

# *SqueezeNet Feature Extraction dan Gradient Boosting untuk Klasifikasi Penyakit Monkeypox pada Citra Kulit*

Putu Desiana Wulaning Ayu<sup>1</sup>, Gede Angga Pradipta<sup>2</sup>

<sup>1,2</sup>Institut Teknologi dan Bisnis STIKOM Bali

e-mail: <sup>1</sup>wulaning\_ayu@stikom-bali.ac.id, <sup>2</sup>angga\_pradipta@stikom-bali.ac.id

Diajukan: 5 Februari 2024; Direvisi: 18 Maret 2024; Diterima: 19 Maret 2024

## **Abstrak**

Pada tahun 2022 WHO menerima laporan dari negara-negara non-endemik tentang kasus penyakit monkeypox (cacar monyet). Saat ini, terdapat 12 negara non-endemik di tiga wilayah WHO yaitu Eropa, Amerika, dan Pasifik Barat yang dilaporkan telah terjangkit virus cacar monyet. Monkeypox menunjukkan gejala serupa dengan cacar tetapi dengan tingkat keparahan yang berbeda, memerlukan identifikasi dan penanganan yang cepat untuk mencegah penularan lebih lanjut. Identifikasi penyakit monkeypox secara cepat dan akurat dapat dilakukan dengan pendekatan kecerdasan buatan yaitu model machine learning. Salah satu metode yang dapat digunakan untuk melakukan analisis data citra medis adalah metode Gradient Boosting. Penelitian ini mengembangkan konsep model klasifikasi penyakit monkeypox dengan menerapkan arsitektur Deep Learning, yaitu SqueezeNet + chi-square, tiga metode Gradient Boosting sebagai metode klasifikasi. Hasil eksperimen menunjukkan kombinasi model SqueezeNet + chi-square + XGBoost menghasilkan performansi yang lebih baik dari kombinasi dua model yang lain, dengan akurasi sebesar 0.943, presisi sebesar 0.942, dan AUC sebesar 0.987.

**Kata kunci:** Monkeypox, Klasifikasi, SqueezeNet, XGBoost.

## **Abstract**

In 2022, the WHO received reports from non-endemic countries about cases of monkeypox disease. Currently, there are 12 non-endemic countries across three WHO regions—Europe, America, and the Western Pacific—that have reported cases of monkeypox. Monkeypox shows symptoms similar to smallpox but with different severity levels, requiring quick and accurate identification and handling to prevent further transmission. The rapid and accurate identification of monkeypox disease can be performed using artificial intelligence approaches, namely machine learning models. One method that can be used for the analysis of medical image data is the Gradient Boosting method. This research develops a concept of monkeypox disease classification model by applying a Deep Learning architecture, SqueezeNet + chi-square, with three Gradient Boosting methods as the classification methods. The experimental results show that the combination of the SqueezeNet + chi-square + XGBoost model provides better performance than the other two model combinations, with an accuracy of 0.943, precision of 0.942, and an AUC of 0.987.

**Keywords:** Monkeypox, Classification, SqueezeNet, XGBoost.

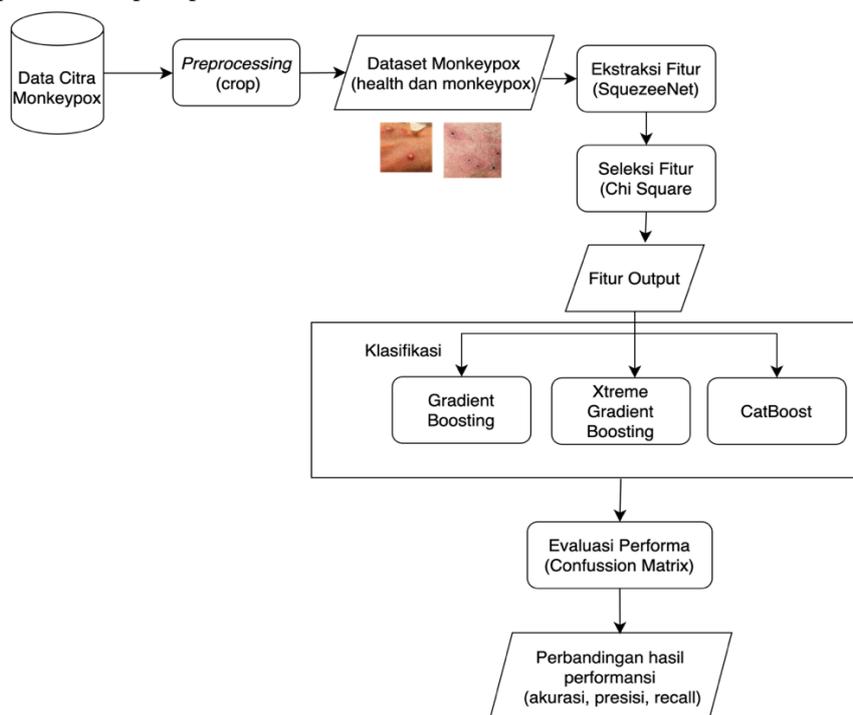
## **1. Pendahuluan**

Dalam beberapa tahun terakhir, dunia kesehatan global telah dihadapkan pada tantangan penanganan wabah penyakit monkeypox (cacar monyet), sebuah zoonis yang berasal dari hewan ke manusia dan memiliki potensi penyebaran yang signifikan di kalangan populasi manusia [1]. Monkeypox menunjukkan gejala serupa dengan cacar tetapi dengan tingkat keparahan yang berbeda, memerlukan identifikasi dan penanganan yang cepat untuk mencegah penularan lebih lanjut. Pada Mei 2022, WHO menerima laporan dari negara-negara non-endemik tentang kasus penyakit cacar monyet ([2]). Saat ini, terdapat 12 negara non-endemik di tiga wilayah WHO yaitu Eropa, Amerika, dan Pasifik Barat yang dilaporkan telah terjangkit virus cacar monyet [3]. Kluster kasus pertama ditemukan di Inggris, di mana kasus pertama ditemukan pada pasien dengan riwayat perjalanan terbaru dari Nigeria pada 7 Mei 2022, di London [4]. Identifikasi penyakit monkeypox secara cepat dan akurat dapat dilakukan dengan pendekatan kecerdasan buatan yaitu model machine learning [1]. Machine learning memiliki berbagai manfaat dalam industri kesehatan, termasuk identifikasi penyakit, diagnosis penyakit, prediksi penyakit, catatan kesehatan cerdas, pencitraan medis, dan lainnya. Oeh karena itu, tidak mengherankan jika banyak orang percaya

bahwa keberadaan machine learning akan sangat bermanfaat. Banyak metode yang tersedia dalam *machine learning* dapat digunakan untuk memroses data. Salah satu metode yang dapat digunakan untuk melakukan analisis data medis adalah metode *Gradient Boosting*[5]. Metode *Gradient Boosting* merupakan sebuah teknik pembelajaran mesin dengan membangun model secara bertahap dengan membangun pohon keputusan. Untuk dapat memprediksi dengan akurat. Metode ini sangat berguna dalam menghadapi kompleksitas data penyakit yang besar dan beragam, dan dapat menganalisis pola serta indikator penyakit dengan lebih efektif[6]. Penelitian dengan objek citra medis menggunakan pendekatan metode *Gradient Boosting* menghasilkan performansi yang cukup signifikan, seperti pada penelitian yang dilakukan oleh Bakasa dkk dalam mendeteksi kanker pankreas dengan menggunakan VGG16 sebagai *feature extractor* dan *XGBoost classifier*, di mana model menghasilkan akurasi sebesar 0.97[6]. Penelitian terkait identifikasi maupun klasifikasi penyakit *monkeypox* antara lain [7], [8], [9], di mana salah penelitian yang dilakukan oleh aldy dkk dengan metode *Gradient Boosting* dengan performa akurasi sebesar 0.71[10]. Mengacu dari penelitian yang dilakukan oleh [10] , sehingga pada penelitian ini menerapkan model klasifikasi berbasis *Gradient Boosting* untuk mengidentifikasi penyakit *monkeypox* pada citra kulit. Kontribusi penelitian ini menerapkan ekstraksi fitur berbasis *deep learning* yaitu menggunakan arsitektur *Squeezenet*, kemudian dilanjutkan dengan proses seleksi fitur dengan menggunakan metode *chi-square*. Untuk mengetahui performa model yang diusulkan, kami membandingkan tiga metode *Gradient Boosting* (*Gradient Boosting*, *XGBoost* (*Xtreme Gradient Boosting*) dan *CatBoosting*[11]). Dengan memanfaatkan dataset yang terdiri dari berbagai fitur klinis, model yang diusulkan diharapkan tidak hanya meningkatkan keakuratan diagnosis tetapi juga memberikan wawasan baru dalam penanganan dan pencegahan penyakit *monkeypox*. Analisis ini penting, tidak hanya untuk meningkatkan respons kesehatan saat ini tetapi juga untuk mempersiapkan dalam menghadapi wabah serupa di masa depan.

## 2. Metode Penelitian

Alur proses penelitian dalam deteksi penyakit *monkeypox* menggunakan metode *Gradient Boosting* dapat dilihat seperti pada Gambar 1, berikut:



Gambar 1. Alur proses klasifikasi penyakit monkeypox.

### 2.1 Akuisisi Data dan Preprocessing

Penelitian ini menggunakan dataset citra kulit yang diperoleh dari penelitian yang dilakukan oleh Bala D dkk [12]. Dataset terbagi menjadi 2 kelas atau kategori, yaitu Normal dan *Monkeypox*. Total Dataset yang digunakan pada penelitian ini sebanyak 925 citra, di mana pada penelitian ini kami membagi data training sebesar 70% dan data testing sebesar 30%. Dilanjutkan dengan proses *preprocessing* citra dengan

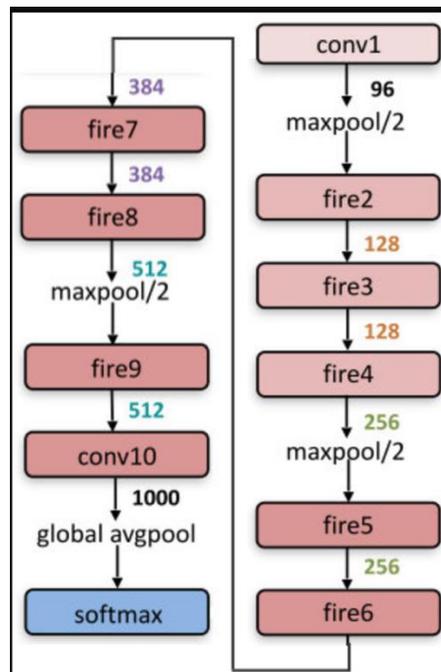
proses *crop* untuk menseragamkan dimensi citra menjadi 224 x 224 piksel. Gambar 2.2 menunjukkan contoh dataset citra.



Gambar 2. Contoh dataset citra dengan kondisi penyakit monkeypox[12].

## 2.2 Ekstraksi Fitur

*SqueezeNet* adalah model jaringan saraf konvolusional yang dirancang untuk memberikan akurasi yang setara dengan model CNN canggih sambil menggunakan jumlah parameter yang sangat sedikit[13]. Model ini menjadi populer karena efisiensinya yang luar biasa. *SqueezeNet* dibangun atas beberapa prinsip desain utama yang memungkinkannya mengurangi jumlah parameter secara signifikan tanpa mengorbankan performa. Berikut adalah komponen utama dari arsitektur *SqueezeNet*[13] dan Gambar 3 menunjukkan arsitektur dari *SqueezeNet*[13].



Gambar 3. Arsitektur dari *SqueezeNet* [13].

### 1. Fire Module

Setiap *Fire Module* dalam *SqueezeNet* terdiri dari tiga bagian utama, yaitu:

- *Squeeze layer*: lapisan ini terdiri dari konvolusi 1x1 Tujuan dari lapisan ini adalah untuk mengurangi kedalaman input, yaitu jumlah saluran atau depth sebelum data tersebut.
- *Expand Layer*: Setelah proses *squeeze*, lapisan *expand* menggunakan dua set filter konvolusi: filter 1x1 dan filter 3x3 yang diterapkan pada *output* dari *squeeze* layer *Output* dari kedua jenis filter ini kemudian digabungkan dalam *depth* untuk membentuk *output* akhir dari *Fire Module*.
- *Activation Function*: Setiap konvolusi diikuti oleh fungsi aktivasi, biasanya ReLU (Rectified Linear Unit), yang membantu model untuk menangkap non-linearitas.

## 2. Downsampling

Untuk mengurangi dimensi spasial dari *feature maps* tanpa kehilangan informasi penting, *SqueezeNet* menerapkan *downsampling* pada titik-titik tertentu dalam jaringan menggunakan teknik seperti *max-pooling*.

## 3. Regularisasi

*SqueezeNet* menggunakan *dropout*, sebuah teknik regularisasi yang mencegah *overfitting* dengan secara acak "mematikan" beberapa neuron selama proses pelatihan. Ini membuat model lebih *robust* karena tidak terlalu bergantung pada setiap neuron tertentu.

## 2.3 Gradient Boosting

Metode *gradient boosting* (juga dikenal sebagai Gradient Boosting Machine atau GBM) diusulkan oleh Friedman dalam dan berfungsi sebagai algoritma dasar untuk XGBoost, dan CatBoost. GBM bertujuan untuk proses optimasi di mana tujuannya adalah untuk meminimalkan suatu fungsi kerugian  $L$ . Persamaan dari GBM dapat dilihat pada (1)-(3)[11].

### 1. Hitung Gradient Negatif

$$r_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_m(x)} \quad (1)$$

Di mana  $r_{im}$  adalah residu pseudo yang menggantikan gradien dari kerugian  $L$  terhadap prediksi saat ini  $F_m(x)$  pada titik data  $i$ .

2. Membangun pohon keputusan / tree yang memprediksi residu pseudo  $r_{im}$  untuk setiap observasi.
3. Menemukan *Multiplier* Terbaik.

Untuk setiap node pada tree, menemukan nilai  $\gamma$  yang meminimalkan kerugian saat nilai  $\gamma$  ditambahkan ke prediksi sebelumnya:

$$\gamma_m = \arg \min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_m(x_i) + \gamma) \quad (2)$$

Di mana  $r_{jm}$  adalah region daun  $j$  pada pohon  $m$ .

### 4. Update Model

$$F_{m+1}(x) = F_m(x) + \nu \sum_{j=1}^J \gamma_{jm} I(x \in R_{jm}) \quad (3)$$

Di mana  $\nu$  adalah laju pembelajaran dan  $I$  adalah fungsi indikator yang bernilai 1 jika  $x$  berada di region  $R_{jm}$ .

Proses ini diulang sampai jumlah iterasi yang diinginkan atau peningkatan model menjadi minimal, di mana menunjukkan model telah mencapai batas kemampuannya dalam meminimalkan fungsi kerugian  $L$ . (Loss Function).

## 2.4 XGBoost (Xtreme Gradient Boosting)

*XGBoost* (Xtreme Gradient Boosting) adalah implementasi pengembangan dari algoritma *gradient boosting* yang dioptimalkan untuk kecepatan dan kinerja. XGBoost menggunakan beberapa teknik matematika untuk meningkatkan efisiensi dan efektivitas, termasuk regularisasi, pemotongan pohon, dan penanganan nilai yang hilang. Fungsi persamaan yang digunakan dalam XGBoost, yang didapat dilihat pada persamaan (4)[11].

$$\text{Obj}(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (4)$$

Di mana:

$\Theta$  adalah parameter model

$n$  adalah jumlah sampel pelatihan

$y_i$  adalah label sebenarnya dari data ke- $i$

$\Omega(f)$  adalah fungsi regulerisasi yang didefinisikan sebagai  $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|_2^2$ , di mana  $T$  adalah jumlah daun pada pohon,  $w$  adalah vektor bobot pada daun,  $\gamma$  dan  $\lambda$  adalah parameter regulerisasi.

$\lambda$  adalah parameter regulerisasi.

### 2.5 CatBoost

*CatBoost* menggunakan pendekatan *gradient boosting*, tetapi dengan beberapa modifikasi untuk mengatasi tantangan khusus dari data kategorikal dan masalah *overfitting* yang sering terjadi pada algoritma *boosting*. Persamaan *Catboost* dilihat pada persamaan (5)[11]. Fungsi tujuan *CatBoost* melibatkan dua komponen utama, yaitu fungsi kerugian  $L$  dan regulerisasi. Fungsi tujuan  $Obj$  untuk iterasi tertentu dapat dinyatakan sebagai:

$$Obj(\Theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \lambda \sum_{j=1}^T \Omega(f_j) \tag{5}$$

$\Theta$  adalah parameter model

$n$  adalah jumlah sampel pelatihan

$y_i$  adalah label sebenarnya dari data ke- $i$

$\Omega(f_j)$  adalah regulasi pohon ke  $-j$ , biasanya berbentuk kompleksitas pohon seperti jumlah pohon atau kedalaman pohon.

$\lambda$  adalah parameter regulerisasi.

$T$  adalah total pohon dalam model

### 2.6 Evaluasi Performansi

Untuk mengevaluasi performansi dari model menggunakan *confussion matrix* dengan metrik evaluasi pada persamaan (6)-(8)[14].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

## 3. Hasil dan Pembahasan

Pada eksperimen ini, pembagian dataset setiap hasil dari model fitur ekstraksi dengan *SqueezeNet* dan seleksi fitur diujicoba dengan mengkombinasikan dengan metode seleksi fitur *Chi-Square*. Eksperimen ini bertujuan untuk mengetahui kombinasi antara metode *SqueezeNet* dan seleksi fitur yang terbaik. Pada proses *learning* pada eksperimen ini menggunakan metode *Gradient Boosing*, *Extreme Gradient Boosting* dan *Catboost*, di mana untuk setiap metode menggunakan parameter jumlah pohon ( $T$ ) = 100, *learning rate* = 0.3, dan parameter regulerisasi  $\lambda$  = 6. Tabel 1-3 menunjukkan *confussion matrix* dari ketiga metode *Gradient Boosting*.

Tabel 1. *Confussion Matrix* pada metode *Gradient Boosting*.

		Predicted		
		Healthy	Monkeypox	Jumlah
Actual	Healthy	351	29	380
	Monkeypox	30	515	545
	Jumlah	381	544	925

Tabel 2. *Confussion Matrix* pada metode *XGBoosting*.

		Predicted		Jumlah
		Healthy	Monkeypox	
Actual	Healthy	349	31	380
	Monkeypox	22	523	545
	Jumlah	371	554	925

Tabel 3. *Confussion Matrix* pada metode *CatBoosting*.

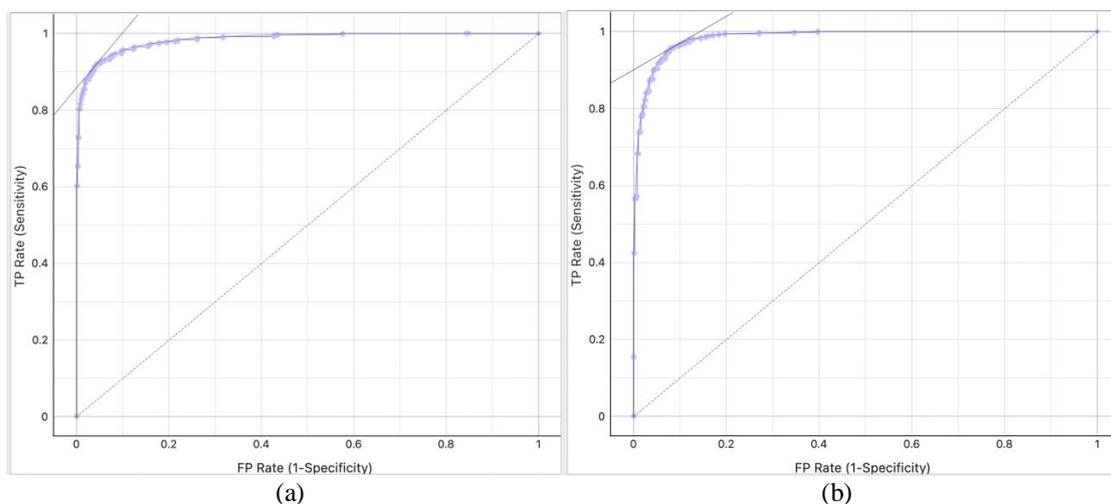
		Predicted		Jumlah
		Healthy	Monkeypox	
Actual	Healthy	347	33	380
	Monkeypox	23	522	545
	Jumlah	370	555	925

Tabel 4. Menunjukkan hasil eksperimen kombinasi model ekstraksi fitur, *seleksi fitur* dan metode *Gradient Boosting*. Metode evaluasi model menggunakan teknik *split* data 70:30. Berdasarkan dari hasil eksperimen, menunjukkan kombinasi antara model *SqueezeNet*, *Chi-Square* dan *XGBoost* mampu memberikan performa terbaik yaitu dengan nilai akurasi sebesar 0.943, presisi sebesar 0.942 dan AUC sebesar 0.987 (ditunjukkan pada Gambar 3).

Tabel 4. Perbandingan performansi pada ketiga metode klasifikasi dengan menggunakan seleksi fitur (Chi-Square).

Metode	Akurasi	Presisi	AUC
<i>Gradient Boosting</i>	0.936	0.931	0.982
<b><i>XGBoosting</i></b>	<b>0.943</b>	<b>0.942</b>	<b>0.987</b>
<i>CatBoost</i>	0.939	0.935	0.984

Kurva ROC (Area Under Curva) digunakan untuk melihat seberapa baik model membedakan antara dua kelas. Area di bawah kurva ROC (AUC) memberikan ukuran tunggal tentang kinerja model secara keseluruhan. Nilai AUC yang lebih tinggi atau mendekati nilai 1 menunjukkan model dengan kinerja prediksi yang lebih baik. Gambar 3. Menunjukkan nilai AUC pada setiap target atau label (*healthy* dan *monkeypox*). Pada penelitian ini terlihat bahwa model *XGBoost* mampu menghasilkan AUC sebesar 0.987.



Gambar 3. Kurva ROC dengan *XGBoosting*. (a) ROC pada label / target *healthy*, (b) ROC pada label/target *monkeypox*.

#### 4. Kesimpulan

Penanganan wabah penyakit *monkeypox* (cacar monyet) dapat dikendalikan dengan deteksi secara cepat dan akurat dapat dilakukan dengan pendekatan kecerdasan buatan yaitu model *machine learning*. Salah satu metode yang dapat digunakan untuk melakukan analisis data citra medis adalah metode *Gradient Boosting*. Pada penelitian ini mengembangkan konsep model klasifikasi penyakit monkeypox dengan menerapkan arsitektur *Deep Learning*, yaitu *SqueezeNet* + chi-square, tiga metode Gradient Boosting sebagai metode klasifikasi. Hasil eksperimen menunjukkan kombinasi model *SqueezeNet* + chi-square + *XGBoost* menghasilkan performansi yang lebih baik dari kombinasi dua model yang lain, dengan akurasi sebesar 0.943, presisi sebesar 0.942, dan AUC sebesar 0.987. Pengembangan penelitian ini dapat dilakukan dengan melakukan analisis lebih mendalam terhadap penyesuaian pada pemilihan parameter pada masing-masing metode.

#### Daftar Pustaka

- [1] M. M. Ahsan, M. R. Uddin, M. Farjana, A. N. Sakib, K. Al Momin, and S. A. Luna, "Image Data collection and implementation of deep learning-based model in detecting Monkeypox disease using modified VGG16," 2022, [Online]. Available: <http://arxiv.org/abs/2206.01862>
- [2] "diagnostics-13-01772-v2.pdf.crdownload."
- [3] S. T. Al Awaidy, F. Khamis, M. Sallam, R. M. Ghazy, and H. Zaraket, "Monkeypox (mpox) Outbreak More queries posed as cases soar globally," *Sultan Qaboos Univ Med J*, vol. 23, no. 1, pp. 1–4, 2023, doi: 10.18295/squmj.8.2022.046.
- [4] H. Hatami *et al.*, "Demographic, Epidemiologic, and Clinical Characteristics of Human Monkeypox Disease Pre- and Post-2022 Outbreaks: A Systematic Review and Meta-Analysis," *Biomedicines*, vol. 11, no. 3, 2023, doi: 10.3390/biomedicines11030957.
- [5] C. J. Tseng and C. Tang, "An optimized XGBoost technique for accurate brain tumor detection using feature selection and image segmentation," *Healthcare Analytics*, vol. 4, no. April, p. 100217, 2023, doi: 10.1016/j.health.2023.100217.
- [6] W. Bakasa and S. Viriri, "VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction," *J Imaging*, vol. 9, no. 7, 2023, doi: 10.3390/jimaging9070138.
- [7] M. Lakshmi and R. Das, "Classification of Monkeypox Images Using LIME-Enabled Investigation of Deep Convolutional Neural Network," *Diagnostics*, vol. 13, no. 9, 2023, doi: 10.3390/diagnostics13091639.
- [8] A. A. Aouragh and M. Bahaj, "Comparison Results of Hybrid CNN-Machine Learning Algorithms Architectures for Monkeypox Images Classification," *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology, IRASET 2023*, pp. 1–6, 2023, doi: 10.1109/IRASET57153.2023.10153062.
- [9] P. E. N. Taruno, G. S. Nugraha, R. Dwiyanaputra, and F. Bimantoro, "Monkeypox Classification based on Skin Images using CNN: EfficientNet-B0," *E3S Web of Conferences*, vol. 465, 2023, doi: 10.1051/e3sconf/202346502031.
- [10] F. Aldi, I. Nozomi, R. B. Sentosa, and A. Junaidi, "Machine Learning to Identify Monkey Pox Disease," *Sinkron*, vol. 8, no. 3, pp. 1335–1347, 2023, doi: 10.33395/sinkron.v8i3.12524.
- [11] Z. He, D. Lin, T. Lau, and M. Wu, "Gradient Boosting Machine: A Survey," pp. 1–9, 2019, [Online]. Available: <http://arxiv.org/abs/1908.06951>
- [12] D. Bala *et al.*, "MonkeyNet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Networks*, vol. 161, pp. 757–775, 2023, doi: 10.1016/j.neunet.2023.02.022.
- [13] N. Fragoulis, "A fast , embedded implementation of a Convolutional Neural Network for Image A fast , embedded implementation of a Convolutional Neural Network for Image Recognition," no. August, pp. 1–4, 2016, doi: 10.13140/RG.2.1.1778.9681.
- [14] P. D. Wulaning Ayu and G. A. Pradipta, "U-Net Tuning Hyperparameter for Segmentation in Amniotic Fluid Ultrasonography Image," *2022 4th International Conference on Cybernetics and Intelligent System, ICORIS 2022*, 2022, doi: 10.1109/ICORIS56080.2022.10031294.