagian dari bagian populasi sebagai komunitas *cluster* mendasarinya dan menghitung posisi pusat *cluster* secara ulang hingga semua bagian data diatur ke dalam setiap *cluster*.

1. Segmentasi K-Means

Dalam segmentasi menggunakan K-Means, langkah pertama adalah menentukan jumlah *cluster* pada citra yang telah diproses sebelumnya dan menghitung *centroid* secara acak. Selanjutnya, hitung jarak setiap piksel ke *centroid* dan kelompokkan piksel-piksel berdasarkan jarak terdekat. Setelah piksel-piksel dikelompokkan berdasarkan jarak terdekat mereka, pusat *cluster* dihitung ulang sebagai pusat massa baru dengan menghitung nilai rata-rata piksel per *cluster* sebagai *centroid* baru dan mengelompokkan kembali piksel sesuai dengan *centroid* tersebut. Jika masih ada piksel yang berpindah *cluster*, maka *centroid* dihitung ulang. Namun, jika tidak ada piksel yang berpindah *cluster*, proses pengelompokan selesai (Febrinanto dkk., 2018).

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah metode klasifikasi yang menentukan kelas suatu objek berdasarkan data pelatihan yang paling dekat dengannya. Data pelatihan diproyeksikan ke dalam ruang multidimensi, di mana setiap dimensi mewakili karakteristik data (Yulianto, Riadi, & Umar 2023). Algoritma KNN sangat sederhana, bekerja dengan menghitung jarak terpendek dari objek *query* ke sampel pelatihan untuk menentukan jumlah k tetangga terdekatnya. Setelah mengidentifikasi k tetangga terdekat, mayoritas dari k tetangga tersebut digunakan untuk memprediksi kelas objek *query*. Tujuan algoritma KNN adalah untuk mengklasifikasikan objek berdasarkan atribut dan sampel pelatihan. Algoritma ini tidak menggunakan model khusus untuk pencocokan, melainkan hanya bergantung pada memori (Yulianto, Riadi, & Umar 2023).

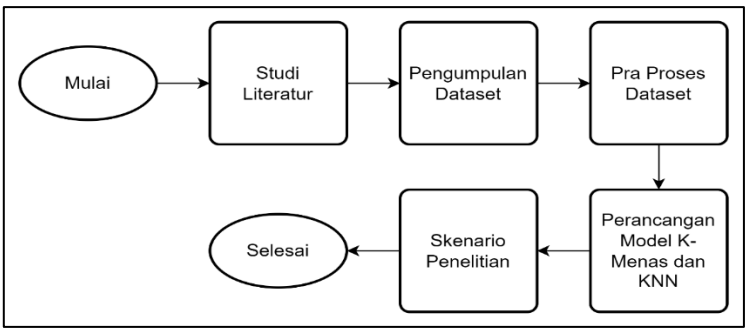
2. *Euclidean Distance*

Euclidean Distance digunakan untuk proses klasifikasi atau identifikasi dengan menghitung jarak antara suatu vektor *training* dan vektor *testing* untuk mengukur jarak antara data dan titik fokus *cluster* digunakan jarak *Euclidean*, maka pada titik tersebut akan diperoleh matriks jarak sebagai berikut :

$$d(x, y) = \sqrt{\sum(xi - yi)^2} \quad (1)$$

Euclidean distance adalah metrik jarak yang umum digunakan dalam implementasi algoritma K-Means dan K-Nearest Neighbors (KNN). Dalam kedua kasus, *Euclidean distance* digunakan untuk mengukur jarak antara titik data dalam ruang fitur.

METODE PENELITIAN



Gambar 1. Alur Penelitian

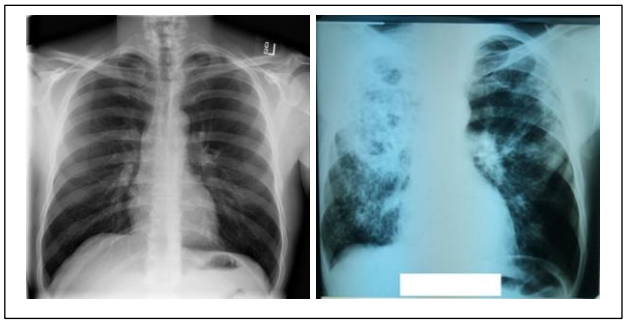
Studi Literatur

Literatur yang telah dikaji pada bagian penelitian terkait pada penelitian ini akan menjadi dasar teori atas studi dan penelitian yang dilakukan. Agar dapat menjadi landasan teori atas dilaksanakannya penelitian ini. Konsep dan dasar teori yang digunakan dapat

meliputi sumber buku, jurnal dari penelitian yang telah dilakukan sebelumnya yang relevan. Referensi dari studi literatur yang telah dipelajari dan digunakan pada penelitian ini dilampirkan pada daftar referensi di bagian akhir.

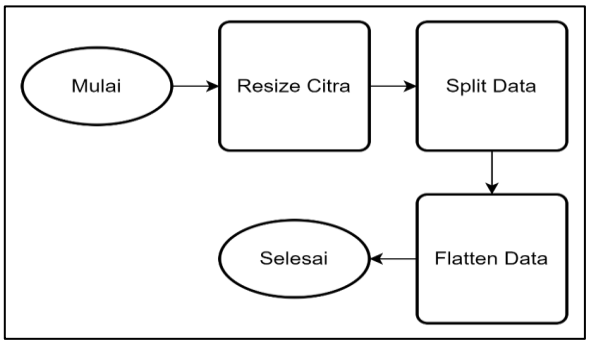
Pengumpulan Dataset

Pada penelitian ini menggunakan data yang didapat dari Kaggle dengan judul "Tuberculosis (TB) Chest X-ray Database" dengan pemilik dataset bernama Tawsifur Rahman, Dr. Muhammad Chowdhury, Amith Khandakar. Dataset ini berisi 4.200 gambar berformat file png yang terbagi menjadi 2 kelas yaitu paru-paru normal dan paru-paru terinfeksi tuberkulosis. Dalam dataset ini ditemukan ukuran citra yang berbeda-beda sehingga memerlukan tahapan pra proses data seperti yang telah disebutkan pada tahap penelitian di atas. Untuk menunjang penelitian, dataset yang digunakan dari website Kaggle yaitu sebanyak 1400 data dengan sebaran data pada kelas normal berisi 700 data citra dan tuberkulosis 700 data citra yang dapat dilihat pada gambar 2.



Gambar 2. Data Citra X-Ray

Pra Proses Dataset

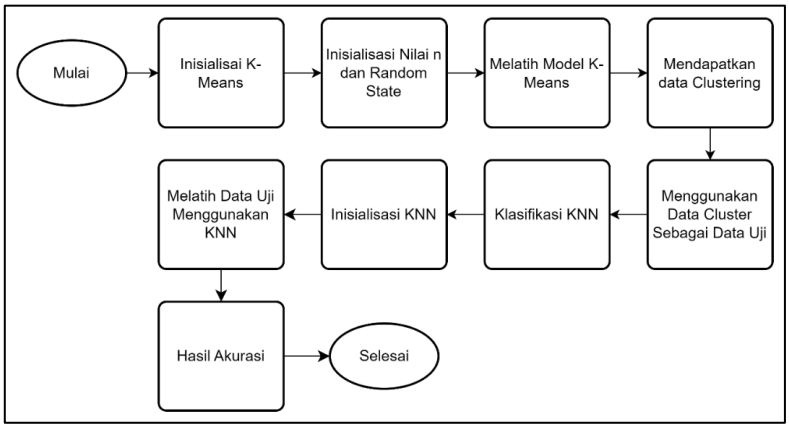


Gambar 3. Pra Proses Dataset

Pra Proses Dataset digunakan untuk menyamakan data citra yang bertujuan untuk menunjang penelitian khususnya pada tahap-tahap selanjutnya. Pada gambar 3, tahapan ini terdiri dari tiga sub tahapan yang dilakukan pada pra proses dataset, antara lain *Resize* Citra, *Split* Data, dan *Flatten* Data. Pada dataset yang telah ditentukan, data

yang berada didalamnya mempunyai ukuran citra 512 x 512 sehingga diperlukan proses *resizing* pada citra menjadi ukuran yang sama yaitu dengan ukuran 64 x 64 untuk mengurangi kompleksitas pada citra sehingga dapat mempercepat waktu komputasi. Setelah citra sudah memiliki bentuk dan tipe yang sama, selajutnya data citra dibagi menjadi 2 bagian yaitu untuk data latih dan data uji. Setelah pembagian data, data kemudian dilakukan proses *flatten* untuk penyesuaian data agar dapat diinputkan pada tahap-tahap selanjutnya dengan menggunakan metode *reshape* yang akan diterapkan pada setiap citra dalam data latih dan data uji yang awalnya berbentuk matriks 2 dimensi menjadi vektor satu dimensi.

Perancangan Model K-Means dan KNN



Gambar 4. Perancangan Model K-Means dan KNN

Pada gambar 4, dapat diilustrasikan sebagai suatu rangkaian langkah yang menggabungkan algoritma K-Means sebagai *clustering* dan K-Nearest Neighbors (KNN) untuk melakukan klasifikasi data uji. Langkah pertama dimulai dari "Mulai" dan melanjutkan ke tahap inisialisasi K-Means sebagai metode awal untuk memulai *clustering*, dimana algoritma K-Means *clustering* di inisialisasi dengan menggunakan nilai n dan *random state*. Pada tahap melatih model K-Means, model dilatih menggunakan algoritma K-Means untuk mengelompokkan data menjadi kelompok berdasarkan fitur-fiturnya yang telah dilakukan pada saat praproses sebelum data inisialisasi kemudian data diperoleh dan dikelompokkan berdasarkan kelompok yang telah terbentuk. Selanjutnya, data yang telah dikelompokkan ini digunakan sebagai fitur yang akan diklasifikasi dengan data uji. Secara bersamaan, ada proses lain yang dimulai dengan tahap inisialisasi KNN dengan mengatur nilai K yang telah ditentukan dalam penelitian ini. Pada tahap melatih data uji menggunakan KNN, klasifikasi dilatih dengan menggunakan data uji yang diperoleh dari proses *clustering* sebelumnya. Evaluasi kinerja

klasifikasi dan hasil akurasi diperoleh pada tahap hasil akurasi Keseluruhan proses berakhir pada tahap selesai.

Skenario Pengujian

Skenario penelitian mengacu pada rencana atau rancangan diambil oleh peneliti untuk mengumpulkan data, menganalisis informasi, dan menyusun kesimpulan berdasarkan hasil penelitian. Skenario penelitian yang dilakukan yaitu mengimplementasikan algortima K-Means dan K-Nearest Neighbor (KNN) untuk mengetahui hasil akurasi dengan menggunakan model gabungan dari kedua algoritma tersebut. Skenario penelitian pada penelitian ini dilakukan dengan membedakan parameter nilai K yang dapat dilihat pada tabel 1.

Tabel 1. Skenario Penelitian

Skenario	Model	Pembagian Data	Nilai K
1	K-means dan KNN	80 : 20	1
2			2
3			3

HASIL DAN PEMBAHASAN

Pengumpulan Dataset

Data yang digunakan pada penelitian ini menggunakan data yang berasal dari Kaggle. Dataset tersebut memiliki total gambar sebanyak 1400 yang terdiri dari 700 data gambar paru-paru normal dan 700 data gambar paru-paru tuberkulosis. Data gambar tersebut kemudian dikumpulkan menjadi satu folder yang telah dibedakan perkelasnya lalu diunggah ke Google Drive untuk memudahkan proses-proses yang akan dilakukan selanjutnya.

Pra Proses Dataset

Setelah berhasil mendapatkan dataset mentah, kemudian dilakukan pra proses dataset dengan tujuan mempersiapkan dataset agar siap untuk masuk ke tahap pelatihan. Tahap pra proses dimulai dengan merubah ukuran citra dari 512 x 512 menjadi 64 x 64 untuk menyederhanakan dimensi citra, meningkatkan efisiensi perhitungan pada tahap analisis atau pelatihan model. Setelah dilakukan proses *resizing*, data kemudian dibagi menjadi ke dalam data latih dan data uji yang akan digunakan pada tahapan selanjutnya. Data dibagi menggunakan rasio 80% untuk data latih dengan total data citra berjumlah 1120 dan 20% untuk data uji dengan total data citra berjumlah 280. Setelah itu, data

dilakukan proses *flatten* dengan mengubah bentuk menjadi satu dimensi. Dengan menerapkan metode `*reshape*`, dimensi saluran warna dan dimensi citra yang sebelumnya tiga dimensi diubah menjadi satu dimensi, yang diwakili oleh -1.

***Clustering* K-Means**

Tabel 2. Nilai *Silhouette Score*

Silhouette Score	Nilai
Data Latih	0.16254882649190985
Data Uji	0.15530865520431072

Skor Silhouette untuk data latih memiliki nilai sekitar 0.16. Ini menandakan bahwa pengelompokan menggunakan algoritma K-Means pada data latih menghasilkan pembagian kluster yang cukup baik, tetapi mungkin terdapat beberapa objek yang tidak terlalu jelas atau ambigu terkait kluster mereka. Skor Silhouette untuk data uji memiliki nilai sekitar 0.15. Hal ini menunjukkan bahwa algoritma K-Means memberikan hasil pengelompokan yang serupa pada data uji seperti pada data latih. Meskipun skornya tidak terlalu tinggi, masih menunjukkan bahwa pengelompokan pada data uji juga memberikan kluster yang relatif baik, walaupun mungkin ada beberapa tingkat ketidakpastian dalam penempatan beberapa objek ke dalam kluster.

Klasifikasi KNN

Tahap klasifikasi KNN dilakukan dengan melakukan klasifikasi menggunakan data latih dan data uji yang telah dibuat pada tahapan sebelumnya. Proses klasifikasi ini dilakukan dengan menggunakan skenario penelitian yang ditetapkan. Sebanyak 1120 data latih dan 280 data uji digunakan untuk mengukur sejauh mana hasil akurasi dari model tersebut.

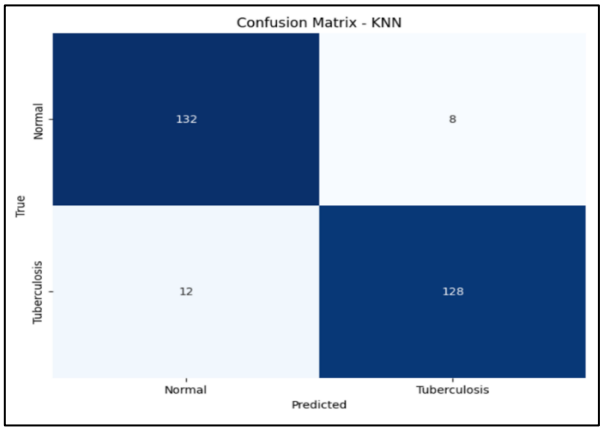
Tabel 3. Akurasi Klasifikasi KNN

Skenario	Model	Pembagian Data	Nilai K	Akurasi
1	K-means dan KNN	80 : 20	1	0.93
2			2	0.90
3			3	0.91

Analisa Hasil Pengujian Skenario Penelitian

Analisa hasil pengujian meliputi evaluasi metrik performa yang terdiri dari hasil *confusion matrix* dan *classification report* yang meliputi nilai akurasi, presisi, *recall*, dan *F1-score* dari skenario penelitian yang telah ditentukan. Laporan klasifikasi yang diberikan menggambarkan kinerja dari model pada dua kelas berbeda, yaitu normal dan tuberkulosis.

1. Skenario Penelitian 1



Gambar 5. *Confusion matrix* Skenario Penelitian 1

Skenario Penelitian 1 berhasil mengklasifikasikan 260 dari 280 gambar yang diuji, sedangkan 20 gambar lainnya tidak berhasil diklasifikasikan dengan benar. Hasil ini menunjukkan performa model yang sudah dilatih dalam mengenali objek baru dari setiap kelas dengan tepat serta bagaimana penyebaran hasil prediksi terhadap seluruh data uji, yaitu sebanyak 280 data.

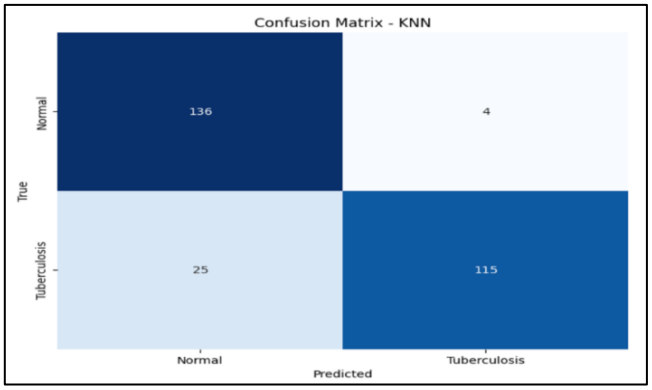
Classification Report – KNN:

	precision	recall	f1-score	support
Normal	0.92	0.94	0.93	140
Tuberculosis	0.94	0.91	0.93	140
accuracy			0.93	280
macro avg	0.93	0.93	0.93	280
weighted avg	0.93	0.93	0.93	280

Gambar 6. *Classification report* Skenario Penelitian 1

Pada Skenario Penelitian 1, kelas "Normal" model memiliki *precision* 0.92, *recall* 0.94, dan *F1-score* 0.93. Untuk kelas "Tuberkulosis", *precision* adalah 0.94, *recall* 0.91, dan *F1-score* 0.93. Akurasi keseluruhan model adalah 0.93. *Macro average* untuk *precision*, *recall*, dan *F1-score* adalah 0.93, 0.93, dan 0.93, sedangkan *weighted average* masing-masing adalah 0.93, 0.93, dan 0.93.

2. Skenario Penelitian 2



Gambar 7. *Confusion matrix* Skenario Penelitian 2

Skenario Penelitian 2 berhasil mengklasifikasikan 251 dari 280 gambar yang diuji, sedangkan 29 gambar lainnya tidak berhasil diklasifikasikan dengan benar. Hasil ini menunjukkan performa model yang sudah dilatih dalam mengenali objek baru dari setiap kelas dengan tepat serta bagaimana penyebaran hasil prediksi terhadap seluruh data uji, yaitu sebanyak 280 data.

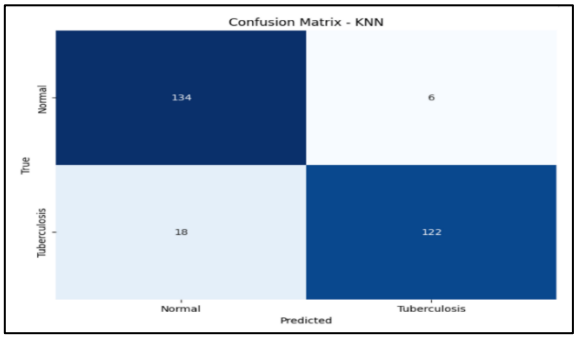
Classification Report – KNN:

	precision	recall	f1-score	support
Normal	0.84	0.97	0.90	140
Tuberculosis	0.97	0.82	0.89	140
accuracy			0.90	280
macro avg	0.91	0.90	0.90	280
weighted avg	0.91	0.90	0.90	280

Gambar 8. *Classification report* Skenario Penelitian 2

Pada Skenario Penelitian 2, kelas "Normal" model memiliki *precision* 0.84, *recall* 0.97, dan *F1-score* 0.90. Untuk kelas "Tuberkulosis", *precision* adalah 0.97, *recall* 0.82, dan *F1-score* 0.89. Akurasi keseluruhan model adalah 0.90. *Macro average* untuk *precision*, *recall*, dan *F1-score* adalah 0.91, 0.90, dan 0.90, sedangkan *weighted average* masing-masing adalah 0.91, 0.90, dan 0.90.

3. Skenario Penelitian 3



Gambar 9. *Confusion matrix* Skenario Penelitian 3

Skenario Penelitian 3, berhasil mengklasifikasikan 256 dari 280 gambar yang diuji, sedangkan 24 gambar lainnya tidak berhasil diklasifikasikan dengan benar. Hasil ini menunjukkan performa model yang sudah dilatih dalam mengenali objek baru dari setiap kelas dengan tepat serta bagaimana penyebaran hasil prediksi terhadap seluruh data uji, yaitu sebanyak 280 data.

Classification Report - KNN:

	precision	recall	f1-score	support
Normal	0.88	0.96	0.92	140
Tuberculosis	0.95	0.87	0.91	140
accuracy			0.91	280
macro avg	0.92	0.91	0.91	280
weighted avg	0.92	0.91	0.91	280

Gambar 10. *Classification report* Skenario Penelitian 3

Pada Skenario Penelitian 3, kelas "Normal" model memiliki *precision* 0.88, *recall* 0.96, dan *F1-score* 0.92. Untuk kelas "Tuberkulosis", *precision* adalah 0.95, *recall* 0.87, dan *F1-score* 0.91. Akurasi keseluruhan model adalah 0.91. *Macro average* untuk *precision*, *recall*, dan *F1-score* adalah 0.92, 0.91, dan 0.91, sedangkan *weighted average* masing-masing adalah 0.92, 0.91, dan 0.91.

KESIMPULAN DAN SARAN

Dari hasil identifikasi yang dilakukan terhadap penyakit tuberkulosis pada paru-paru dari data citra *x-ray*, didapatkan bahwa penggunaan metode K-Means sebagai *clustering* dan KNN sebagai klasifikasi menunjukkan hasil yang baik. Penggunaan pembagian data dengan rasio 80 : 20 dengan nilai K = 1 pada skenario penelitian 1 mendapatkan nilai akurasi terbaik yaitu sebesar 93%. Nilai *classification report* yang didapatkan juga menunjukkan hasil yang baik yaitu kelas "Normal" *precision* sebesar 92%, *recall* 94%, dan *F1-score* 93%. Untuk kelas "Tuberkulosis", *precision* sebesar 94%, *recall* 91%, dan *F1-score* 93%. Hal tersebut menunjukkan bahwa model tersebut dapat diterapkan dengan baik dan efektif untuk melakukan identifikasi penyakit tuberkulosis pada paru-paru.

Penelitian selanjutnya disarankan untuk menggunakan dataset yang persebaran datanya seimbang pada setiap kelasnya sehingga tidak terjadi perbedaan jumlah data pada tiap kelasnya yang dapat mempengaruhi pada hasil akurasi sistem. Penambahan jumlah dataset juga dapat dilakukan agar model dapat mempunyai fitur yang kaya dan dapat

menambah nilai akurasi dari hasil keseluruhan model. Penelitian lain dapat dilakukan dengan menggunakan jenis algoritma gabungan lainnya seperti CNN-SVM untuk membandingkan nilai akurasi dan kinerja modelnya, sehingga dapat menghasilkan analisis identifikasi yang lebih akurat.

DAFTAR REFERENSI

Abdullah, Dahlan, Meutia Rahmi, and Zara Yunizar. 2022. "Implementasi Sistem Pakar Diagnosa Awal Penyakit Tuberculosis Paru Menggunakan Fuzzy Tsukamoto." *VARIASI : Majalah Ilmiah Universitas Almuslim* 13(3):153–57. doi: 10.51179/vrs.v13i3.860.

Agung Mujiono, Alfinas, Kartini Kartini, and Eva Yulia Puspaningrum. 2024. "Implementasi Model Hybrid Cnn-Svm Pada Klasifikasi Kondisi Kesegaran Daging Ayam." *JATI (Jurnal Mahasiswa Teknik Informatika)* 8(1):756–63. doi: 10.36040/jati.v8i1.8855.

Febrinanto, Falih Gozi, Candra Dewi, Anang Tri Wiratno, Balai Penelitian, Tanaman Jeruk, Buah Subtropika, and Badan Litbang Pertanian. 2018. "Implementasi Algoritme K-Means Sebagai Metode Segmentasi Citra Dalam Identifikasi Penyakit Daun Jeruk." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2(11):5375–83.

Kusuma, Diki Andita, and Chairani Chairani. 2015. "Rancang Bangun Sistem Pakar Pendiagnosa Penyakit Paru-Paru Menggunakan Metode Case Based Reasoning." *Jurnal Informatika,Telekomunikasi Dan Elektronika* 6(2):57–62. doi: 10.20895/infotel.v6i2.74.

Maysanjaya, I. Md. Dendi. 2020. "Klasifikasi Pneumonia Pada Citra X-Rays Paru-Paru Dengan Convolutional Neural Network." *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* 9(2):190–95. doi: 10.22146/jnteti.v9i2.66.

Munantri, Nadzir Zaid, Herry Sofyan, and Mangaras Yanu Florestiyanto. 2020. "Aplikasi Pengolahan Citra Digital Untuk Identifikasi Umur Pohon." *Telematika* 16(2):97. doi: 10.31315/telematika.v16i2.3183.

Mustopa, Ali, Hendri Mahmud Nawawi, Sarifah Agustiani, Siti Khotimatul Wildah, Sistem Informasi, Kampus Kota, Universitas Bina Sarana, Teknik Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Teknologi Komputer, Universitas Bina, Sarana Informatika, Kota Jakarta Pusat, Mechine Learning, and Ekstraksi Fitur. 2022. "Ekstraksi Fitur Dengan Classifer Random Forest Untuk Memprediksi Covid 19 Berdasarkan Hasil Rontgen Thorax Feature Extraction with Random Forest Classifier to Predict Covid 19 Based on Results Thorax X - Ray." *Jurnal Sistem Informasi* 11:515–25.

Nurmalinda Noviansyah, Nur Eni Lestari, and Eka Rokhmiati. 2021. "Hubungan Perilaku Orang Tua Dan Faktor Lingkungan Dengan Kejadian Tuberkulosis Paru Pada Anak Di Desa Bangunjaya Tahun 2020." *Indonesian Scholar Journal of Nursing*

and Midwifery Science (ISJNMS) 1(04):149–56. doi: 10.54402/isjnms.v1i04.72.

Rahmadewi, Reni, and Rahmadi Kurnia. 2016. "Klasifikasi Penyakit Paru Berdasarkan Citra Rontgen Dengan Metoda Segmentasi Sobel." *Jurnal Nasional Teknik Elektro* 5(1):7. doi: 10.25077/jnte.v5n1.174.2016.

Rosmita Ritonga, Eli, and Muhammad Dedi Irawan. 2023. "Sistem Pakar Diagnosa Penyakit Paru-Paru." *Journal Of Computer Engineering, System And Science* 2(1):193–200.

Yulianto, Rahmat Ardila Dwi, Imam Riadi, and Rusydi Umar. 2023. "Perancangan Klasifikasi Pasien Stroke Dengan Metode K-Nearest Neighbor." *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab* 8(2):262–68. doi: 10.36341/rabit.v8i2.3454.