

Support Vector Machine Optimization Using Grid Search And GLCM Feature Extraction For Identification Of Lung Cancer

Optimasi Support Vector Machine Menggunakan Grid Search Dan Ekstraksi Fitur GLCM Untuk Identifikasi Kanker Paru-Paru

***Muhammad Abdul Ghani¹, Bambang Yuwono², Dessyanto Boedi Prasetyo³**

^{1,2,3} Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

^{1*}123180020@student.upnyk.ac.id ²bambangy@upnyk.ac.id ³dess@upnyk.ac.id

Informasi Artikel

Received: -

Revised: -

Accepted: -

Published: -

Abstract

Purpose:

The purpose of this study was to determine the effect of the Grid Search method on the results of SVM accuracy in identifying lung cancer images.

Design/methodology/approach:

This study uses the grid search algorithm to optimize the support vector machine method.

Findings/result:

The results of the study show that the Grid Search algorithm provides an increase in the accuracy value of 7.5% from 79.9% to 87.4%. An increase also occurred in the precision value of 8.1% from 81.6% to 89.7%. In addition, there was also an increase in recall by 4.1% from 82.7% to 88.9% and f1-score by 6.2% from 82.7% to 88.9%. The Grid Search algorithm also reduces the error value by 7.5% from 20% to 12.5%.

Originality/value/state of the art:

This study optimizes support vector machines using grid search and GLCM feature extraction to identify lung cancer.

Abstrak

Tujuan:

Tujuan dari penelitian ini adalah untuk mengetahui pengaruh metode Grid Search terhadap hasil akurasi SVM dalam identifikasi citra kanker paru-paru.

Perancangan/metode/pendekatan:

Penelitian ini menggunakan algoritma grid search untuk optimasi metode support vector machine.

Hasil:

Hasil penelitian menunjukkan bahwa algoritma Grid Search

*Keywords:*Support Vector Machine;

Grid Search; GLCM; Lung Cancer

Kata kunci: support vector machine;

Grid Search; GLCM; Kanker Paru-paru

memberikan peningkatan pada nilai *accuracy* sebesar 7.5% dari 79.9% menjadi 87.4%. Peningkatan juga terjadi pada nilai *precision* sebesar 8.1% dari 81.6% menjadi 89.7%. Selain itu, peningkatan juga terjadi pada *recall* sebesar 4.1% dari 82.7% menjadi 88.9% dan *f1-score* sebesar 6.2% dari 82.7% menjadi 88.9%. Algoritma *Grid Search* juga memberi penurunan terhadap nilai *error* sebesar 7.5% dari 20% menjadi 12.5%.

Keaslian/ *state of the art*:

Penelitian ini melakukan optimasi terhadap *support vector machine* menggunakan *grid search* dan ekstraksi fitur GLCM untuk identifikasi kanker paru-paru.

1. Pendahuluan

Mendiagnosis suatu penyakit adalah pekerjaan yang sangat kompleks dan diperlukan banyak pengujian pada pasien untuk mencapai kesimpulan yang tepat. Hal ini dapat mengarahkan untuk menggunakan perangkat analitik, yang direncanakan untuk membantu para dokter dalam mengambil keputusan. Penentuan dini mengurangi waktu perawatan dan dapat menyelamatkan nyawa. Salah satu penyakit ini adalah pertumbuhan kanker ganas paru-paru, yang terjadi ketika sel-sel di jaringan paru-paru berkembang secara tidak terkendali[1]. Kasus-kasus ini sering disebabkan oleh kombinasi faktor genetik dan paparan gas radon, asbes, perokok pasif, atau bentuk polusi udara lainnya. Kanker paru-paru dapat dilihat pada radiografi dada dan *computed tomography (CT) scan*. Diagnosis dipastikan dengan biopsi yang biasanya dilakukan dengan *bronchoscopy* atau *CT-guidance*. Keganasan paru merupakan salah satu penyakit yang menyebabkan 1,61 juta kematian di dunia setiap tahunnya[10]. Kecerdasan Buatan membuat mesin mampu secara cerdas menjawab pertanyaan saat ini di bidang teknik yang sangat luas. Pembelajaran mesin memberikan teknik kecerdasan buatan terbaik seperti klasifikasi, regresi, pembelajaran *supervised* dan *unsupervised*, dan banyak lagi[11].

Penelitian yang berhubungan dengan klasifikasi kanker paru-paru pernah dilakukan dengan metode *Naive Bayes* dengan ekstraksi fitur GLCM (*Gray Level Co-occurrence Matrix*) oleh (Yuniarto et al., 2021). 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%[13]. Terdapat pula penelitian lain yang pernah dilakukan oleh Anifah (2017). Deteksi kanker dilakukan dengan metode ANN (*Artificial Neural Network Back-Propagation*). Dari database kanker paru-paru, total 50 gambar dipertimbangkan dan gambar-gambar ini dibagi menjadi dua kelompok gambar kanker dan nonkanker. Fitur diekstraksi dengan bantuan

GLCM dan ini diberikan ke ANN. Kanker terdeteksi dan diperoleh akurasi 80%[3]. Namun, pada penelitian tersebut belum di terapkan algoritma pemilihan parameter atau optimasi parameter pada metode SVM yang digunakan.

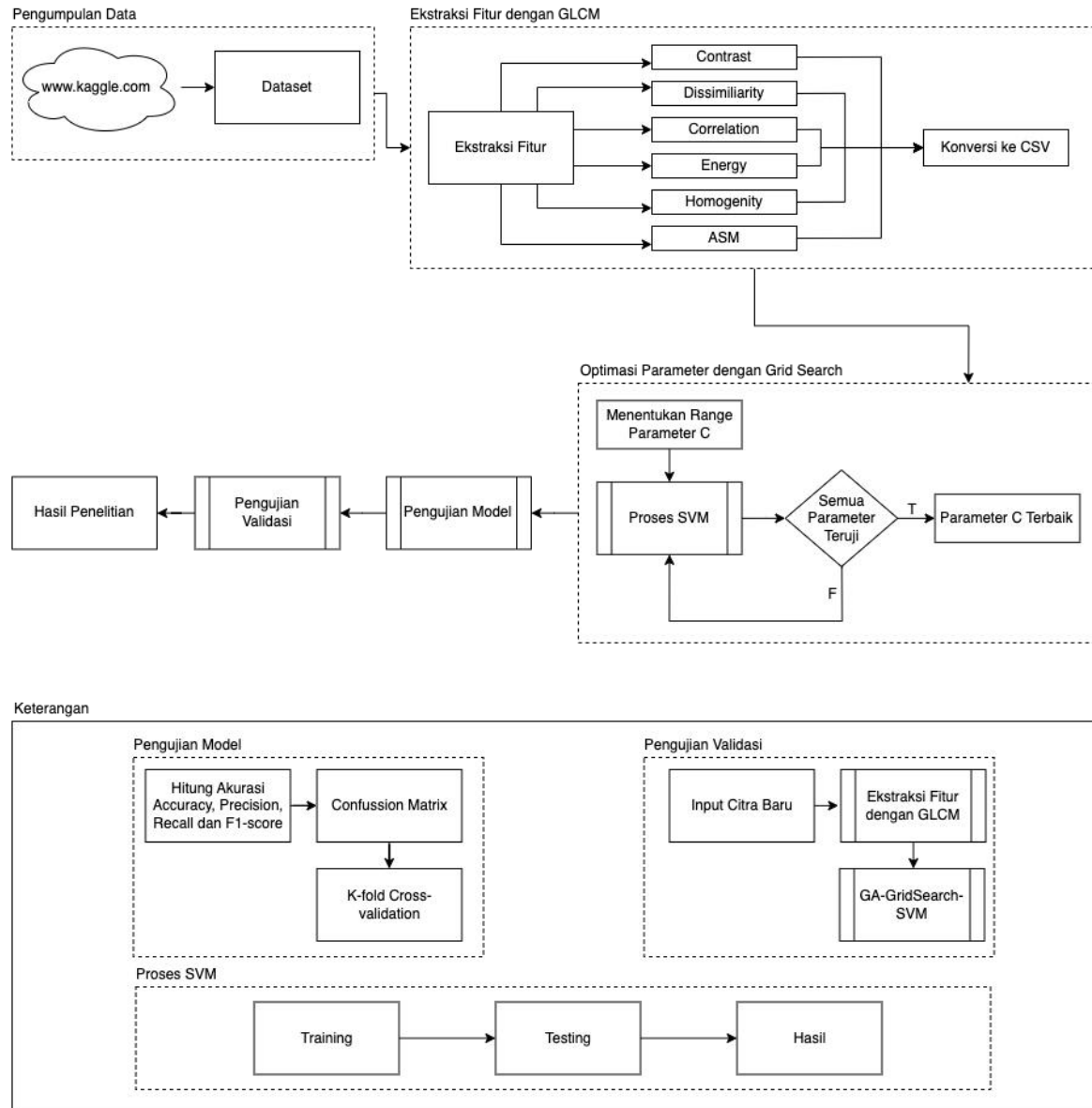
SVM merupakan metode yang memiliki kelebihan mampu melakukan generalisasi data dengan hasil akurasi yang tinggi[6]. SVM berkerja dengan memisahkan kelas satu dengan kelas lainnya menggunakan hyperplane yang optimal[6]. SVM merupakan metode *machine learning* yang baik digunakan untuk klasifikasi teks[6]. Kemampuan SVM dalam melakukan klasifikasi teks pengaruhi kernel yang di gunakan[7]. Penelitian terkait klasifikasi teks menggunakan metode SVM dengan beberapa kernel pernah dilakukan oleh (Yuniarto, et al, 2020)[8]. Hasil penelitian menunjukan SVM dengan kernel RBF mendapatkan nilai 0.815 sedangkan SVM kernel linear mendapatkan nilai 0.775[8]. Maka dari hal tersebut, pada penelitian yang sedang dilakukan akan digunakan SVM dengan kernel RBF.

SVM (*Support Vector Machine*) mampu melakukan generalisasi data dengan hasil akurasi yang tinggi [9]. SVM bekerja dengan memisahkan kelas satu dengan kelas lainnya menggunakan hyperplane yang optimal[9]. SVM mampu memberikan kinerja klasifikasi suatu citra dengan baik [3]. Akan tetapi SVM memiliki kelemahan yaitu kesulitan dalam menentukan nilai parameternya. Sedangkan pemilihan parameter yang tepat pada SVM akan sangat berpengaruh terhadap hasil akurasi klasifikasi yang dilakukan [12]. Untuk mendapatkan hasil yang baik menggunakan metode SVM, perlu dilakukan optimasi dengan memilih parameter SVM yang paling optimal. Penelitian yang berkaitan dengan penggunaan metode SVM dalam melakukan klasifikasi pernah dilakukan (Apriani et al., 2022) yang mengoptimasi SVM dengan Grid Search. Penelitian ini menghasilkan akurasi sebesar 94%[4].

Pada penelitian ini akan dilakukan optimasi parameter SVM menggunakan algoritma *grid search* untuk identifikasi kanker paru-paru. Pemilihan parameter C dilakukan dari *range* 0,01 sampai 1000 dengan kelipatan 10.

2. Metode/Perancangan

Penelitian ini dilakukan dengan metode kuantitatif dan menggunakan data sekunder. Tahapan penelitian diilustrasikan pada Gambar 1.



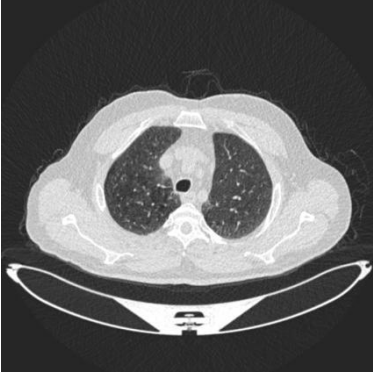

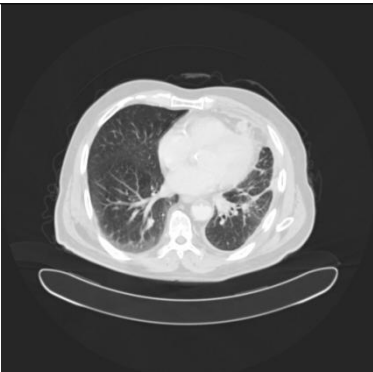
Gambar 1. Metodologi Penelitian


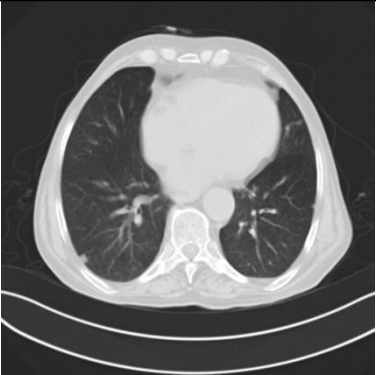
2.1. Persiapan Dataset

Proses pengumpulan data untuk penelitian ini menggunakan dataset sekunder dari situs Kaggle yang menyediakan sumber data terbuka. Pada situs ini telah dipilih dataset kanker paru-paru dengan total 977 data. Data tersebut terbagi menjadi 416 citra paru-paru normal dan 561 citra paru-paru terkena kanker. Format data dapat dilihat pada Tabel 1.

Tabel 1 Dataset Citra CT-scan Kanker Paru-paru

Sumber: <https://www.kaggle.com/>

| Citra | Nama | Kelas |
|---|------------------------|--------|
|  | Normal case (1).jpg | Normal |
|  | Normal case (2).jpg | Normal |
|  | Malignant case (1).jpg | Cancer |

| | | |
|---|------------------------|--------|
|  | Malignant case (2).jpg | Cancer |
|  | Malignant case (9).jpg | Cancer |

2.2. Ekstraksi Fitur dengan GLCM

Dataset yang awalnya berupa citra CT-scan diekstraksi menjadi nilai matriks menggunakan metode GLCM. Ada enam fungsi yang digunakan GLCM untuk mengekstrak dataset yaitu: Contrast, Energy, Correlation, Dissimilarity, Homogeneity, dan ASM.

2.3. Splitting Data

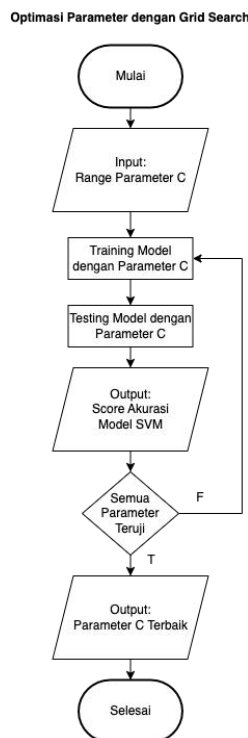
Dataset kemudian dibagi menjadi data latih dan data uji. Penelitian ini menggunakan rasio 7 banding 3. Pembagian dataset dijelaskan pada Tabel 2.

Tabel 2 Pembagian data training dan data testing

| No | Nama | Jenis | Jumlah |
|----|------------------------|----------|--------|
| 1 | Citra Paru-paru Normal | Training | 291 |
| 2 | Citra Paru-paru Kanker | Training | 393 |
| 3 | Citra Paru-paru Normal | Testing | 125 |
| 4 | Citra Paru-paru Kanker | Testing | 168 |

2.4. Permodelan Arsitektur Grid Search-SVM

Proses training SVM yang teroptimasi dengan *Grid Search* melewati beberapa tahap. Berikut flowchart dari proses training model SVM yang teroptimasi *Grid Search* diilustrasikan pada Gambar 2 berikut :



Gambar 2 Model Arsitektur *Grid Search*-SVM

Proses optimasi yang pertama dilakukan adalah menggunakan Grid Search. Dimana langkah pertama adalah menentukan grid atau range parameter SVM yang akan dioptimasi, dalam hal ini adalah parameter C. Range parameter yang akan diuji adalah, $C=[0.01, 0.1, 1, 10, 100, 1000]$. Setelah menentukan rentang parameter yang akan dioptimalkan, training dijalankan terhadap SVM yang dibuat dan data pelatihan diuji dalam proses *cross validation*.

2.5. Validasi dan Pengujian

Dataset yang awalnya berupa citra CT-scan diekstraksi menjadi nilai matriks menggunakan metode GLCM. Ada enam fungsi yang digunakan GLCM untuk mengekstrak dataset yaitu: *Contrast*, *Energy*, *Correlation*, *Dissimilarity*, *Homogeneity*, dan *ASM*. Pengujian model dilakukan dengan menggunakan Confusion Matrix dan *K-fold Cross Validation*. Model yang akan diuji adalah model SVM yang sudah teroptimasi maupun yang belum teroptimasi *Grid Search*. Pengujian dengan *Confusion Matrix* dilakukan Terhadap model *Support Vector Machine* dengan optimasi parameter menggunakan *Grid Search* dan model *Support Vector Machine* tanpa optimasi parameter.

3. Hasil dan Pembahasan

Penelitian ini menghasilkan 2 pengujian yaitu pengujian model SVM tanpa optimasi parameter dan pengujian model SVM dengan optimasi parameter menggunakan algoritma *grid search* (GS-SVM).

3.1. Pengambilan Data

Pengambilan data dilakukan untuk menyimpan data di *Google Colaboratory*. Dalam hal ini data diambil dari situs *Kaggle*, disimpan di *Google Drive* dan diakses melalui *Google Colaboratory*.

3.2. Ekstraksi Fitur GLCM

Ekstraksi fitur bertujuan untuk mengubah citra menjadi angka melalui proses matematis. Metode yang digunakan untuk proses ekstraksi fitur adalah metode GLCM yang kemudian akan diubah bentuk menjadi *dataframe* sehingga didapatkan hasil sebagai berikut.

Tabel 3. Hasil Ekstraksi Fitur GLCM

| | name | label | contrast | energy | correlation | dissimilarity | homogeneity | ASM |
|-----|--------------------------|-------|-----------------|-----------------------|------------------|---------------|------------------|------------------------|
| 0 | Normal case (253).jpg | 0 | 2770.2508 71 | 0.0541 260531 5 | 0.7541659 657 | 36.60716377 | 0.102421820 3 | 0.00292 962963 |
| 1 | Malignant case (131).jpg | 1 | 693.04153 95 | 0.0685 340545 3 | 0.9400858 433 | 10.25252327 | 0.287793738 9 | 0.00469 691663 |
| 2 | Normal case (251).jpg | 0 | 2783.1182 05 | 0.0545 289535 2 | 0.7540835 519 | 36.79577324 | 0.103092171 2 | 0.00297 340677 2 |
| 3 | Malignant case (500).jpg | 1 | 691.30605 51 | 0.1102 60075 | 0.9451223 048 | 11.82108682 | 0.243816968 8 | 0.01215 728414 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

3.3. Pembagian Data Training dan Data Testing

Pembagian data dilakukan untuk mengecek kinerja dari metode yang digunakan. Data training digunakan untuk melatih metode klasifikasi. Data testing digunakan untuk mengukur hasil akurasi pada data pelatihan. Data akan dibagi dengan rasio 70% untuk data training dan 30% data testing.

3.4. Pengujian Model SVM

Hasil pengujian model SVM tanpa optimasi parameter untuk mengidentifikasi kanker paru-paru pada citra CT-Scan didapatkan nilai *accuracy* sebesar 79.9%, *precision* sebesar 81.6%, *recall* sebesar 84% dan *f1-score* sebesar 82.7% serta nilai *error* sebesar 20%.

3.5. Pencarian Parameter C Terbaik dengan *Grid Search*

Parameter yang diuji adalah range 0.1 sampai dengan 1000 dengan kelipatan 10. Hasil uji menggunakan *Grid Search* ditampilkan pada tabel berikut.

Tabel 4. Hasil *Grid Search*

| No | C | Score | Rata-Rata |
|----|------------|--------------|--------------|
| 1 | 0.01 | 0.715 | 0.767 |
| 2 | 0.01 | 0.759 | |
| 3 | 0.01 | 0.810 | |
| 4 | 0.01 | 0.779 | |
| 5 | 0.01 | 0.772 | |
| 6 | 0.1 | 0.730 | 0.768 |
| 7 | 0.1 | 0.745 | |
| 8 | 0.1 | 0.825 | |
| 9 | 0.1 | 0.787 | |
| 10 | 0.1 | 0.757 | |
| 11 | 1 | 0.788 | 0.797 |
| 12 | 1 | 0.781 | |
| 13 | 1 | 0.832 | |
| 14 | 1 | 0.794 | |
| 15 | 1 | 0.794 | |
| 16 | 10 | 0.869 | 0.836 |
| 17 | 10 | 0.810 | |
| 18 | 10 | 0.847 | |
| 19 | 10 | 0.838 | |
| 20 | 10 | 0.816 | |
| 21 | 100 | 0.905 | 0.855 |
| 22 | 100 | 0.832 | |
| 23 | 100 | 0.869 | |
| 24 | 100 | 0.846 | |
| 25 | 100 | 0.824 | |
| 26 | 1000 | 0.869 | 0.847 |
| 27 | 1000 | 0.832 | |
| 28 | 1000 | 0.854 | |
| 29 | 1000 | 0.838 | |
| 30 | 1000 | 0.846 | |

Dapat dilihat bahwa rata-rata *score* tertinggi didapat saat C bernilai 100 dengan rata-rata *score* sebesar 0.855.

3.6. Pengujian Model *Grid Search*-SVM

Hasil pengujian model SVM dengan optimasi parameter menggunakan *Grid Search* untuk mengidentifikasi kanker paru-paru pada citra CT-Scan didapatkan nilai *accuracy* sebesar 87.4%, *precision* sebesar 89.7%, *recall* sebesar 88.1% dan *f1-score* sebesar 88.9% serta nilai *error* sebesar 12.5%.

3.7. Pembahasan

Pengaruh *grid search* terhadap keakuratan metode SVM dapat dilihat dengan membandingkan hasil model SVM dengan model GS-SVM. Perbandingan hasil dapat dilihat tabel 1. Pengaruh *Grid Search* terhadap model SVM dapat dilihat pada Tabel 3 berikut.

Tabel 5. Perbandingan SVM dan *Grid Search*-SVM

| Pengujian | Nilai | | Perubahan |
|------------------|-------|-------------------------|-----------|
| | SVM | <i>Grid Seach</i> - SVM | |
| <i>Accuracy</i> | 79.9% | 87.4% | + 7.5% |
| <i>Precision</i> | 81.6% | 89.7% | + 8.1% |
| <i>Recall</i> | 84% | 88.1% | + 4.1% |
| <i>F1-Score</i> | 82.7% | 88.9% | + 6.2% |
| <i>Error</i> | 20% | 12.5% | - 7.5% |

Dari Tabel 3 dapat dilihat bahwa algoritma *Grid Search* memberikan peningkatan pada nilai *accuracy* sebesar 7.5% dari 79.9% menjadi 87.4%. Peningkatan juga terjadi pada nilai *precision* sebesar 8.1% dari 81.6% menjadi 89.7%. Selain itu, peningkatan juga terjadi pada *recall* sebesar 4.1% dari 82.7% menjadi 88.9% dan *f1-score* sebesar 6.2% dari 82.7% menjadi 88.9%. Algoritma *Grid Search* juga memberi penurunan terhadap nilai *error* sebesar 7.5% dari 20% menjadi 12.5%.

4. Kesimpulan dan Saran

Dalam pengujian model yang telah dilakukan, dapat disimpulkan bahwa penggunaan *Grid Search* yang merupakan algoritma optimasi parameter dapat memengaruhi hasil akurasi pada SVM untuk identifikasi kanker paru-paru.

Melalui *Grid Search*, didapatkan hasil bahwa parameter terbaik dalam uji ini adalah dengan nilai C=100 pada kernel linear. Pengaruh *Grid Search* terhadap SVM dapat dilihat pada peningkatan nilai *accuracy* sebesar 7.5% dari 79.9% menjadi 87.4%. Ini menunjukkan bahwa model *Grid Search*-SVM semakin baik dalam memprediksi keluaran yang benar. Peningkatan juga terjadi pada nilai *precision* sebesar 8.1% dari 81.6% menjadi 89.7% yang berakibat pada semakin sedikitnya objek positif yang salah diklasifikasikan sebagai objek negatif. Selain itu, peningkatan juga terjadi pada *recall*

sebesar 4.1% dari 82.7% menjadi 88.9% yang menunjukkan semakin sedikitnya nilai positif yang terlewatkan. Terakhir adalah peningkatan terhadap f1-score sebesar 6.2% dari 82.7% menjadi 88.9%. yang menunjukkan bahwa model Grid Search-SVM semakin baik dalam mengklasifikasikan objek positif dan negatif dengan akurasi tinggi. Algoritma Grid Search juga memberi penurunan terhadap nilai error sebesar 7.5% dari 20% menjadi 12.5%.

Sebagai saran untuk peneliti selanjutnya agar melakukan penelitian menggunakan kernel lain seperti RBF dan polynomial, serta melakukan pengujian terhadap jenis citra lainnya.

Daftar Pustaka

- [1] Al-Kadi, O. S. (2015a). A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumors. *Computerized Medical Imaging and Graphics*, 41(December), 67-79. <https://doi.org/10.1016/j.compmedimag.2014.05.013>
- [2] Ankita, R., Kumari, C. U., Mehdi, M. J., Tejashwini, N., Pavani, T. (2019). Lung cancer image-feature extraction and classification using GLCM and SVM classifier. *International Journal of Innovative Technology and Exploring Engineering*, 8(11), 2211–2215. <https://doi.org/10.35940/ijitee.K2044.0981119>
- [3] Anifah, L..., Haryanto, Harimurti, R.P., Permatasari, Z., Rusimamto, P.W., & Muhamad, A.R. (2017). Cancer lungs detection on CT scan image using artificial neural network backpropagation based gray level co-occurrence matrices feature. 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 327-332.
- [4] Apriani, J. H. Jaman, and R. I. Adam, "Optimasi SVM menggunakan algoritme grid search untuk identifikasi citra biji kopi robusta berdasarkan circularity dan eccentricity," *Jurnal Teknologi dan Sistem Komputer*, vol. 10, no. 1, pp. 12-19, Jan. 2022. <https://doi.org/10.14710/jtsiskom.2021.13807>
- [5] Aranguren, I., Valdivia, A., Morales-Castaneda, B., Oliva, D., Abd Elaziz, M., & Perez-Cisneros, M. (2021). Improving the segmentation of magnetic resonance brain images using the LSHADE optimization algorithm. *Biomedical Signal Processing and Control*, 64(October 2020). <https://doi.org/10.1016/j.bspc.2020.102259>
- [6] Doan, Q. H., Le, T., & Thai, D. K. (2021). Optimization strategies of neural networks for impact damage classification of RC panels in a small dataset. *Applied Soft Computing*, 102, 107100, <https://doi.org/10.1016/j.asoc.2021.107100>
- [7] Elnakib, A., Amer, H. M., Abou-Chadi, F. E. Z. (2020). Early lung cancer detection using deep learning optimization. *International Journal of Online and Biomedical Engineering*, 16(6), 82–94. <https://doi.org/10.3991/ijoe.v16i06.13657>
- [8] Eshan, S. C., & Hasan, M. S. (2018). An application of Machine Learning to detect abusive Bengali text. In 20th International Conference of Computer and Information Technology, ICCIT 2017 (Vols. 2018-Janua, pp. 1–6). IEEE. <https://doi.org/10.1109/ICCITECHN.2017.8281787>
- [9] Karayigit, H., Inan Aci, C., & Akdagli, A. (2021). Detecting abusive Instagram comments in Turkish using convolutional neural network and *Machine Learning* methods. *Expert Systems with Applications*, 174(February). <https://doi.org/10.1016/j.eswa.2021.114802>

- [10]Li, J., Usevich, K., & Comon, P. (2018). Globally convergent Jacobi-type algorithms for simultaneous orthogonal symmetric tensor diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 39(1), 1-22. [https://doi.org/10,1137/17M1116295](https://doi.org/10.1137/17M1116295).
- [11]Putra, T. D., Utami, E., & Kurniawan, M. P. (2022). *Klasifikasi penderita kanker Paru Paru Menggunakan Algoritma Artificial Neural Network (ANN)* 13 *Klasifikasi penderita kanker Paru Paru Menggunakan Algoritma Artificial Neural Network (ANN)*.
- [12]Soumaya, Z., Drissi Taoufiq, B., Benayad, N., Yunus, K., & Abdelkrim, A. (2021). The detection of Parkinson disease using the genetic algorithm and SVM classifier. *Applied Acoustics*, 171, 107528. [https://doi.org/10,1016/j.apacoust.2020,107528](https://doi.org/10.1016/j.apacoust.2020.107528)
- [13]Yunianto, M., Anwar, F., Nur Septianingsih, D., Dwi Ardyanto, T., & Farits Pradana, R. (2021). KLASIFIKASI KANKER PARU PARU MENGGUNAKAN NAIVE BAYES DENGAN VARIASI FILTER DAN EKSTRAKSI CIRI GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM). *Indonesian Journal of Applied Physics*, 11(2).