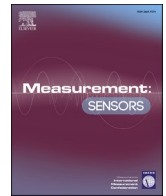


Daftar isi tersedia di [Sains Langsung](#)

Pengukuran: Sensor

beranda jurnal: www.sciencedirect.com/journal/measurement-sensors

Klasifikasi kanker paru-paru berdasarkan peningkatan pembelajaran mendalam menggunakan data ekspresi gen

V. Yuvaraj^{A,B,*}, D. Maheswari^C^ADepartemen Ilmu Komputer, Sekolah Tinggi Seni dan Sains Rathnavel Subramaniam (RVS), Sullur, Coimbatore, India^BDepartemen Aplikasi Komputer, Dr.INGP Seni dan Sains, Coimbatore, 641048, India^CKepala dan Koordinator Penelitian, PG Departemen Ilmu Komputer, Sekolah Tinggi Seni dan Sains Rathnavel Subramaniam (RVS), Sullur, Coimbatore, India

INFO PASAL

Kata kunci:

Kanker paru-paru

Klasifikasi

Seleksi gen

Teknologi microarray dan normalisasi

ABSTRAK

Kanker paru-paru merupakan salah satu kanker yang mengancam jiwa secara global. Perkiraan World Cancer Research Fund International bahwa pada tahun 2022, penyakit ini telah mencatat diagnosis 1,8 juta kasus baru penyakit ini. Para profesional medis dapat merawat pasien dengan aman dan efektif ketika mereka membuat diagnosis proaktif dan mengklasifikasikan kondisi mereka. Ketersediaan teknologi microarray telah membuka jalan bagi penyelidikan gen dan hubungannya dengan banyak penyakit, termasuk kanker paru-paru. Penelitian sebelumnya telah menyarankan pemilihan gen menggunakan IWOA (Improved Whale Optimization Algorithm). Setelah itu, Jaringan MLSTM (Modified Long Short-Term Memory) diperhitungkan untuk kategorisasi kanker paru-paru. Namun demikian, karena data masukan direpresentasikan dalam skala yang berbeda, LSTM cenderung mengalami overfitting dan sifat yang menonjol (dengan skala yang menurun) mungkin menjadi tidak berguna karena fitur lain memiliki nilai pada skala yang lebih tinggi. Untuk mengatasi masalah tersebut dalam penelitian ini, perkenalkan model klasifikasi kanker paru yang disempurnakan. Dalam pekerjaan ini, awalnya dilakukan pra-pemrosesan data untuk normalisasi skala data menggunakan normalisasi min max. Seleksi gen dilakukan dengan menggunakan IWOA. Terakhir, ECNN (Enhanced convolution neural network) dipertimbangkan untuk mengklasifikasikan kanker paru-paru. Hasil eksperimen dieksekusi pada Matlab 2013b menggunakan Kent Ridge Bio-Medical Dataset, kerangka ICNN yang diusulkan dalam penelitian ini dibandingkan dengan teknik MIMOGA, SMO, dan MLSTM yang sudah ada sebelumnya. Hasil dari model yang diusulkan ini menunjukkan efektivitasnya ketika dievaluasi menggunakan ukuran kinerja seperti nilai recall, presisi, akurasi, dan f-measure.

1. Perkenalan

Baru-baru ini, terjadi peningkatan jumlah kematian akibat kanker paru-paru berdasarkan laporan dari berbagai organisasi medis. Dibandingkan dengan permulaan abad ini, terdapat peningkatan besar dalam jumlah kematian akibat kanker [1]. Terapi yang ada seperti kemoterapi, radiasi, dan pembedahan, meskipun hasilnya kurang memuaskan. Kategorisasi yang akurat dari berbagai kategori kanker paru-paru sangat penting untuk efektivitas terapi dan pengurangan dampak buruknya pada pasien. Dengan penggunaan teknologi microarray, jutaan gen dapat dipelajari untuk mendapatkan rincian penting tentang cara kerja sel. Data ini dapat digunakan untuk mendiagnosis dan memprediksi kanker [2].

Sangat penting untuk menciptakan pendekatan berkualitas baik yang memilih subset gen yang signifikan terutama karena karakteristik data gen seperti ukuran sampel yang lebih kecil, tingkat kebisingan yang lebih tinggi, dan dimensi yang lebih besar), yang berguna dalam mengkategorikan kanker dengan lebih baik. Metode ini membantu dokter mengkategorikan a

sampel yang lebih kecil dari tumor berbasis gen yang penting secara fisiologis dan menyoroti beberapa gen khusus untuk pembangunan uji coba berbiaya rendah selain mampu mengurangi biaya komputasi [3,4], serta manfaat tambahan dari diagnosis dini kanker dan terapi terkaitnya. Selain itu, penelitian telah membuktikan kegunaan informasi sejarah biologis dalam evaluasi statistik dalam hal mengidentifikasi karakteristik biologis yang unik sekaligus menghilangkan hubungan yang tidak jelas di antara karakteristik tersebut [5,6].

Selain itu, informasi masa lalu dapat digunakan untuk memeriksa konsistensi pengetahuan saat ini, selain verifikasi data eksperimen untuk menggantikan kemungkinan kesenjangan atau memasukkan informasi tambahan. Dalam beberapa tahun terakhir, beberapa profil ekspresi gen telah terbukti secara akurat membedakan subkategori subtype tumor yang berbeda. Para peneliti telah menemukan bahwa penggunaan data biologis dapat mencegah hasil sepihak dari pengujian terpisah dan tidak hanya berhasil mengurangi atau mengurangi gangguan pada chip gen. Namun demikian,

* Penulis yang sesuai. Departemen Ilmu Komputer, Sekolah Tinggi Seni dan Sains Rathnavel Subramaniam (RVS), Sullur, Coimbatore, India. *Alamat email:* yuvarajvelliangiri@gmail.com (V.Yuvaraj), maheswari@rvsgroup.com (D.Maheswari).

<https://doi.org/10.1016/j.measen.2023.100902>

Diterima 10 April 2023; Diterima dalam bentuk revisi 16 Agustus 2023; Diterima 27 September 2023

Tersedia online 9 Oktober 2023

2665-9174/© 2023 Penulis. Diterbitkan oleh Elsevier Ltd. Ini adalah artikel akses terbuka di bawah lisensi CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

pentingnya data kanker dari klasifikasi sebelumnya belum ditangani secara menyeluruh. Sistem jaringan saraf dapat berguna dalam mengatasi masalah yang disebutkan di atas [7].

Efektivitas ANN bergantung pada kapasitas pembelajarannya. Jaringan saraf yang populer adalah MLP (Multi-Layer Perceptron) yang merupakan mekanisme feed forward konvensional dan menggunakan pelatihan yang diawasi. Teknik terawasi terkenal lainnya selain MLP adalah algoritma berbasis gradien dan stokastik. Propagasi mundur dan variannya sering digunakan oleh para akademisi sebagai contoh standar metode berbasis gradien, meskipun ada beberapa kelemahan algoritma berbasis gradien, seperti kecenderungannya untuk hanya menemukan nilai minimum lokal, kelambatannya, dan meningkatnya ketergantungan pada nilai minimum lokal. parameter awal.

Untuk menghindari masalah ini, penelitian yang ada memperkenalkan model seleksi gen yang efektif yang bergantung pada IWOA. Selanjutnya, Jaringan MLSTM dipertimbangkan untuk klasifikasi kanker paru-paru. Namun, data masukan disajikan dalam berbagai skala dan hal ini dapat membuat atribut penting (dengan skala yang diperkecil) menjadi kurang efektif karena atribut lain memiliki nilai pada skala yang lebih tinggi dan LSTM rentan terhadap overfitting. Untuk mengatasi masalah tersebut dalam penelitian ini memperkenalkan model klasifikasi kanker paru yang disempurnakan. Dalam pekerjaan ini, awalnya dilakukan pra-pemrosesan data untuk normalisasi skala data menggunakan normalisasi min max. Seleksi gen dilakukan dengan menggunakan IWOA. Terakhir, ECNN (Enhanced convolution neural network) diterapkan untuk klasifikasi kanker paru-paru.

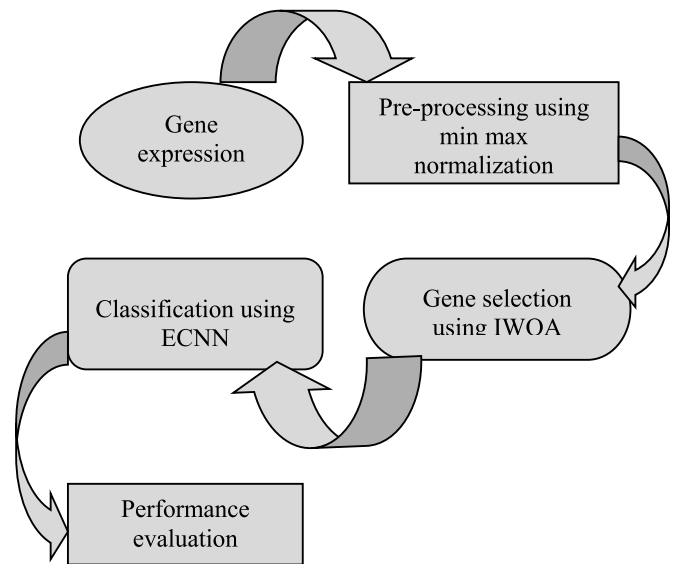
2. Tinjauan Pustaka

Arun Kumar dan Ramakrishnan [2018] [8] mengusulkan metrik kesamaan yang memilih karakteristik menggunakan pendekatan reduksi cepat kasar fuzzy. Pendekatan reduksi cepat kasar fuzzy yang disarankan menentukan ukuran kesamaan yang disesuaikan untuk mengidentifikasi jumlah gen bermanfaat terendah pada tahap awal. Entropi berdasarkan IG (Information Gain) membantu mengurangi dimensi pada tahap berikutnya dan selanjutnya pengklasifikasi RF (hutan acak) digunakan untuk memeriksa kumpulan data ekspresi gen untuk mengidentifikasi kanker paru-paru/ ovarium dan leukemia. Teknik yang disarankan berhasil mencapai akurasi klasifikasi masing-masing sebesar 99,45%/99,6% dan 97,22%. R, paket perangkat lunak sumber terbuka digunakan untuk mendemonstrasikan teknik di mana nilai metrik kinerja termasuk perolehan, presisi, akurasi, pengukuran-f, dan wilayah karakteristik menunjukkan hasil yang lebih baik dibandingkan dengan teknik klasifikasi yang ada.

Rafii dan Hassani [2015] [9] mengusulkan penggunaan MLP untuk mengklasifikasi kanker paru-paru dari data microarray. Studi ini menggunakan PCA untuk pertama-tama mengurangi dimensi data dan selanjutnya jaringan MLP (jaringan saraf feed-forward) mengklasifikasi instance. Penilaian mereka terhadap keakuratan klasifikasi pada kumpulan data ekspresi gen kanker paru-paru menunjukkan efektivitas teknik yang mereka usulkan. Studi ini juga menggambarkan nilai komparatif dari teknik lain dengan menggunakan kumpulan data yang sama.

Azzawi dkk. [2018] [10] mengusulkan peningkatan prediksi jaringan saraf MLP dengan menggunakan IMPSO (peningkatan Particle Swarm Optimization). Studi ini menghitung bobot/bias MLP menggunakan IMPSO-nya. Studi ini meningkatkan akurasi klasifikasi dalam memprediksi kanker paru-paru dari data ekspresi gen menggunakan skema diskriminasi yaitu MLP-IMPSO. Dengan menggunakan kumpulan data kanker paru-paru microarray yang sebenarnya, akurasi prediksi model yang direkomendasikan diperiksa secara menyeluruh dan dibandingkan dengan teknik pembelajaran mesin lainnya termasuk jaringan saraf berdasarkan fungsi basis radial, C4.5, SVM (mesin vektor pendukung), MLP, dan NB (Naive Bayes). Keandalan evaluasi mereka ditingkatkan dengan menggunakan validasi silang data. Studi ini menemukan peningkatan kinerja pendekatan yang mereka sarankan ketika informasi sebelumnya dipertimbangkan. Keberhasilan telah dibuktikan, menunjukkan bagaimana metode ini meningkatkan keakuratan diagnosis kanker paru-paru dengan menggunakan data biologis historis. Hasil penelitian memperjelas bahwa strategi yang disarankan berhasil dalam mengidentifikasi kanker paru-paru.

Salem dkk. [2015] [11] memperkenalkan metode unik untuk mengkategorikan penyakit kanker pada manusia berdasarkan karakteristik ekspresi gen. Metode yang disarankan menggabungkan DGA (Deep Genetic Algorithm) dengan IG



Gambar 1. Diagram blok khas dari model yang diusulkan.

dimana seleksi awal didasarkan pada IG diikuti dengan penggunaan GA (Algoritma Genetika) untuk mengurangi jumlah fitur. Pada tahap akhir, GP (genetic Programming) mengkategorikan jenis kanker. Studi tersebut mengevaluasi skema mereka bersama dengan teknik yang ada pada tujuh kumpulan data kanker yang berbeda untuk mendapatkan hasil yang memuaskan.

Li dkk. [2018] [12] mengusulkan klasifikasi data ekspresi gen kanker paru-paru menggunakan hukuman laso kelompok tumpang tindih yang jarang dan regresi multinomial adaptif. Studi ini menawarkan pendekatan pengelompokan tumpang tindih yang dapat dipahami secara biologis yang menekankan pentingnya kelompok gen dari kelas yang kurang terwakili. Signifikansi gen dalam kelompok dievaluasi dengan informasi timbal balik bersyarat, dan bobot berorientasi data. Regresi multinomial adaptif yang diatur diperkenalkan sesuai dengan mekanisme pengelompokan dan bobot yang dibangun, dan algoritma penyelesaian dirancang. Metode ini dapat digunakan untuk melakukan klasifikasi kelas jamak dengan memilih kelompok gen signifikan untuk setiap kelas, dan juga dapat membantu dalam seleksi adaptif gen signifikan dalam setiap kelompok. Dari hasil percobaan terbukti bahwa kinerja yang dicapai oleh skema tersebut jauh meningkat jika dibandingkan dengan 6 teknik lainnya dalam hal metrik kinerja dalam klasifikasi termasuk akurasi, dan gen terpilih merupakan gen yang bertanggung jawab atas kejadian kanker paru-paru.

Anand dan Bansal [2021] [13] menyarankan teknik untuk memilih subset gen terbaik yang sangat penting untuk prediksi yang lebih akurat. Repositori Kumpulan Data Biomedis Kent Ridge telah menyediakan data tentang ekspresi gen kanker paru-paru. Tujuan survei ini adalah untuk meningkatkan nilai AUC ROC dan F-measure sehingga model dapat dievaluasi dengan tepat saat bekerja dengan kumpulan data yang tidak seimbang.

Singh dan Susan [2022] [14] menganjurkan penggunaan pengklasifikasi FMM (Fuzzy Min-Max) untuk mengkategorikan data microarray ekspresi gen kanker paru-paru yang umumnya diabaikan karena dimensi dan biaya pemrosesan yang lebih tinggi. LASSO (Least Absolute Shrinkage and Selection Operator), meningkatkan akurasi dan kecepatan pengklasifikasi FMM dalam memilih subset gen terbaik untuk mengklasifikasi kanker paru-paru. Penelitian ini juga membandingkan kinerja FMM-LASSO dengan SVM, RF, KNN (K-nearest Neighbour), NB, dan LR (Logistic Regression) dengan dan tanpa LASSO. Hasil mereka menunjukkan bahwa FMM-LASSO mencapai hasil yang lebih baik dibandingkan teknik lainnya.

3. Metodologi yang diusulkan

Bagian ini mempelajari model yang diusulkan secara ekstensif. Model yang diusulkan terdiri dari tiga tahap, yang pertama adalah pra-pemrosesan menggunakan min max

normalisasi, tahap kedua adalah seleksi gen yang bergantung pada IWOA dan tahap ketiga adalah klasifikasi kanker paru-paru berbasis ECNN. Desain umum skema yang diusulkan digambarkan sebagai Gambar 1.

3.1. Pra-pemrosesan menggunakan normalisasi min-max

Normalisasi min-maks adalah teknik sederhana yang mengubah skala nilai data ke kisaran antara 0 dan 1, menggunakan nilai minimum dan maksimum dari data asli. Teknik ini mempertahankan urutan dan jarak relatif dari titik data, namun juga mengurangi varians dan memperbesar efek outlier.

Fitur input data gen mungkin hadir dengan skala berbeda yang dapat mengakibatkan hasil yang salah dan untuk mencegah masalah ini, skala data perlu dinormalisasi. Dalam penelitian ini, model Normalisasi Min -max digunakan dan proses normalisasi melibatkan penerjemahan nilai numerik ke rentang lain dengan menerapkan fungsi matematika. Dalam penelitian yang diusulkan ini, normalisasi min-maks berguna dalam normalisasi kumpulan data paru-paru [15,16]. Normalisasi min-maks adalah salah satu cara utama untuk normalisasi data. Normalisasi nilai yang ada dalam kumpulan data dilakukan dengan baik dalam rentang nilai minimum dan maksimum tertentu dari kumpulan data dan semua nilai diganti sesuai dengan ekspresi di bawah ini.

$$ay = \frac{v - \text{min}_A}{\text{maks}_A - \text{min}_A} (\text{maks}_B - \text{min}_B) + \text{min}_B \quad (1)$$

Di mana,

A - Data atribut,
Min(A), Max(A) menandakan nilai absolut minimum dan maksimum A.

ay menunjukkan nilai baru setiap entri data v menandakan nilai Lama setiap entri data. New_max(A), new_min(A) mewakili nilai maks dan min dari rentang (yaitu nilai batas rentang yang diperlukan) secara bersamaan.

3.2. Seleksi gen berdasarkan IWOA

Setelah normalisasi, penting untuk memilih gen terbaik karena hanya sebagian kecil dari masukan data ekspresi gen yang menunjukkan pola diferensial untuk kelompok atau sampel yang berbeda. IWOA yang digunakan dalam penelitian ini membantu dalam pemilihan gen yang tepat. Metode WOA didasarkan pada dinamika populasi dan dipengaruhi oleh alam. Mereka menggunakan teknik yang dikenal sebagai "perburuan jaring gelembung" untuk mencari mangsa dan meniru perilaku paus bungkuk. WOA menggunakan optimasi stokastik melalui agen pencariannya yang membantu menemukan solusi terhadap masalah optimasi. Tiga langkah dasar WOA meliputi menggiling mangsa, menyerang dengan jaring gelembung, dan mengawasi mangsa terbaik.

Ekspresi matematika inti dari WOA diberikan dalam Persamaan (2) dan (3).

$$X(t+1) = X(t) - A \cdot |C \cdot X(t) - X(t)| \text{ if } p < 0.5 \quad (2)$$

$$X(T+1) = CX(T) - X(T) \cdot e^{ik} \text{ karena } (2\pi\theta) + X(T) \text{ if } p \geq 0.5 \quad (3)$$

Nilai tetap yang menentukan bentuk spiral logaritmik, b, diinisialisasi ke 1 untuk reduksi linier vektor koefisien a, sehingga A = 2a. (ra); C = 2.r, dari 2 hingga 0 sepanjang iterasi yang merupakan solusi optimal, keluaran sebesar X*. X* mengacu pada vektor yang mewakili posisi paus. Nilai antara 0 dan 1 dialokasikan ke vektor acak, r. Saat lokasi paus diperbarui, nomor acak apa pun di antaranya - 1 dan 1 digunakan untuk beralih antar persamaan (2) dan (3); dalam persamaan (2) dan (3), probabilitasnya adalah 50% dan 50%, yang berarti paus memilih salah satu jalur secara acak dan dengan kecepatan yang sama selama tahap pengoptimalan. Selama fase jaring gelembung, vektor A dapat memiliki nilai acak antara [-1, 1], namun selama fase pencarian, vektor A juga dapat memiliki nilai di atas atau di bawah 1. Persamaan (4) menampilkan pencarian

teknik dalam tindakan

$$X(t+1) = X_{\text{acak}} - \text{SEBUAH} \cdot |C \cdot X_{\text{acak}} - X(t)| \quad (4)$$

Algoritma WOA melakukan pencarian global dengan menggunakan pendekatan pencarian acak ini. Pada awalnya, solusi acak dihasilkan ketika pencarian WOA dimulai. Selanjutnya, teknik ini digunakan untuk merevisi solusi ini secara berulang. Hingga jumlah iterasi tertinggi yang telah ditentukan tercapai, pencarian akan dilanjutkan.

3.2.1. Algoritma Pengoptimalan Paus yang Ditingkatkan

Algoritma pengoptimalan perlu memiliki trade-off yang tepat antara eksplorasi dan eksploitasi untuk memberikan hasil yang efektif. Oleh karena itu, ada pengurangan ukuran langkah seiring dengan peningkatan iterasi di WOA. Parameter A mengontrol ukuran langkah ini. Namun, ditemukan bahwa selama iterasi WOA, divergensi yang buruk membuat jebakan terbatas pada local optima.

Untuk mengatasi masalah ini, penelitian ini memanfaatkan IWOA. Fungsi retribusi terbang dapat digunakan untuk mengubah nilai A. Hasilnya, bakat eksplorasi dan eksploitasi WOA diperkuat secara bersamaan.

Ekspresi matematis dari distribusi Retribusi adalah seperti yang diberikan.

$$L(s, \gamma, \mu) = \begin{cases} \sqrt{\gamma} \left[\frac{\gamma}{2\pi} \frac{1}{\gamma^2 - \mu^2} \right] \frac{1}{(s - \mu)^{3/2}} & \text{jika } 0 < \mu < \infty \\ 0 & \text{jika } s \leq 0 \end{cases} \quad (5)$$

Di Sini μ, γ , Dan s mewakili parameter posisi, parameter skala yang mengatur skala distribusi dan kumpulan sampel.

3.2.2. Algoritma untuk meningkatkan optimasi paus

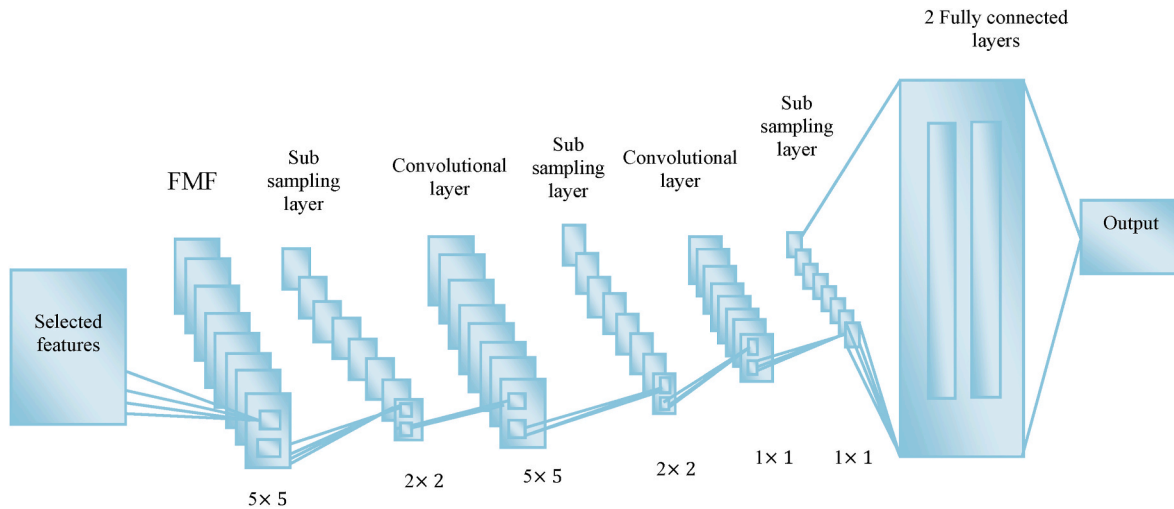
AWAL

1. Impor data
2. Inisialisasi posisi populasi paus X
3. menghitung kebugaran masing-masing paus
4. Inisialisasi a dan r, Hitung A dan C
5. Tentukan X* sebagai lokasi paus pemburu terbaik
6. Tetapkan T=1
7. sementara T ≤ iterasi maksimal dapat dilakukan
8. untuk setiap paus yang berburu, lakukan
9. jika $|A| < 0.5$
10. jika $|A| < 1$
11. perbarui lokasi permohonan paus berburu saat ini (2)
12. lain jika $|A| \geq 1$
13. pilih beberapa agen pencarian secara acak
14. perbarui posisi pelamar paus pemburu saat ini (4)
15. berakhir jika
16. lain jika $P \geq 0.5$
17. meningkatkan lokasi permohonan paus berburu yang ada (3)
18. berakhir jika
19. akhir untuk
20. tingkatkan X* jika ada solusi terbaik
21. T = T + 1
22. berakhir sementara
23. keluaran X*

AKHIR

3.3. Klasifikasi kanker paru menggunakan ECNN

Setelah pemilihan fitur selesai, penting untuk mengklasifikasikan fitur yang dipilih. Dalam penelitian ini, Improved CNN (Convolution Neural Networks) digunakan untuk prediksi kanker paru-paru [17,18]. CNN dan jaringan saraf tiruan umum memiliki struktur yang beragam. Berbeda dengan ANN tradisional yang menyebarkan masukan ke vektor, lapisan dalam CNN dipilih sedemikian rupa sehingga memiliki kecocokan spasial dengan data masukan. CNN terdiri dari satu atau lebih lapisan independen atau beberapa lapisan yang sepenuhnya terhubung, satu



Gambar 2. Arsitektur CNN.

atau lebih jumlah konvolusi, urutan lapisan sub-sampling, dan lapisan keluaran [19,20].

3.3.1. Kelemahan CNN tradisional

Dalam model CNN tradisional, kernel konvolusi digunakan untuk pemetaan fitur yang terkadang menyebabkan hilangnya informasi. Untuk mengatasi permasalahan tersebut, pada penelitian ini digunakan fungsi keanggotaan fuzzy pada lapisan masukan.

3.3.2. CNN yang ditingkatkan

CNN mencakup Konvolusi, sub-sampling, dan lapisan yang tertaut sepenuhnya. Diagram blok umum CNN ditunjukkan pada Gambar 2. Bagian berikut menawarkan ringkasan singkat dari setiap jenis lapisan.

3.3.3. Lapisan konvolusi

Dalam fase konvolusi ini, fitur masukan dan kernel (filter) digabungkan menjadi satu. Hasil konvolusi fitur masukan dan kernel digunakan untuk menghasilkan peta fitur keluaran dengan ukuran $i \times j$ dimana kernel yang termasuk dalam matriks konvolusi biasanya disebut filter.

Pada lapisan ini Fuzzy Membership Function (FMF) digunakan untuk pemetaan fitur yang dapat didefinisikan sebagai ($w_1 = 0.3, w_2 = 0.4, w_3 = 0.5, w_4 = 0.7$) dan akan dihitung sebagai berikut

$$H_{aku} = \frac{1}{M} \sum_{j=1}^M \mu_{aku}(A_{2j}) \quad (6)$$

Di mana $\mu_{aku}(A_{2j})$ adalah fungsi keanggotaan $\mu_{aku}(A_{2j}) : R \rightarrow [0,1]$, $aku = 1, 2, \dots, M$, $j = 1, 2, \dots, N$. Menggunakan fungsi keanggotaan Gaussian.

Output yang diperoleh dari lapisan konvolusi ke- i , direpresentasikan sebagai

$$C_{aku}^{(i)} = \sum_{j=1}^N B_{aku}^{(i)} K_{(i-1)}^{(j)} * C_{aku}^{(i-1)} \quad (7)$$

Di mana, $B_{aku}^{(i)}$ menunjukkan matriks bias dan $K_{(i-1)}^{(j)}$ singkatan dari konvolusi filter atau kernel memiliki ukuran $A \times A$ yang menyediakan koneksi antara j -peta fitur di lapisan $(aku-1)$ dan $Saya$ -Peta fitur ada di lapisan yang sama.

Hasil $C_{aku}^{(i)}$ lapisan menyertakan peta fitur. Dalam (10), yang pertama lapisan konvolusi $C_{aku}^{(i)}$ membentuk ruang input, yaitu, $C_{(0)}^{(aku)} = X_{Saya}$. Peta fitur dihasilkan oleh kernel. Setelah konvolusi, fungsi aktivasi digunakan dalam lapisan untuk konversi non-linier dari keluaran lapisan konvolusi.

$$Y_{aku}^{(aku)} = YC_{aku}^{(aku)} \quad (8)$$

Di mana, $Y_{aku}^{(aku)}$ mengacu pada output yang dihasilkan fungsi aktivasi dan $C_{aku}^{(aku)}$ menunjukkan masukan yang diterimanya.

3.3.4. Sub sampling atau lapisan pengumpulan

Lapisan konvolusi sebelumnya membantu mengekstraksi tujuan utama lapisan ini, yaitu meminimalkan dimensi peta fitur secara geografis. Peta fitur digunakan sebagai mask, dan fungsi sub sampling digunakan di antara keduanya. Di antara teknik subsampling baru yang diperkenalkan adalah pengumpulan rata-rata, pengumpulan jumlah, dan pengumpulan maksimum. Penggabungan maksimal, yang menghasilkan fitur keluaran yang cocok dari nilai terbesar di setiap blok, merupakan penggabungan dominan. Karena memungkinkan lapisan konvolusi menahan rotasi dan konversi antar masukan, lapisan subsampling harus dipertimbangkan.

3.3.5. Lapisan yang terhubung sepenuhnya

Lapisan terakhir CNN adalah jaringan feed forward konvensional yang memiliki satu atau banyak lapisan tersembunyi. Di lapisan keluaran, fungsi aktivasi SoftMax digunakan:

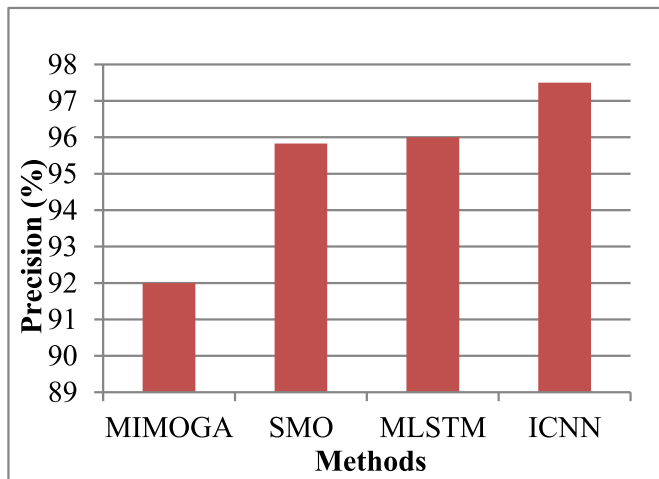
$$Y_{aku}^{(aku)} = fZ_{aku}^{(aku)} \quad (9)$$

$$\text{Dimana } Z_{aku}^{(aku)} = \sum_{saya=1}^{M_{saya}^{(aku)}} W_{aku}^{(aku)} \quad (10)$$

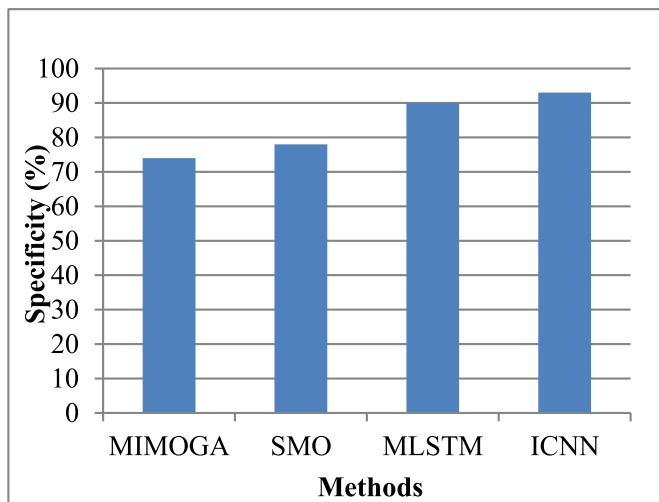
Di mana, $W_{aku}^{(aku)}$ menunjukkan bobot seluruh lapisan yang terhubung sepenuhnya harus menyesuaikan agar setiap kelas dapat terwakili dan f menunjukkan fungsi transfer yang menetapkannya sebagai nonlinier. Perlu dicatat bahwa sepenuhnya terhubung neuron lapisan menyertakan properti nonlinier, bukan lapisan tunggal seperti lapisan konvolusi dan penggabungan.

4. Hasil dan pembahasan

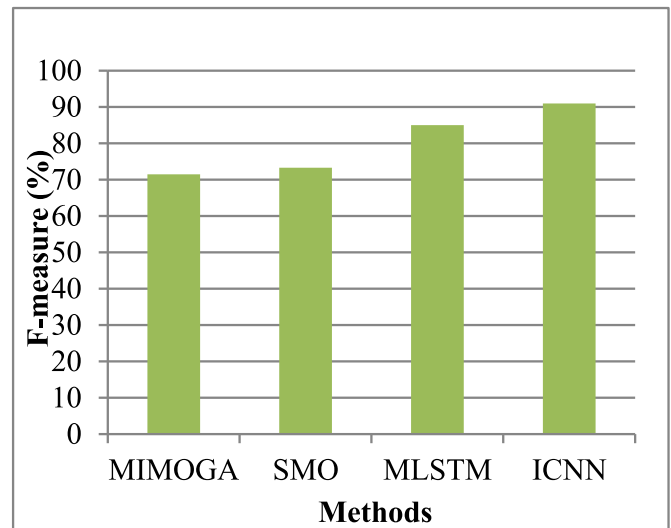
Di sini, hasil eksperimen dari pengujian model yang disarankan yang dilakukan pada Matlab 2013b. Dengan menggunakan Kent Ridge Bio-Medical Dataset, kerangka kerja ICNN yang diusulkan dalam penelitian ini dibandingkan dengan model MIMOGA, SMO, dan MLSTM yang sudah ada sebelumnya berdasarkan metrik seperti recall, presisi, akurasi, dan f-measure. Repositori Kumpulan Data Bio-Medis Kent Ridge menawarkan data tentang ekspresi gen kanker paru-paru. Basis data ini berisi informasi ekspresi gen yang dikumpulkan



Gambar 3. Perbandingan hasil presisi.



Gambar 4. Perbandingan hasil kekhususan.



Gambar 5. Perbandingan hasil pengukuran F.

4.2. Kekhususan

Kekhususan diberikan oleh rasio negatif nyata, yang prediksinya dilakukan sebagai negatif (atau negatif sebenarnya). Itu dihitung seperti di bawah ini:

$$\text{Kekhususan} = \frac{\text{Negatif Sejati}}{\text{Negatif sejati} + \text{Positif palsu}} \quad (12)$$

Gambar 4 mengilustrasikan hasil yang berkaitan dengan analisis perbandingan kinerja untuk ICNN yang diusulkan dan metode MIMAGA, SMO, dan MLSTM terbaru dalam hal Kekhususan. Sesuai grafik di atas, berbagai metodologi klasifikasi dicantumkan di sepanjang sumbu x dan sumbu y mewakili nilai Spesifisitas. Pekerjaan yang diusulkan menggunakan model optimasi paus untuk seleksi gen di mana parameter ukuran langkah akan dihitung berdasarkan fungsi penerbangan retribusi dan meningkatkan tingkat positif sebenarnya yang secara otomatis meningkatkan Spesifisitas. Dapat disimpulkan dari gambar di atas bahwa algoritma yang diusulkan menunjukkan peningkatan nilai Spesifisitas yaitu 93% sedangkan algoritma metode MIMAGA, SMO dan MLSTM yang ada memberikan hasil Spesifisitas yang lebih rendah yaitu 74%, 78% dan 90% (lihat Gambar 5).

4.3. F-ukuran

Skor-F menghitung rata-rata harmonik dari nilai presisi dan perolehan kembali suatu sistem. F1-Skor dihitung sebagai:

$$F1 - \text{skor} = \frac{2 \times \text{presisi} \times \text{mengingat}}{\text{presisi} + \text{penarikan kembali}} \quad (13)$$

Hasil perbandingan kinerja metode ICNN yang diusulkan dan metode MIMAGA, SMO dan MLSTM yang ada ditinjau dari f-measure ditunjukkan pada gambar di atas. Pada grafik di atas, beberapa metodologi klasifikasi dicantumkan di sepanjang sumbu x dan sumbu y mewakili nilai ukuran-f. Hal ini ditunjukkan dari gambar di atas bahwa algoritma ICNN yang diusulkan menghasilkan nilai f-measure yang lebih baik yaitu 91% sedangkan metode MIMAGA, SMO dan MLSTM yang ada menghasilkan nilai f-measure sebesar 71,5%, 73,3% dan 85%.

4.4. Ketepatan

Akurasi mengukur sejauh mana teknik tersebut benar secara keseluruhan dan dihitung dengan rasio antara total parameter klasifikasi asli (True positif + Benar negatif) dan jumlah parameter klasifikasi (benar positif + benar negatif + positif palsu + negatif palsu.).

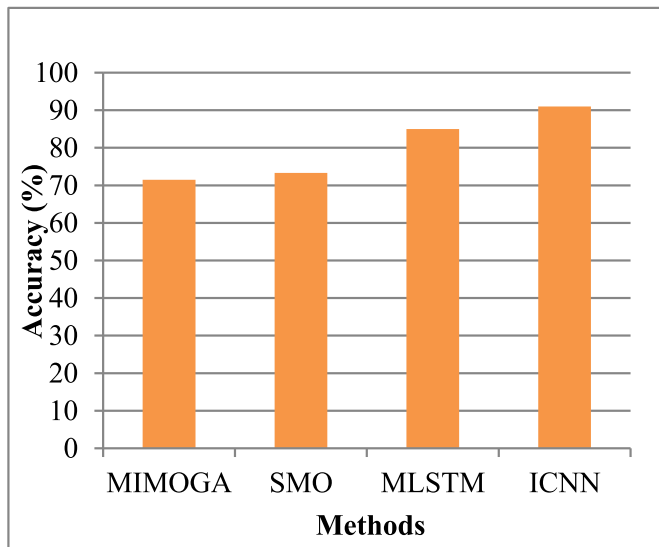
dari 86 sampel adenokarsinoma paru primer dan 10 sampel paru non-neoplastik. 7129 gen hadir di setiap sampel. Kumpulan data lengkap digunakan untuk menghasilkan 30% set pengujian dan 70% set pelatihan. Hasilnya divalidasi menggunakan kumpulan data pelatihan dan pengujian.

4.1. Presisi

Presisi adalah ukuran keakuratan atau kualitas. Algoritma yang memiliki presisi lebih baik biasanya memiliki relevansi yang lebih tinggi dibandingkan algoritma yang memiliki presisi lebih rendah. Misalnya, presisi dapat dihitung dengan rasio jumlah positif sebenarnya dan jumlah totalnya dari positif. Itu dihitung sebagai,

$$\text{Presisi} = \frac{\text{Benar-benar positif}}{\text{Positif benar} + \text{Positif palsu}} \quad (11)$$

Di atas Gambar 3, menggambarkan evaluasi perbandingan pendekatan yang diusulkan dan teknik terkini dalam hal presisi. Metodologi tercantum dalam sumbu x dan nilai presisi diambil alih sumbu y. Pada gambar di atas, metode ICNN yang diusulkan adalah dibandingkan dengan pendekatan MIMAGA, SMO dan MLSTM yang ada. Ini saya simpulkan dari hasil usulan kerangka kerja ICNN ini menghasilkan lebih tinggi A, SMO dan Metode MLSTM menghasilkan masing-masing 92%, 95,83%, 96%.



Gambar 6. Perbandingan hasil akurasi.

$$\text{Akurasi} = \left(\frac{T_P + T_N}{T_P + T_N + F_P + F_N} \right) \quad (14)$$

Perbandingan metrik akurasi untuk pengklasifikasi seperti MIMOGA, SMO dan MLSTM ditunjukkan pada Gambar 6. Sumbu x mencantumkan metodologi dan nilai akurasi diambil sepanjang sumbu y. Dapat dipastikan dari gambar di atas bahwa algoritma ICNN yang diusulkan memberikan hasil peningkatan akurasi sebesar 93% sedangkan metode MIMOGA, SMO dan MLSTM yang ada memberikan hasil akurasi yang lebih rendah yaitu masing-masing 77,77%, 85,33% dan 90,23%.

5. Kesimpulan dan pekerjaan masa depan

Kanker paru-paru adalah salah satu jenis kanker yang paling umum di seluruh dunia dan menjadi penyebab lebih dari satu juta kematian setiap tahunnya. Mengidentifikasi gen yang mewakili secara berbeda dan berbagai pola ekspresi gen dapat menjelaskan interpretasi subjek normal dan abnormal pada tingkat transkriptomik. Dalam penelitian ini, pra-pemrosesan dihitung menggunakan min-max untuk normalisasi skala data. Seleksi gen dilakukan dengan menggunakan IWOA. ECNN diterapkan untuk klasifikasi kanker paru-paru. Hasil eksperimen menunjukkan bahwa model yang diusulkan ini menghasilkan hasil akurasi sebesar 93% yang meningkat jika dibandingkan dengan model lain yang ada saat ini. Namun model yang diusulkan tidak dapat diterapkan pada penyakit lain dan hal ini dapat difokuskan pada masa mendatang.

Pernyataan pendanaan

Penulis penelitian ini tidak mendapatkan dana khusus.

Deklarasi kepentingan bersaing

Para penulis menyatakan bahwa mereka tidak memiliki pesaing yang diketahui yang keuangan kepentingan atau hubungan pribadi yang bisa saja muncul pengaruh

pekerjaan yang dilaporkan dalam makalah ini.

Ketersediaan data

<https://www.kaggle.com/datasets/yusufdede/lung-cancer-dataset>

Referensi

- [1] SA Ludwig, S. Picek, D. Jakobovic, Klasifikasi data kanker: menganalisis data ekspresi gen menggunakan algoritma pohon keputusan fuzzy, dalam: Aplikasi Riset Operasi dalam Manajemen Perawatan Kesehatan, Springer, Cham, 2018, hlm..
- [2] P. Chaudhari, H. Agrawal, K. Kotecha, Augmentasi data menggunakan MG-GAN untuk meningkatkan klasifikasi kanker pada data ekspresi gen, *Soft Comput.* 24 (15) (2020) 11381–11391.
- [3] L. Sun, Z. Xiao-Yu, Q. Yu-Hua, JC Xu, SG Zhang, Y. Tian, Metode seleksi gen berbasis entropi lingkungan gabungan dengan skor Fisher untuk klasifikasi tumor, *Appl. Intel.* 49 (4) (2019) 1245–1259.
- [4] R. Alanni, J. Hou, H. Azzawi, Y. Xiang, Algoritma pemilihan gen baru untuk klasifikasi kanker menggunakan kumpulan data microarray, *BMC Med. Genom.* 12 (1) (2019) 1–12.
- [5] MM Babu, Pengantar analisis data microarray, dalam: *Computational Genomics: Theory and Application*, 2004, hlm.225–249.
- [6] A Alzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, C. Thaventhiran, Peningkatan klasifikasi ansambel jaringan saraf untuk diagnosis penyakit kanker paru-paru, *Appl. Komputasi Lembut.* 80 (2019) 579–591.
- [7] H. Lu, H. Gao, M. Ye, X. Wang, Algoritma ansambel hibrid yang menggabungkan AdaBoost dan algoritma genetika untuk klasifikasi kanker dengan data ekspresi gen, *IEEE ACM Trans. Hitung. biologi. Bioinf* 18 (3) (2019) 863–870.
- [8] C. Arunkumar, S. Ramakrishnan, Pemilihan atribut menggunakan ukuran kesamaan khusus berbasis fuzzy roughset untuk data ekspresi gen microarray kanker paru-paru, *Jurnal Komputasi dan Informatika Masa Depan* 3 (1) (2018) 131–142.
- [9] F. Rafii, BDR Hassani, Jaringan MLP untuk prediksi kehadiran kanker paru-paru berdasarkan data microarray, dalam: *Konferensi Dunia Ketiga tentang Sistem Kompleks (WCCS) 2015*, IEEE, 2015, November, hlm..
- [10] H. Azzawi, J. Hou, R. Alanni, Y. Xiang, Pendekatan jaringan saraf hibrid untuk klasifikasi kanker paru-paru dengan kumpulan data ekspresi gen dan pengetahuan biologis sebelumnya, dalam: *Konferensi Internasional tentang Pembelajaran Mesin untuk Jaringan*, Springer, Cham, 2018, November, hlm.279–293.
- [11] H. Salem, G. Attiya, N. El-Fishawy, Profil ekspresi gen berdasarkan klasifikasi penyakit kanker manusia, dalam: *Teknik Komputer Internasional ke-11 tahun 2015 Konferensi (ICENCO)*, IEEE, 2015, Desember, hlm. 181–187.
- [12] J. Li, Y. Wang, X. Song, H. Xiao, Regresi multinomial adaptif dengan kelompok yang tumpang tindih untuk klasifikasi kelas jamak kanker paru-paru, *Comput. biologi. medis.* 100 (2018) 1–9.
- [13] M. Anand, N. Bansal, Prediksi kanker paru-paru menggunakan teknik pembelajaran mesin dengan pendekatan statistik untuk seleksi gen, dalam: *Soft Computing and Signal Processing*, Springer, Singapura, 2021, hlm.273–284.
- [14] Y. Singh, S. Susan, Seleksi gen optimal dan klasifikasi data microarray menggunakan jaringan saraf fuzzy min-max dengan LASSO, dalam: *Konferensi Internasional tentang Sistem Cerdas dan Fuzzy*, Springer, Cham, 2022, hlm. 777–784.
- [15] VG Raju, KP Lakshmi, VM Jain, A. Kalidindi, V. Padma, Mempelajari pengaruh proses normalisasi/transformasi terhadap keakuratan klasifikasi yang diawasi, dalam: *Konferensi Internasional Ketiga tentang Sistem Cerdas dan Inventif 2020 Teknologi (ICSSIT)*, IEEE, 2020, Agustus, hlm.729–735.
- [16] J. Wang, G. Wen, S. Yang, Y. Liu, Estimasi masa manfaat yang tersisa dalam prognostik menggunakan jaringan saraf LSTM dua arah yang dalam, dalam: *Konferensi Prognostik dan Manajemen Kesehatan Sistem 2018 (PHM-Chongqing)*, IEEE, 2018, Oktober, hal.1037–1042.
- [17] MA Albahar, Klasifikasi lesi kulit menggunakan jaringan saraf konvolusi dengan pengatur baru, *IEEE Access* 7 (2019) 38306–38313.
- [18] A. Jafari, A. Ganesan, CSK Thalisetty, V. Sivasubramanian, T. Oates, T. Mohsenin, Sensorsnet: jaringan saraf konvolusi dalam yang skalabel dan berdaya rendah untuk klasifikasi data multimodal, *IEEE Transactions on Circuits and Systems I: Regular Papers* 66 (1) (2018) 274–287.
- [19] W. Zhang, K. Ma, J. Yan, D. Deng, Z. Wang, Penilaian kualitas gambar buta menggunakan jaringan saraf konvolusi bilinear dalam, *IEEE Trans. Lingkaran. sistem. Teknologi Video.* 30 (1) (2018) 36–47.
- [20] Y. Liu, Y. Zhong, Q. Qin, Klasifikasi pemandangan berdasarkan jaringan saraf konvolusi multiskala, *IEEE Trans. Geosains. Rem. Sens.* 56 (12) (2018) 7109–7121.