JITE (Journal of Informatics and Telecommunication Engineering)

Available online <http://ojs.uma.ac.id/index.php/jite> DOI : 10.31289/jite.v5i2.6216

Received: 30 November 2021

Accepted: 24 January 2022

Published: 26 January 2022

Identification of Pneumonia using The K-Nearest Neighbors Method using HOG Feature Extraction

Nurul Khairina¹⁾, Theofil Tri Saputra Sibarani¹⁾, Rizki Muliono¹⁾,
Zulfikar Sembiring¹⁾, Muhathir¹⁾*

¹⁾Prodi Teknik Informatika, Fakultas Teknik, Universitas Medan Area, Indonesia

*Corresponding Email: muhathir@staff.uma.ac.id

Abstrak

Pneumonia merupakan penyakit paru-paru basah. Pneumonia pada umumnya disebabkan oleh virus, bakteri atau jamur. Tak jarang Pnumonia dapat menyebabkan kematian. Metode K-Nearest Neighbors adalah metode klasifikasi yang menggunakan nilai mayoritas dari kategori nilai k yang terdekat. Pada saat ini masyarakat tidak terlalu khawatir dengan pneumonia karena penyakit pneumonia ini memiliki gejala seperti batuk biasa. Namun, informasi yang cepat dan akurat dari pakar kesehatan juga sangat diperlukan agar gejala pneumonia dapat dikenali lebih dini dan cara menanganinya juga dapat lakukan lebih cepat. Pada penelitian ini, peneliti akan mendiagnosis pneumonia untuk memperoleh informasi dengan cepat tentang gejala pneumonia. Informasi ini akan mengadopsi pengetahuan manusia ke komputer yang dirancang untuk menyelesaikan permasalahan identifikasi penyakit pneumonia. Pada penelitian ini, metode K-Nearest Neighbors akan dikombinasikan dengan Fitur Ekstraksi HOG untuk mengidentifikasi pneumonia yang lebih akurat. Klasifikasi KNN yang digunakan adalah Fine KNN, Cosine KNN, dan Cubic KNN. Dimana akan dilihat berapa nilai akurasi, presisi, recall, dan f1-score. Hasil penelitian menunjukkan klasifikasi dapat berjalan dengan baik pada metode Fine KKN, Cosine KNN, maupun Cubic KNN. Fine KNN menghasilkan tingkat akurasi sebesar 80,67, Cosine KNN menghasilkan tingkat akurasi sebesar 84,93333, dan Cubic KNN menghasilkan tingkat akurasi sebesar 83,13333. Nilai presisi, recall dan f1-score dari Fine KNN secara berurutan adalah 0.794842, 0.923706, dan 0.854442. Nilai presisi, recall dan f1-score dari Cosine KNN secara berurutan adalah 0.803048, 0.954039, dan 0.872056. Nilai presisi, recall dan f1-score dari Cubic KNN secara berurutan adalah 0.73388, 0.964561, dan 0.833555. Dari hasil pengujian, identifikasi positif dan negatif Pneumonia ditemukan lebih akurat dengan klasifikasi Cosine KNN yang dapat mencapai 84,93333.

Kata Kunci: Identifikasi Penyakit, Pneumonia, K-Nearest Neighbors, Ekstraksi HOG, Fitur HOG

Abstract

Pneumonia is a wet lung disease. Pneumonia is generally caused by viruses, bacteria or fungi. Not infrequently Pneumonia can cause death. The K-Nearest Neighbors method is a classification method that uses the majority value from the closest k value category. At this time people are not too worried about pneumonia because this pneumonia has symptoms like a normal cough. However, fast and accurate information from health experts is also very necessary so that pneumonia symptoms can be recognized early and how to deal with them can also be done faster. In this study, researchers will diagnose pneumonia to obtain information quickly about the symptoms of pneumonia. This information will adopt human knowledge into computers designed to solve the problem of identifying pneumonia. In this study, the K-Nearest Neighbors method will be combined with the HOG Extraction Feature to identify pneumonia more accurately. The KNN classification used is Fine KNN, Cosine KNN, and Cubic KNN. Where will be seen how the value of accuracy, precision, recall, and f1-score. The results showed that the classification could run well on the Fine KKN, Cosine KNN, and Cubic KNN methods. Fine KNN has an accuracy rate of 80.67, Cosine KNN has an accuracy rate of 84,93333, and Cubic KNN has an accuracy rate of 83,13333. Fine KNN has precision, recall and f1-score values of 0.794842, 0.923706, and 0.854442. Cosine KNN has precision, recall and f1-score values of 0.803048, 0.954039, and 0.872056. Cubic KNN has precision, recall and f1-score values of 0.73388, 0.964561, and 0.833555. From the test results, positive and negative identification of pneumonia was found to be more accurate with the Cosine KNN classification which reached 84,93333.

Keywords: Disease Identification, Pneumonia, K-Nearest Neighbors, HOG Extraction, HOG Features

How to Cite: Khairina, N., Sibarani, T. T. S., Muliono, R., Sembiring, Z., & Muhathir, M. (2022). Identification of Pneumonia using The K-Nearest Neighbors Method using HOG Feature Extraction. *2JITE (Journal of Informatics and Telecommunication Engineering)*, 5(2), 563–569.

I. PENDAHULUAN

Penularan penyakit dapat terjadi secara langsung maupun melalui media tertentu, misalnya kontak langsung melalui sentuhan kulit, udara, dan air (Dharmayanti & Tjandararini, 2017). Pneumonia adalah penyakit menular yang mengkhawatirkan, pasien dengan pneumonia membutuhkan perawatan di rumah sakit dan memerlukan perawatan intensif (Z.H.Z & W.Lim, 2006). Di Indonesia pneumonia juga menempati urutan kedua dalam proporsi penyebab kematian. Oleh karena itu, tampaknya pneumonia adalah masalah kesehatan utama di Indonesia (Wahyuni & Ramadhan, 2019). Pneumonia dapat terjadi pada anak-anak dan juga orang dewasa (Abubakar, S, Fatimawali, & Y., 2019). Penelitian ini bertujuan untuk membantu tenaga medis memberikan diagnosis yang benar berdasarkan gejala yang dialami oleh pasien (A.Romadhony & Sulistiyo, 2014). Berdasarkan hal di atas, perlu untuk mengembangkan sistem prediksi untuk diagnosis pneumonia dengan menggunakan metode KNN (K-Nearest Neighbors) dengan memanfaatkan ekstraksi HOG, karena algoritma K-Nearest Neighbor mudah diterapkan dalam mendiagnosis suatu penyakit, dengan memanfaatkan ekstraksi HOG, karena metode HOG dikembangkan untuk mendeteksi objek lain (A, E.Santoso, & C.Dewi, 2019).

Klasifikasi KNN terhadap objek baru dilakukan berdasarkan atribut data latih serta data uji. Nilai mayoritas akan digunakan dalam menentukan jenis kelas dari contoh data selanjutnya (S.M & S.R, 2012). Fitur HOG akan digunakan untuk menghitung nilai gradient citra untuk mengenali ciri dari objek – objek tertentu (Rinaldi & Hendra, 2016). Tidak hanya itu, metode *Histogram of Oriented Gradient* (HOG) juga bisa melakukan deteksi fitur yang hasil akhirnya dapat menghasilkan tingkat akurasi yang baik (Yohannes, 2019). Penelitian terkait dengan HOG pernah dilakukan oleh Sovia, namun sistem dibangun menggunakan perancangan UML dan use case (R.Sovia & Mandala, 2019). Penelitian (Muhatir, Theofil, & Al-Khowarizmi, 2020) menyimpulkan bahwa metode HOG mampu mengidentifikasi penyakit Tuberkulosis dengan menggunakan metode K-NN.

Berdasarkan latar belakang yang telah di paparkan sebelumnya. Metode KNN dan HOG telah diterapkan pada identifikasi penyakit Tuberkulosis. Namun pada penelitian ini, peneliti tertarik untuk mengidentifikasi penyakit Pneumonia. Peneliti akan melakukan ekstraksi fitur *Histogram of Oriented Gradient* (HOG) pada citra yang diduga positif Pneumonia serta memanfaatkan metode KNN (*K-Nearest Neighbors*) dengan klasifikasi Fine KNN, Cosine KNN, Cubic KNN. Proses awal dilakukan dengan pre-processing, kemudian dilakukanlah proses ekstraksi ciri menggunakan ekstraksi fitur HOG, kemudian proses diagnosis dilakukan dengan metode KNN (*K-Nearest Neighbors*). Proses selanjutnya nilai hasil ekstraksi yang memuat ciri citra akan dipetakan untuk mendeteksi penyakit Pneumonia.

II. STUDI PUSTAKA

A. *Pneumonia*

Pneumonia merupakan penyakit yang menyerang jaringan paru-paru (*alveoli*) dan pada umumnya disebabkan oleh bakteri, virus maupun jamur (Rasyid, 2013). Adapun beberapa gejalanya, yaitu : napas sesak, peradangan paru-paru secara mendadak, demam, sesak napas, dahak berwarna kehijauan atau seperti karet (Abubakar, S, Fatimawali, & Y., 2019) (Heru, Onny, & Tri, 2012).

Angka kematian yang disebabkan oleh Pneumonia sangat tinggi. Pneumonia sering terjadi di beberapa negara Eropa. Di Amerika, Pneumonia yang dapat menyerang anak balita dan orang dewasa menjadi penyakit yang paling ditakuti (Christian, Ari, & Audrey, 2016).

B. *Histogram of Oriented Gradient*

Fitur HOG sering digunakan oleh *computer vision* dan *image processing*. HOG memiliki keunggulan, yaitu mampu menangkap tepi atau struktur gradient yang sangat karakteristik dari bentuk sebenarnya (N & T.B, 2005). Cara kerja HOG adalah dengan menghitung nilai *gradient* dari suatu daerah tertentu dari gambar. Setiap gambar memiliki gambar karakteristik yang di tunjukkan oleh *gradient* nilai yang diperoleh dengan membagi suatu citra menjadi wilayah terkecil yang disebut sel (Girsang & Muhathir, 2021) (Tanjung & Muhathir, 2020).

Algoritma *Histogram of Oriented Gradient* (HOG) sering digunakan untuk mengidentifikasi objek. HOG membuat orientasi *gradient* dari *pixel* citra. *Gradient* setiap *pixel* citra akan dihitung dan dibagi menjadi *cell*.

Setiap *cell* akan dibentuk histogramnya dan digabungkan ke kelompok yang lebih besar. Kelompok ini disebut *block*. Selanjutnya, normalisasi akan dilakukan pada tiap blok (N & B, 2005)(Muhathir et al., 2019).

C. K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor* sering digunakan untuk mengklasifikasikan objek (Larasati & Muhathir, 2021)(Sandy et al., 2019)(F.Liantoni, 2015). Algoritma *K-Nearest Neighbor* memiliki beberapa jenis klasifikasi yang sering digunakan, yaitu (Johnson & Yadav, 2018) :

1. Fine KNN : Pengklasifikasi tetangga terdekat yang membuat distinctions antara kelas dengan jumlah tetangga diatur ke 1.
2. Medium KNN : Kelas tetangga terdekat yang membuat perbedaan lebih sedikit dari pada Fine KNN dengan jumlah tetangga diatur ke 10.
3. Coarse KNN : Pengklasifikasi tetangga terdekat yang membuat perbedaan kasar antar kelas, dengan jumlah tetangga diatur ke 100.
4. Cosine KNN : Pengklasifikasi tetangga terdekat yang menggunakan jarak cosinus metrik.
5. Cubic KNN : Kelas tetangga terdekat yang menggunakan jarak kubik metrik.
6. Weighted KNN : Sebuah kelas tetangga terdekat yang menggunakan pembobotan jarak.

Jarak terdekat antara data *training* akan diukur dengan *euclidean distance* dan *manhattan distance*. Menurut penelitian terdahulu, *euclidean distance* mempunyai hasil lebih akurat dari pada *manhattan distance*. Adapun *euclidean distance* dapat dihitung menggunakan persamaan (Rasyid, 2013) (Rivan & Y.Yohannes, 2019).

III. METODE PENELITIAN

A. Dataset

Data set diambil dari *platform Kaggle* yang terdiri dari 5.840 gambar dengan kombinasi 1.575 gambar paru-paru normal dan 4.265 gambar paru-paru positif Pneumonia.

B. Tahapan Penelitian

Berikut ini tahapan penelitian *training* pola dan *testing* pola dengan algoritma KNN dan fitur HOG.

Gambar 1. Tahapan Penelitian (Muhathir, Theofil, & Al-Khowarizmi, 2020)

Gambar 1 terdiri dari dua proses, proses pelatihan pertama adalah proses pengolahan data *grayscale* serta mengekstraksi dengan memanfaatkan fitur (HOG). Setelah melakukan ekstraksi, proses kedua yaitu melakukan proses klasifikasi dengan algoritma KNN. Dari hasil klasifikasi dengan algoritma KNN, seseorang dapat dinyatakan sebagai Positif Pneumonia atau Negatif Pneumonia.

IV. HASIL DAN PEMBAHASAN

A. *X-Ray Set Pneumonia*

Sampel X-ray Set Pneumonia ini diambil dari platform Kaggle, Gambar 2 melampirkan beberapa sampel ke Set X-ray Pneumonia.

(a)

(b)

Gambar 2. Sampel X-ray Set Pneumonia (a) Negatif, (b) Positif.

B. *Deteksi HOG*

Hasil deteksi fitur HOG dapat dilihat pada Gambar 3 berikut ini. Gambar 3 menunjukkan hasil deteksi fitur HOG dengan 70 nilai kuat di gambar X-ray Pneumonia Set.

Gambar 3. Deteksi HOG Fitur

Dalam penelitian ini, penulis melakukan klasifikasi menggunakan tiga klasifikasi KNN yaitu Fine KNN, Cosine KNN, dan Cubic KNN. Akurasi klasifikasi mencapai 84.93333 dengan klasifikasi cosinus KNN seperti

yang ditunjukkan pada Tabel 1. Diagram Hasil Klasifikasi Pneumonia menggunakan KNN juga dapat dilihat pada Gambar 4.

Tabel 1. Klasifikasi KNN, Cosen KNN, Cubic KNN

Klasifikasi			Akurasi	Presisi	Recall	F1 Score
Fine KNN						
Known	Normal	Pneumonia	80.67	0.794842	0.923706	0.854442
Normal	82.07	17.93				
Pneumonia	20.73	79.27				
Cosine KNN						
Known	Normal	Pneumonia	84.93333	0.803048	0.954039	0.872056
Normal	89.53333	10.46667				
Pneumonia	19.66667	80.33333				
Cubic KNN						
Known	Normal	Pneumonia	83.13333	0.73388	0.964561	0.833555
Normal	92.53333	7.466667				
Pneumonia	26.26667	73.73333				

Tabel 2. Classification Accuracy

Akurasi Klasifikasi	Fine KNN	Cosine KNN	Cubic KNN
Normal	82.07	89.53	92.53
Pneumonia	79.27	80.33	73.73
Average Accuracy	80.67	84.93	83.13

Gambar 4. Diagram Hasil Klasifikasi Pnemonia menggunakan KNN

Dalam tabel menunjukkan hasil klasifikasi menggunakan metode menggunakan tiga klasifikasi yaitu Fine KNN menghasilkan tingkat akurasi 80,67, cosine KNN menghasilkan tingkat akurasi 84.93333, dan KNN Cubik dengan tingkat akurasi 83.13333.

V. SIMPULAN

Identifikasi lebih detail tentang penyakit Pneumonia dapat bermanfaat pada perkembangan ilmu pengetahuan khususnya dunia kedokteran. Gejala penyakit Pneumonia yang diawali dari batuk biasa, sering sekali dianggap hal yang ringan dan tidak membutuhkan penanganan yang serius, padahal penyakit Pneumonia dapat menyebabkan kematian. Penelitian ini mengidentifikasi gejala penyakit Pneumonia dengan metode *K-Nearest Neighbor* dengan tiga klasifikasi yaitu Fine KNN, Cosine KNN dan Cubic KNN serta dikombinasikan dengan fitur *Histogram of Oriented Gradient*. Hasil penelitian menunjukkan tingkat akurasi klasifikasi set gambar X-ray Pneumonia menggunakan metode KNN dan HOG memiliki hasil yang berbeda-beda. Fine KNN mencapai akurasi sebesar 80,67, Cosine KNN mencapai akurasi sebesar 84.93333, dan Cubic KNN mencapai akurasi sebesar 83.13333. Dari tiga klasifikasi KNN, dapat dilihat bahwa klasifikasi Cosine KNN memiliki nilai yang paling akurat yaitu sebesar 84.93333.

DAFTAR PUSTAKA

- Santoso, A. P., E. & Dewi, C. (2019). Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor dan Naive Bayes. *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(4), 3319-3324.
- Romadhony, B. W., & Sulistiyo, M. (2014). Analisis dan Implementasi Sistem Pendiagnosis Penyakit Tuberculosis Menggunakan Metode Case-Based Reasoning. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, (pp. 1907-5022).
- Bakar, A. P., Fatimawali, & Y.P.V.Y. (2019). Uji Daya Hambat Ekstrak Rimpang Lengkuas Merah (Alpinia purpurata K.Schum Terhadap Pertumbuhan Bakteri *Klebsiella pneumoniae* Isolat Sputum pada Penderita Pneumoniaresisten Antibiotik Seftriakson. *PHARMACON Jurnal Ilmiah Farmasi-UNSRAT*, 8(1), 11-21.
- Abubakar, S. P. M., Fatimawali, & Y., Y. P. (2019). Uji Daya Hambat Ekstrak Rimpang Lengkuas Merah (Alpinia purpurata K.Schum) Terhadap Pertumbuhan Bakteri *Klebsiella pneumoniae* Isolat Sputum Pada Penderita Pneumoniaresisten Antibiotik Seftriakson. *PHARMACON Jurnal Ilmiah Farmasi - UNSRAT*, 8(1), 11-21.
- Christian, T. K., Ari, L. R., & Audrey, M. I. (2016, Desember). Gambaran karakteristik pneumonia pada anak yang dirawat di ruang perawatan intensif anak RSUP Prof. Dr. R. D. Kandou Manado periode 2013 – 2015. *Jurnal e-Clinic (eCl)*, 4(2).
- Dalal, N., & Triggs, B. (2005). Histogram of Oriented Gradient for Human Detection," *Computer Vision and Pattern. 1*.
- Dharmayanti, I., & Tjandrarini, D. (2017). Identifikasi Indikator dalam Indeks Pembangunan Kesehatan Masyarakat (IPKM) untuk Meningkatkan Nilai Sub-Indeks Penyakit Menular. *JKP*, 5(3), 249-257.
- F.Liantoni. (2015). Klasifikasi Daun dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *ULTIMATICS*, 7(2).
- Girsang, N. D., & Muhathir, M. (2021). *Classification Of Batik Images Using Multilayer Perceptron With Histogram Of Oriented Gradient Feature Extraction*. 4(February), 197-204.
- Heru, P., Onny, S., & Tri, J. (2012, Oktober). Hubungan Faktor-Faktor Lingkungan Fisik Rumah dengan Kejadian Pneumonia pada Balita di Wilayah Kerja Puskesmas Jatibarang Kabupaten Brebes. *Jurnal Kesehatan Lingkungan Indonesia*, 11(2), 194-198.
- Johnson, J. M., & Yadav, A. (2018). Fault Detection and Classification Technique for HVDC Transmission Lines Using KNN. *Information and Communication Technology for Sustainable Development*. Springer.
- Josefa, R., Sovia, R., & Mandala, E. P. (2019). Sistem Pakar Diagnosa Penyakit Pneumonia Pada Anak Menggunakan Metode Case Based Reasoning. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 868 - 872.
- Larasati, D. A., & Muhathir, M. (2021). Application of the K-NN Method and GLCM Feature Extraction in Classifying Formalin Fish Images. *Journal Of Research Computer Science*, 1(1), 1-13.
- Liantoni, F. (2015, Desember). Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor. *ULTIMATICS*, 7(2).
- M, S., T, T., & R, S. (2012). Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients. *International Journal of Information and Education Technology*, 2(3), 220-223.
- Muhathir, Theofil, T. S., & Al-Khowarizmi. (2020). Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction. *Jurnal Teknik Informatika dan Sistem Informasi*, 1(1), 33-38.

- Muhathir, Rizal, R. A., Sihotang, J. S., & Gultom, R. (2019). Comparison of SURF and HOG extraction in classifying the blood image of malaria parasites using SVM. *2019 International Conference of Computer Science and Information Technology, ICoSNIKOM 2019*. <https://doi.org/10.1109/ICoSNIKOM48755.2019.9111647>
- N, D., & B, T. (2005). Histograms of Oriented Gradients for Human Detection. *EEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893.
- N, D., & T.B. (2005). Histogram of Oriented Gradients for Human Detection. *EEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 11(2), 886-893.
- Sovia, R., & Mandala, E. (2019). Sistem Pakar Diagnosa Penyakit Pneumonia Pada Anak Menggunakan Metode Case Based Reasoning. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, (pp. 868-872).
- Sandy, B., Siahaan, J. K., Permana, P., & Muhathir, M. (2019). Klasifikasi Citra Wayang Dengan Menggunakan Metode k-NN & GLCM. *Prosiding Seminar Nasional Teknologi Informatika*, 2(1), 71–77.
- Tanjung, J. P., & Muhathir, M. (2020). Classification of facial expressions using SVM and HOG. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, 3(2), 210–215. <https://doi.org/10.31289/jite.v3i2.3182>