

Optimisasi Parameter VGGNet melalui Bayesian Optimization untuk Klasifikasi Nodul Paru

Kadek Eka Sapta Wijaya¹⁾, Gede Angga Pradipta²⁾, Dadang Hermawan³⁾

Program Studi Magister Sistem Informasi
Institut Teknologi dan Bisnis STIKOM Bali
Denpasar, Indonesia

e-mail: 222012029@stikom-bali.ac.id¹⁾, angga_pradipta@stikom-bali.ac.id²⁾,
dadang@stikom-bali.ac.id³⁾

Abstrak

Kanker paru menjadi fokus utama dalam kesehatan masyarakat global karena tingkat kematian yang signifikan. Untuk menganalisis kanker paru, pemindaian radiografi seperti X-ray, MRI, atau CT scan umumnya digunakan. Pendekatan untuk mendeteksi nodul paru dapat dikelompokkan menjadi dua kategori utama, yaitu: metode *machine learning* dan berbasis *deep learning*. Metode *machine learning* memerlukan ekstraksi fitur manual dari nodul, sedangkan pendekatan berbasis *deep learning*, terutama dengan menggunakan *Convolutional Neural Networks (CNN)*. Metode CNN semakin populer karena kemampuannya untuk belajar fitur-fitur yang relevan secara otomatis dari data. Membangun CNN melibatkan penyesuaian konfigurasi yang tidak dapat ditentukan secara otomatis dari data. Hasil dari pencarian Bayesian Optimization dapat membantu dalam menentukan hiperparameter yang optimal untuk jaringan CNN. Namun, penelitian ini menemukan bahwa menerapkan parameter optimal pada VGGNet untuk klasifikasi multi kelas menghasilkan kecenderungan *overfitting*. Meskipun akurasi model saat pelatihan mencapai 96,8%, akurasi pada data uji hanya mencapai 35,40%, menunjukkan bahwa model cenderung tidak generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

Kata kunci: *Hyperparameter, Bayesian Optimization, Deep Learning, VGGNet.*

1. Pendahuluan

Kanker telah menjadi fokus utama dalam domain kesehatan masyarakat dan merupakan penyebab kematian yang signifikan secara global. Penyakit ini memiliki potensi untuk muncul di berbagai bagian tubuh dan dapat mengenai individu dari beragam rentang usia. Dikenal adanya lebih dari 100 jenis kanker yang berbeda, tiap-tiap jenis memiliki ciri khas dan perilaku yang unik. Di antara variasi tersebut, kanker paru-paru, payudara, kolorektal, hati, dan perut tercatat sebagai yang paling umum terjadi. Paru-paru adalah organ berpasangan yang terletak di dada yang bertanggung jawab atas proses respirasi, yakni pertukaran gas oksigen dan karbon dioksida dalam tubuh. Data pada tahun 2021 mengungkapkan bahwa sekitar 1,80 juta individu telah meninggal akibat kanker paru-paru [1].

Pada umumnya, kanker paru dianalisis melalui pemindaian radiografi seperti X-ray, MRI, atau CT scan. Keberhasilan utama dari proses ini bergantung pada keahlian radiologis dalam mengidentifikasi lesi yang mencurigakan di paru-paru, termasuk nodul kecil. Tugas ini menjadi krusial dan menantang, terutama dalam mendeteksi nodul kecil. Menurut literatur, sekitar 68% nodul paru dapat diidentifikasi dengan tepat oleh seorang radiologis, namun angka tersebut dapat ditingkatkan menjadi 82% dengan melibatkan pendapat dari radiologis lainnya [2].

Pendekatan untuk mendeteksi nodul paru dapat dikelompokkan menjadi dua kategori utama: metode *machine learning* dan berbasis *deep learning*. Metode tradisional seperti *thresholding* dan pengelompokan memerlukan ekstraksi fitur secara manual dari nodul. Di sisi lain, pendekatan berbasis *deep learning*, terutama dengan menggunakan *Convolutional Neural Networks (CNN)*, mampu secara otomatis mempelajari fitur-fitur nodul paru dari gambar input dengan kekuatan ekstraksi fitur yang signifikan. Pendekatan ini menjanjikan tingkat ketahanan dan generalisasi yang lebih baik dalam deteksi nodul medis [3]. Membangun *Convolutional Neural Network (CNN)* melibatkan penyesuaian sejumlah konfigurasi yang tidak dapat ditentukan secara otomatis dari data, dan harus diatur secara manual oleh peneliti di bidang *machine learning*. Faktor-faktor seperti struktur jaringan dan parameter yang diterapkan dalam pelatihan CNN dikenal sebagai hiperparameter [4].

Hiperparameter merupakan nilai-nilai yang harus ditentukan secara manual sebelum fase pelatihan model. Proses penyetapan hiperparameter, yang dikenal sebagai optimisasi, memiliki peranan yang sangat

penting dalam upaya meningkatkan kinerja algoritma pembelajaran mesin. Meskipun hubungan antara hiperparameter dan performa model tidak selalu transparan, optimisasi hiperparameter melibatkan pencarian set nilai yang dapat menghasilkan kinerja terbaik. Proses ini mencakup eksperimen dengan berbagai kombinasi nilai hiperparameter serta melatih beberapa model untuk memilih yang paling optimal [5]. Oleh karena itu, diperlukan optimisasi hiperparameter pada *Convolutional Neural Network* (CNN).

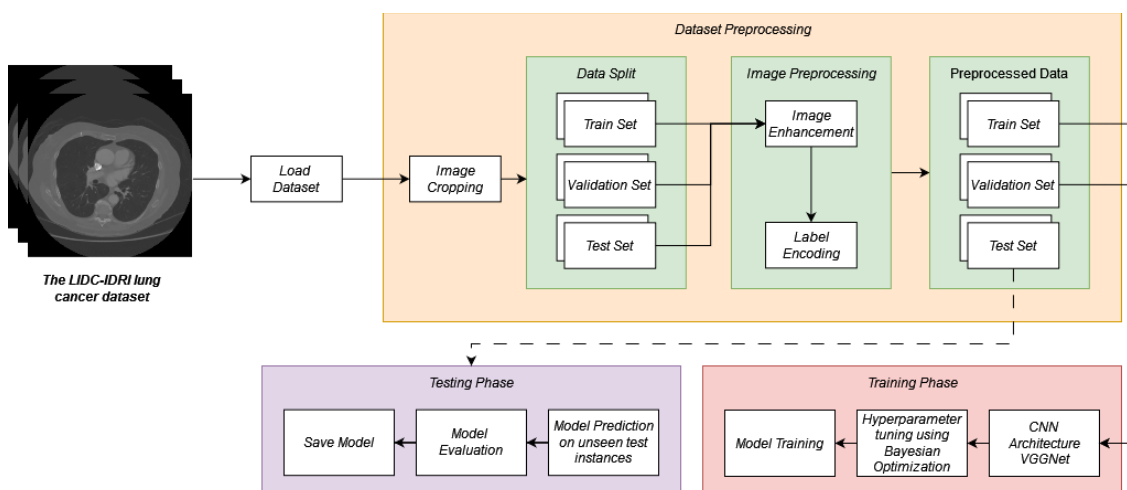
Menurut beberapa penelitian yang telah dilakukan sebelumnya oleh Zheng M, dkk. [6] mengusulkan penggunaan *Multi-Level Convolutional Neural Network* (ML-CNN) untuk klasifikasi nodul paru-paru, dengan konfigurasi hiperparameter yang ditentukan melalui *Bayesian Optimization*. Pendekatan ini memanfaatkan model *surrogate* probabilistik seperti *Gaussian processes* untuk mengoptimalkan fungsi kesalahan validasi dari konfigurasi hiperparameter. Hasil eksperimen menunjukkan bahwa algoritma yang diusulkan mengungguli metode penyetalan manual serta metode optimisasi hiperparameter lainnya seperti *Random Search*, *Gaussian Processes* dengan kernel stasioner, dan metode *Hyperparameter Optimization* via RBF and *Dynamic Coordinate Search* (HORD). Adapun penelitian lainnya yang dilakukan oleh Hanifi S, dkk. [7] melakukan perbandingan antara teknik optimisasi hiperparameter seperti *scikit-opt*, *optuna*, dan *hyperopt* pada model *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory Network* (LSTM) untuk memprediksi kekuatan angin. Evaluasi dilakukan dengan memanfaatkan *Root Mean Square Error* (RMSE) dan waktu komputasi sebagai parameter. Hasil eksperimen menunjukkan bahwa *Optuna* dengan metode *Tree-structured Parzen Estimator* (TPE) dan *Expected Improvement* (EI) menunjukkan efisiensi terbaik untuk kedua model CNN dan LSTM. Meskipun demikian, dalam hal akurasi, model CNN menunjukkan hasil yang serupa, sedangkan pada model LSTM, teknik *hyperopt* memberikan tingkat akurasi tertinggi.

Bayesian Optimization merupakan pendekatan terkemuka dalam *tuning* hiperparameter, menawarkan strategi global tanpa gradien untuk mengidentifikasi hiperparameter optimal dengan iterasi yang minim. Metode ini efektif untuk menangani fungsi objektif kompleks dan situasi di mana fungsi target sulit dibedakan terhadap variabel independen. Prosesnya memecah masalah menjadi sub-masalah, membangun model distribusi parameter menggunakan proses Gaussian, dan mencari solusi optimal melalui optimisasi pengganti. Dengan meningkatkan jumlah sampel, model semakin mendekati solusi optimal [8].

Berdasarkan uraian tersebut, metode *Bayesian Optimization* (BO) digunakan untuk secara otomatis menemukan parameter optimal pada CNN. BO bertugas mengidentifikasi kombinasi parameter terbaik dengan tujuan mengurangi waktu yang diperlukan untuk mencapai model yang akurat.

2. Metode Penelitian

Bagian ini akan difokuskan pada proses perancangan dan implementasi model yang akan dipakai untuk mencapai tujuan penelitian. Langkah-langkah metodologi yang akan diuraikan meliputi pemilihan dan pengolahan dataset, desain dan konfigurasi model, serta evaluasi performa model, dapat dilihat pada Gambar 1.



Gambar 1. Gambaran Alur Sistem

2.1 Analisis kebutuhan perangkat keras dan perangkat lunak

Menganalisis kebutuhan perangkat keras dan perangkat lunak menjadi langkah penting untuk menjamin konsistensi eksperimen. Konfigurasi lingkungan yang akan diterapkan guna memastikan stabilitas pelaksanaan dan optimalisasi pemanfaatan sumber daya komputasi, dapat dilihat pada Tabel 1.

Tabel 1. Platform Eksperimen

Nama Item	Spesifikasi
CPU	16 vCPU
RAM	167 GB
GPU	RTX A6000
DISK	30 GB
Language	Python
IDE	JupyterLab

2.2 LIDC-IDRI Dataset

Dataset LIDC-IDRI menyediakan 244.527 gambar CT scan toraks dari 1.010 kasus yang dapat diakses publik. Setiap gambar dilengkapi dengan informasi koordinat dan batas nodul yang tersedia dalam file anotasi XML, bersama dengan penilaian diagnostik dari empat ahli radiologi toraks yang berpengalaman. Penilaian ahli radiologi mencakup tingkat keganasan nodul paru, yang diklasifikasikan dalam lima kategori berbeda. Data ini digunakan untuk menentukan anotasi yang sesuai dengan nodul terkait berdasarkan file XML dan daftar anotasi yang diberikan [2].

Dataset dalam penelitian ini akan dikelompokkan menjadi tiga kategori: *benign*, *malignant*, dan *normal*. Pemindaian CT awalnya tersedia dalam format DICOM dengan resolusi 512×512 , namun kemudian diubah menjadi format JPG untuk mempermudah akses.

2.3 Dataset Preprocessing

Sebelum memasuki tahap pra-pemrosesan, dilakukan *filter* dengan mengecualikan penilaian 3, yang menunjukkan probabilitas menengah antara nodul yang bersifat jinak atau ganas. Tahap pra-pemrosesan terdiri dari dua langkah utama. Pertama, gambar-gambar diubah menjadi ukuran 256×256 piksel. Selanjutnya, ada tiga proses lanjutan, yaitu data split, pra-pemrosesan gambar, dan data yang telah diproses sebelumnya.

1. Data Split

Pada tahap ini, data dipisahkan menjadi tiga bagian yang berbeda, dengan 60% digunakan untuk pelatihan model, 20% untuk validasi, dan 20% untuk pengujian.

2. Image Preprocessing

Pada tahap ini, gambar-gambar akan ditingkatkan kualitasnya dengan menerapkan *filter Gaussian* dan *Gabor*. Setelah proses peningkatan gambar selesai, kelas-kelas dalam dataset akan diubah menjadi representasi numerik menggunakan teknik *encoding label*, di mana kelas *benign* akan diwakili sebagai 0, kelas *malignant* sebagai 1, dan kelas *normal* sebagai 2.

3. Preprocessed Data

Pada tahap ini, akan menghasilkan tiga set data yang telah di proses dari tahapan *data split* dan *image preprocessing*. Set pertama terdiri dari data pelatihan yang akan digunakan sebagai proses pelatihan, set kedua terdiri dari data validasi yang akan digunakan sebagai optimalisasi model, dan terakhir yaitu set ketiga terdiri dari data pengujian yang digunakan untuk menilai performa model.

2.4 Training Phase

Pada tahapan *training phase* ini dilakukan penerapan model *Convolutional Neural Network* (CNN) dengan menggunakan arsitektur VGGNet. VGG *Network* terdiri dari beberapa varian, termasuk tiga model CNN (VGG-F, VGG-M, dan VGG-S), serta jaringan dalam VGG-VD. VGG-VD terbagi menjadi VGG-16, yang terdiri dari 13 lapisan konvolusi dan 3 lapisan Fully Connected (FC), dan VGG-19, yang memiliki 16 lapisan konvolusi dan 3 lapisan FC [9]. Setelah menerapkan model CNN dengan arsitektur VGGNet, selanjutnya dilakukan penyeteraan hiperparameter dengan menggunakan teknik *Bayesian Optimization* (BO). Setelah melakukan proses optimasi hiperparameter, selanjutnya dilakukan proses pelatihan yang melibatkan dataset pelatihan, validasi, dan pengujian.

2.5 Testing Phase

Pada tahap pengujian ini, model digunakan untuk memprediksi data uji yang belum pernah dilihat sebelumnya. Data uji berperan sebagai ujian akhir untuk menguji kemampuan model dalam mengklasifikasikan nodul paru yang baru. Evaluasi kinerja model dilakukan menggunakan metrik seperti *Accuracy*, yang merupakan proporsi dari jumlah sampel yang diklasifikasikan dengan tepat oleh model terhadap total jumlah sampel dalam dataset evaluasi [10]. *Precision* menggambarkan proporsi dari sampel yang diklasifikasikan dengan benar terhadap total sampel yang dinyatakan sebagai positif oleh model [10]. *Recall*, yang juga dikenal sebagai sensitivitas atau *True Positive Rate (TPR)*, mengindikasikan proporsi dari sampel positif yang berhasil diklasifikasikan dengan benar oleh model [10]. *F1-Score* merupakan rata-rata harmonis dari presisi dan recall, menunjukkan bahwa skor ini memberikan penalti terhadap nilai yang ekstrem dari keduanya [10]. Setelah evaluasi selesai dan model terbukti memiliki performa yang baik, langkah terakhir adalah menyimpan model.

3. Hasil dan Pembahasan

Bagian ini menguraikan hasil dari penelitian yang telah dilaksanakan dan akan dijabarkan dengan detail proses *preprocessing*, fase pelatihan, dan pengujian yang telah ditempuh dalam penelitian ini.

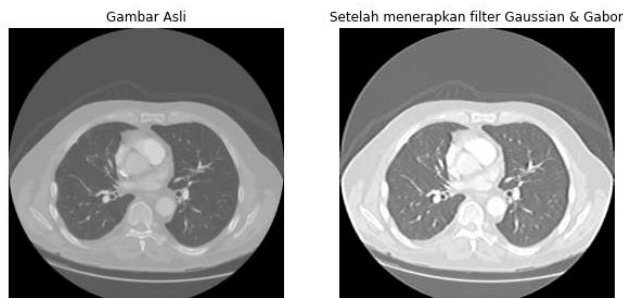
3.1 Preprocessing

Pada tahap ini, dilakukan tahap *preprocessing* sebelum diterapkan pada model *training*, *validation* dan *testing*. Semua gambar dilakukan *cropping* ke resolusi $256 \times 256 \times 1$ untuk menyesuaikan input shape dari *tensor* yang ada pada model. Kemudian gambar akan diacak terlebih dahulu untuk memastikan pelatihan yang tidak bias, lalu dibagi menjadi perbandingan 60:20:20 antara data *train*, *validation*, dan data *test*. Dataset pelatihan akan terdiri dari 60% gambar, yang akan digunakan untuk melatih model. Sebanyak 20% gambar akan dialokasikan sebagai dataset validasi untuk memonitor kinerja model selama pelatihan. Sementara itu, 20% sisanya akan disimpan sebagai dataset pengujian, digunakan untuk menguji model terhadap data yang tidak pernah dilihat sebelumnya. Distribusi pembagian data dapat dilihat pada Tabel 2.

Tabel 2. Distribusi kelas setelah data dibagi dengan rasio 60:20:20

Kelas	Split	Gambar	Total
<i>Benign</i>	<i>Train</i>	1964	9244
<i>Malignant</i>		3149	
Normal		4131	
<i>Benign</i>	<i>Validation</i>	655	3082
<i>Malignant</i>		1050	
Normal		1377	
<i>Benign</i>	<i>Test</i>	655	3082
<i>Malignant</i>		1050	
Normal		1377	

Setelah pembagian set data, langkah selanjutnya adalah melakukan *enhancement* pada gambar menggunakan *filter Gaussian* dan *Gabor*. Hasil dari proses *filter* ini dapat dilihat pada Gambar 2.



Gambar 2. Hasil *Image Enhancement*

Dataset pelatihan, validasi, dan pengujian yang telah meningkatkan kualitas gambarnya dengan teknik *enhancement* kemudian akan dilakukan *encoding* pada label kelas menggunakan metode *label encode* 0, 1, dan 2. Angka-angka ini mewakili kelas *benign*, *malignant*, dan *normal*.

3.2 Hyper-parameters settings

1. Pencarian parameter dengan *bayesian optimization*

Pada bagian ini pengaturan hiperparameter dan optimasi bayesian, didefinisikan ruang pencarian untuk hiperparameter. Ruang pencarian ini mencakup *learning rate* dengan rentang dari $1e-6$ hingga $1e-2$, menggunakan distribusi *log-uniform*. Selain itu, pilihan *optimizer* terdiri dari *adam*, *sgd*, dan *rmsprop*. Setelah menjalankan 10 iterasi dengan total waktu *running time* 3 jam 37 menit 53 detik, ditemukan pengaturan hiperparameter optimal sebagai berikut: *optimizer* menggunakan metode Adam, dengan *learning rate* sebesar 0.0000687.

2. Parameter optimal untuk arsitektur VGGNet

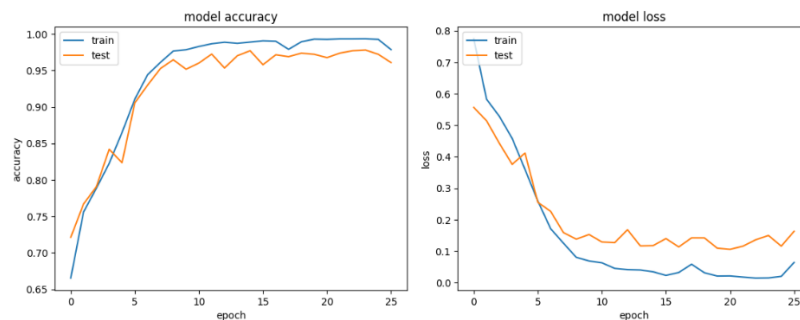
Dalam upaya meningkatkan kinerja arsitektur VGGNet untuk tugas klasifikasi tiga kelas, diterapkan pengaturan hiperparameter yang telah didapatkan dari hasil pencarian parameter yang optimal oleh *bayesian optimization*. Pengaturan tersebut meliputi penggunaan aktivasi *softmax* dan *optimizer* Adam, yang diatur dengan *learning rate* 0.0000687 dan *loss function* menggunakan *categorical cross-entropy*. Dengan pengaturan ini, tujuan adalah untuk meningkatkan akurasi dan stabilitas model dalam membedakan antara kelas-kelas yang berbeda.

3.3 Analisis hasil

Pembahasan hasil evaluasi kinerja model VGGNet setelah melalui proses pencarian hiperparameter menggunakan *bayesian optimization* ditampilkan dalam Tabel 3. Hasil evaluasi menunjukkan bahwa model VGGNet ini mencapai akurasi pengujian sebesar 35.40%. Meskipun nilai presisi, recall, dan F1-score relatif seragam, yakni sekitar 35.48%, 35.40%, dan 35.44% secara berturut-turut, namun secara keseluruhan, akurasi model tergolong rendah. Penemuan ini menunjukkan keterbatasan model dalam membedakan dengan tepat antara kelas-kelas yang ada.

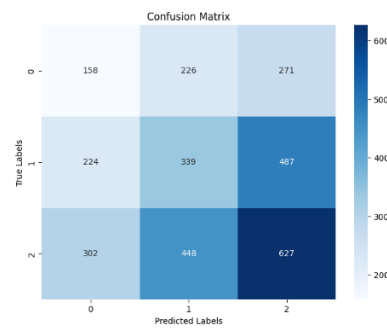
Tabel 3. Hasil evaluasi model

Model	Test accuracy	Precision	Recall	F1-Score
VGGNet dengan pencarian hiperparameter dengan <i>Bayesian Optimization</i>	35.40%	35.48%	35.40%	35.44%



Gambar 3. Grafik model *train* dan *validation* serta model *loss*.

Grafik menampilkan metrik kinerja model selama proses pelatihan, diukur dalam bentuk akurasi dan *loss* terhadap data pelatihan (*train*) dan validasi (*test*) seiring dengan jumlah *epoch*. Dari grafik akurasi, terlihat bahwa model berhasil meningkatkan akurasi terhadap data pelatihan secara konsisten, mencapai puncaknya sekitar *epoch* ke-10. Namun, akurasi terhadap data validasi cenderung stagnan atau bahkan menurun setelah beberapa *epoch* awal, menandakan kemungkinan *overfitting*. Grafik *loss* memperlihatkan bahwa *loss* terhadap data pelatihan secara stabil menurun seiring bertambahnya *epoch*, sementara *loss* terhadap data validasi cenderung stagnan atau meningkat setelah beberapa *epoch* awal.



Gambar 4. *Confusion matrix*

Confusion matrix yang terlihat pada Gambar 2, kita dapat mengamati distribusi prediksi model terhadap kelas sebenarnya. Sebagai contoh, model dengan tepat memprediksi 158 sampel dari kelas 0 (*benign*), 339 sampel dari kelas 1 (*malignant*), dan 627 sampel dari kelas 2 (*normal*). Namun, terdapat kesalahan prediksi dimana model salah mengklasifikasikan beberapa sampel dari setiap kelas ke kelas lainnya. Misalnya, model memprediksi 226 sampel dari kelas 0 (*benign*) sebagai kelas 1 (*malignant*), dan 302 sampel dari kelas 2 (*normal*) sebagai kelas 0 (*benign*).

4. Kesimpulan

Penelitian ini menyajikan pendekatan pemilihan parameter secara otomatis dengan menggunakan *bayesian optimization* sebagai alat seleksi parameter pada CNN, khususnya dalam arsitektur VGGNet untuk klasifikasi nodul paru. Metode yang diusulkan menunjukkan bahwa pencarian parameter optimal memerlukan waktu sekitar 3 jam 37 menit 53 detik dengan 10 iterasi, menghasilkan *optimizer* terbaik, yaitu Adam, dengan *learning rate* sebesar 0.0000687. Namun, saat menerapkan parameter tersebut pada VGGNet untuk klasifikasi multi kelas (*benign*, *malignant*, dan *normal*), terjadi kecenderungan *overfitting*, yang tercermin dari akurasi model yang rendah, hanya mencapai 35.40% dibandingkan dengan akurasi pada saat *training* sebesar 96,8%.

Daftar Pustaka

- [1] R. Raza *et al.*, “Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images,” *Eng Appl Artif Intell*, vol. 126, Nov. 2023, doi: 10.1016/j.engappai.2023.106902.
- [2] I. Ali, M. Muzammil, I. U. Haq, A. A. Khaliq, and S. Abdullah, “Efficient lung nodule classification using transferable texture convolutional neural network,” *IEEE Access*, vol. 8, pp. 175859–175870, 2020, doi: 10.1109/ACCESS.2020.3026080.
- [3] X. Wu, H. Zhang, J. Sun, S. Wang, and Y. Zhang, “YOLO-MSRF for lung nodule detection,” *Biomed Signal Process Control*, vol. 94, Aug. 2024, doi: 10.1016/j.bspc.2024.106318.
- [4] N. M. Aszemi and P. D. D. Dominic, “Hyperparameter Optimization in Convolutional Neural Network using Genetic Algorithms,” 2019. [Online]. Available: www.ijacsa.thesai.org
- [5] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, and S. H. Deng, “Hyperparameter optimization for machine learning models based on Bayesian optimization,” *Journal of Electronic Science and Technology*, vol. 17, no. 1, pp. 26–40, Mar. 2019, doi: 10.11989/JEST.1674-862X.80904120.
- [6] M. Zhang, H. Li, J. Lyu, S. H. Ling, and S. Su, “Multi-level CNN for lung nodule classification with Gaussian Process assisted hyperparameter optimization,” Jan. 2019, [Online]. Available: <http://arxiv.org/abs/1901.00276>
- [7] S. Hanifi, A. Cammarono, and H. Zare-Behtash, “Advanced hyperparameter optimization of deep learning models for wind power prediction,” *Renew Energy*, vol. 221, Feb. 2024, doi: 10.1016/j.renene.2023.119700.
- [8] L. Zhang *et al.*, “CNN-LSTM Model Optimized by Bayesian Optimization for Predicting Single-Well Production in Water Flooding Reservoir,” *Geofluids*, vol. 2023, 2023, doi: 10.1155/2023/5467956.
- [9] S. Pang *et al.*, “VGG16-T: A novel deep convolutional neural network with boosting to identify pathological type of lung cancer in early stage by ct images,” *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 771–780, 2020, doi: 10.2991/ijcis.d.200608.001.
- [10] S. A. Hicks *et al.*, “On evaluation metrics for medical applications of artificial intelligence,” *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-09954-8.