

DOI <http://dx.doi.org/10.36722/sst.v7i3.1284>

Klasifikasi Kanker Tumor Payudara Menggunakan Arsitektur *Inception-V3* Dan Algoritma *Machine Learning*

Arif Supriyanto^{1*}, Wisnu Ananta Kusuma¹, Hendra Rahmawan²¹Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor,
Jalan Raya Dramaga, Kampus IPB Dramaga Bogor, 16680²Pusat Studi Biofarmaka Tropika, Institut Pertanian Bogor,
Jalan Raya Dramaga, Kampus IPB Dramaga Bogor, 16151Penulis untuk Korespondensi/E-mail: arif.supriyanto@if.uai.ac.id

Abstract – Breast cancer is a disease that arises due to breast tissue cells that grow abnormally and continuously. This disease is a disease with a large increase in number of around 13 million per year, with a mortality rate of 9.6% from a total of 65,858 cases. Early detection of breast cancer for prevention needs to be done, with the hope that breast cancer is easier to treat and cure and can even be prevented before it enters an advanced stage. In this research, build a model with transfer learning technique for breast cancer classification. There are 4 methods tested, namely Inception-V3 feature extraction with the Radial Basic Function Neural Network classification method, FeedForward Neural Network, Logistic Regression and feature extraction by making changes to the hyperparameter layer. This study compares the four models to get the best one to solve the problem of breast cancer classification. The data used in this study are breast cancer image data with a zoom scale of 40X, 100X, 200X and 400X. The dataset was sourced from The Laboratory University of Parana with P&D Laboratory Pathological Anatomy and Cytopathology, Parana, Brazil. The results of this study indicate that the Inception-V3 feature extraction method with the Logistic Regression classification method on the 40X zoom scale data provides the best accuracy (93.00%), precision (94.00%), and recall (91.00%) F1-score (92.00%).

Abstrak - Kanker payudara adalah penyakit yang timbul diakibatkan oleh sel jaringan payudara yang tumbuh secara tidak normal dan terus menerus. Penyakit ini merupakan penyakit dengan jumlah kenaikan yang besar sekitar 13 juta pertahun, dengan tingkat kematian persentase 9,6% dari total 65.858 kasus. Deteksi kanker payudara sejak dini untuk pencegahan perlu dilakukan, dengan harapan kanker payudara lebih mudah ditangani dan disembuhkan bahkan bisa dicegah sebelum memasuki stadium lanjut. Pada penelitian ini membangun model dengan teknik transfer learning untuk klasifikasi kanker payudara. Terdapat 4 metode yang diujikan yaitu ekstraksi fitur *Inception-V3* dengan metode klasifikasi *Radial Basic Function Neural Network*, *FeedForward Neural Network*, *Logistic Regression* dan ekstraksi fitur dengan melakukan perubahan pada hyperparameter layer. Penelitian ini membandingkan ke empat model tersebut untuk mendapatkan yang terbaik untuk mengatasi masalah klasifikasi kanker payudara. Data yang digunakan pada penelitian ini adalah data citra kanker payudara dengan zoom scale 40X, 100X, 200X dan 400X. Dataset bersumber dari The Laboratory University of Parana dengan P&D Laboratory Pathological Anatomy and Cytopathology, Parana, Brazil. Hasil penelitian ini menunjukkan metode ekstraksi fitur *Inception-V3* dengan metode klasifikasi *Logistic Regression* pada data zoom scale 40X memberikan akurasi terbaik (93,00%), precision (94,00%), dan recall (91,00%) F1-score (92,00%).

Keywords - Breast cancer, transfer learning, Convolutional Neural network, Inception-V3 dan algoritme machine learning

PENDAHULUAN

Kanker payudara adalah suatu penyakit yang timbul diakibatkan oleh sel jaringan payudara yang tumbuh secara tidak normal dan terus menerus. Menurut [1], penyakit kanker payudara akan meningkat 13 juta pertahun ditahun 2030. Hal ini merupakan suatu jumlah kenaikan yang cukup besar. Diikuti dengan tingkat kematian akibat kanker payudara sekitar 22.430 kasus dengan persentase 9.6% dari total 65.858 kasus [2]

Kanker payudara menjadi ancaman serius bagi wanita di seluruh dunia, termasuk Indonesia saat ini. Kemenkes menyatakan lebih dari 80.00% kasus kanker payudara ditemukan pada pasien stadium lanjut, dimana upaya pengobatan sulit dilakukan [3]. Oleh sebab itu, deteksi kanker payudara sejak dini perlu dilakukan sebagai pencegahan [3].

Penelitian klasifikasi kanker payudara sebelumnya telah dilakukan [4]. Tujuan penelitiannya yaitu klasifikasi *binary* (*Benign* dan *Malignant*) menggunakan citra Hispathological. Metode yang digunakan untuk representasi tekstur citra adalah LBP, CLBP, LPQ, GLCM, TAS dan ORB dan metode klasifikasi dengan *Support Vector Machine* (SVM), K-NN, *Random Forest*, *Neural network*. Hasil penelitian tersebut menghasilkan akurasi 85%. Namun cenderung memprediksi kelas *malignant*. Pada penelitian selanjutnya Novar dan Made [5] dan Nugroho [6]. Penelitian tersebut menggunakan metode *K-Means*, GLCM, dan SVM. Hasil penelitian tersebut menghasilkan akurasi 80% dan 71,8%. Pada penelitian sebelumnya ekstraksi fitur dilakukan secara manual, sehingga mempengaruhi performa dalam klasifikasi citra kanker payudara.

Ekstraksi fitur secara otomatis dapat dilakukan oleh metode *Convolutional Neural network* (CNN) [7]. CNN mampu untuk mengekstraksi fitur-fitur yang abstrak yang ada pada data citra. Model CNN salah satunya yaitu *Inception-V3* mampu mengenali 1000 kelas yang berbeda dari jutaan gambar. Namun penggunaan model CNN membutuhkan dataset yang besar untuk menghindari *over fit* dan menghasilkan akurasi yang baik. Penelitian [8], memanfaatkan model CNN untuk dataset yang kecil dengan menggunakan teknik *transfer learning* dengan sedikit perubahan pada lapisan klasifikasi citra.

Transfer learning adalah solusi yang bisa dimanfaatkan pada dataset yang tidak cukup ideal untuk melakukan proses training dari awal [9]. Penelitian *transfer learning* dan *fine tuning* telah

dilakukan [10], penelitian tersebut menghasilkan akurasi sebesar 95,12% dan 95,81% setelah dilakukan *fine tuning*.

Sebuah penelitian mengatakan bahwa nilai akurasi menggunakan *deep learning* lebih tinggi apabila diganti *classifier* KNN pada *fully connected* [11], penelitian klasifikasi dengan menggunakan algoritme *machine learning* telah dilakukan [12], tujuan penelitian tersebut untuk melakukan komparasi 5 metode algoritme klasifikasi *neural network*, *Logistic Regression*, KNN untuk melihat metode terbaik. Akurasi peling baik ditunjukkan pada algoritme *neural network* dan *Logistic Regression*. Penelitian [13], menyebutkan *radial basis function neural network* (RBFNN) merupakan salah satu metode handal untuk permasalahan regresi dan klasifikasi.

Berdasarkan permasalahan yang ada pada penelitian sebelumnya, penelitian ini akan mengaplikasikan *transfer learning* model CNN *Inception-V3* untuk klasifikasi kanker payudara. Perubahan dilakukan pada arsitektur jaringan *Inception-V3*, dengan mengurangi jumlah layer konvolusi dan memotong jaringan *fully connected* digantikan algoritme *machine learning* yaitu Radial Basis Function *Neural network* (RBFNN), *Feed Forward Neural network* (FFNN) dan *Logistic Regression*.

METODE PENELITIAN

Data Penelitian

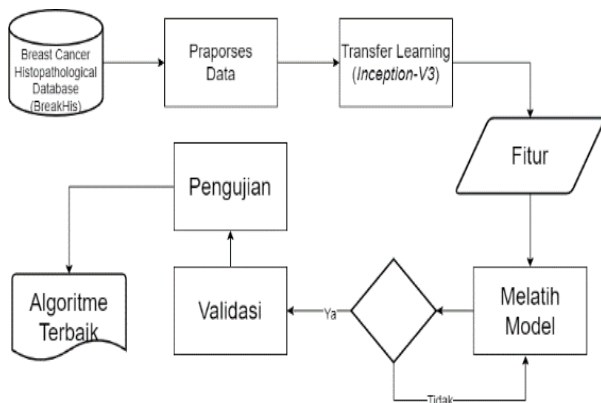
Penelitian ini menggunakan data citra kanker payudara sebagai data input. Data tersebut berasal dari kerja sama The Laboratory of Vision, Robotics and Imaging, Department of Informatics of the Federal University of Parana dengan P&D Laboratory Pathological Anatomy and Cytopathology, Parana, Brazil.

Alat dan Bahan

Penelitian ini dilakukan menggunakan perangkat keras dan perangkat lunak sebagai berikut. Adapun Perangkat lunak yang digunakan antara lain: (1). Google Colab. (2). Visual Studio Code sebagai code editor. (3). Python Data Analysis.

Tahapan Penelitian

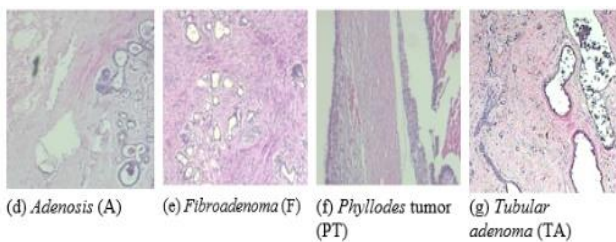
Tahapan penelitian yang dilakukan mencakup beberapa tahapan, ditunjukkan pada gambar.



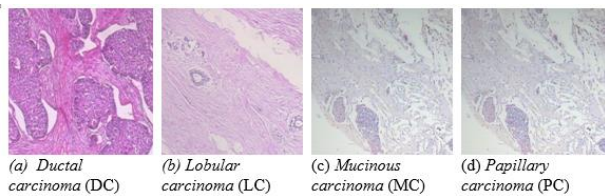
Gambar 1. Tahapan Penelitian

Dataset BreakHis

Penelitian ini menggunakan dataset *breakHis* sebagai data input. Dataset *breakHis* terbagi menjadi dua kelas yaitu *benign* (jinak) dan *malignant* (ganas) dengan masing-masing terbagi menjadi empat subkelas yaitu *Ductal carcinoma* (DC), *Lobular carcinoma* (LC), *Mucinous carcinoma* (MC), dan *Papillary carcinoma* (PC) pada kelas *malignant*. *Adenosis* (A), *Fibroadenoma* (F), *Phyllodes tumor* (PT), dan *Tubular adenoma* (TA) pada kelas *benign*. Setiap *subclass* terbagi menjadi empat *zoom scale* yaitu 40x, 100x, 200x, 400x. Kemudian terbagi menjadi tiga channel RGB (*red, green, blue*) yang berukuran 700 x 460 *pixels*. Pembagian jumlah dataset *breakhis* ditunjukkan pada Tabel 3. Gambar 2 menunjukkan sampel data citra *benign* (jinak). Gambar 3 menunjukkan sampel data citra *malignant* (ganas).



Gambar 2. Datacitra tipe *benign*



Gambar 3. Datacitra tipe *malignant*

Tabel 1. Informasi dataset BreakHis

Magnification	Benign	Malignant	Total
40X	625	1370	1995
100X	644	1437	2081
200X	623	1390	2013

Magnification	Benign	Malignant	Total
400X	588	1232	1820

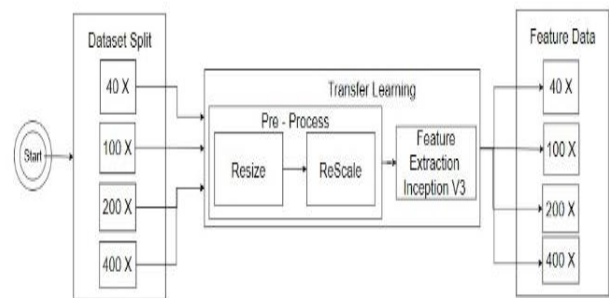
Tabel 1 menunjukkan dataset yang digunakan memiliki total 7.909 sampel citra yang terdiri dari 82 pasien. Total sampel citra *benign* yaitu 2,480 data dan *malignant* yaitu 5,429 data yang terdiri dari *zoom scale* yang berbeda.

Data Preprocessing

Preprocessing merupakan langkah awal pengolahan data yang dilakukan dengan tujuan mendapatkan data yang optimal. Pada penelitian ini proses *preprocessing* yang dilakukan adalah *resize* dan *rescale*. *Resize* pada citra bertujuan untuk mengubah resolusi dan ukuran horizontal dan vertikal suatu citra digital. Pada penelitian ini citra di *resize* berukuran 224 x 224 x 3 sesuai ketentuan pada arsitektur *Inception-V3*, bertujuan untuk mempercepat dan memudahkan proses perhitungan. *Rescale* citra bertujuan untuk mengubah ukuran citra. Pada penelitian ini citra di *rescale* dengan membagi *pixel* terkecil dengan *pixel* terbesar. *rescale* digunakan untuk mengubah ukuran *pixel* RGB gambar (0-255) menjadi rentang angka (0-1) untuk mempermudah proses training data.

Transfer learning

Metode dalam penelitian ini menggunakan arsitektur CNN *Inception-V3*. Arsitektur CNN *Inception-V3* terbagi atas dua tahapan yaitu, *feature learning* dan *classification*. Namun pada penelitian ini hanya menggunakan tahapan *feature learning*. Rancangan dari arsitektur *transfer learning* pada penelitian ini diilustrasikan pada Gambar 4.



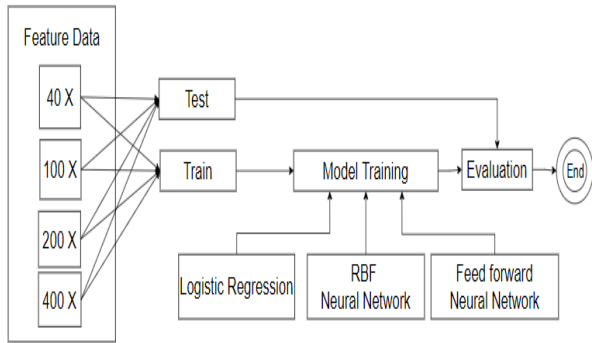
Gambar 4. Arsitektur *transfer learning*

Tahapan *feature learning* terbagi atas 3 lapisan yakni, *convolutional layer*, *activation layer* dan *pooling layer*. Tiga layer tersebut memiliki urutan yang tidak selalu sama, *layer* tersebut dimodifikasi sesuai dengan kebutuhan. umumnya proses *feature learning* ini diawali proses *convolusi* antara *matriks*

input. Proses ini dilakukan berulang kali sampai didapatkan *ekstraksi fitur* berupa *feature map*.

Klasifikasi

Pada tahapan klasifikasi, penelitian ini menggunakan algoritme *regression logistik*, RBFNN, FFNN. Tahapan training klasifikasi ditunjukkan pada Gambar 5.



Gambar 5. Tahapan training klasifikasi

Proses klasifikasi dibagi menjadi dua, yaitu proses training dan proses testing. Proses training digunakan untuk membuat model arsitektur CNN dan proses testing digunakan untuk menguji model yang telah dibuat.

Tahapan Evaluasi

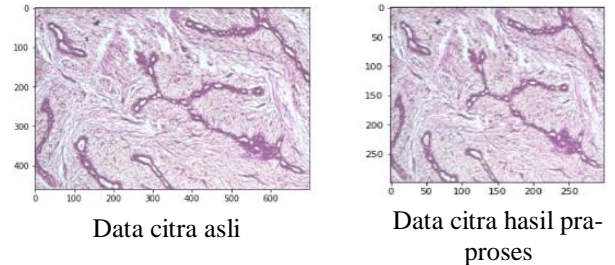
Tahap evaluasi merupakan tahap pemeriksaan performa hasil pengujian yang didesain menjadi beberapa skenario pengujian. Pada tahapan evaluasi pada penelitian ini menggunakan akurasi, *precision*, *recall* dan *F1 score* untuk mengukur kinerja algoritme.

HASIL DAN PEMBAHASAN

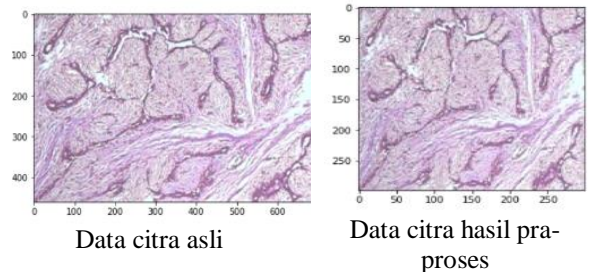
Praproses Data

Praproses citra pada penelitian ini bertujuan untuk mempercepat waktu pengolahan citra. Praproses data breakhis diperlakukan pada dua jenis kelas data yaitu *malignant* dan *benign* yang belum dilakukan proses *cropping*. Praproses data dilakukan dengan *rescale* (skala ulang) dan *resize* (perubahan ukuran) agar kedua jenis data siap dimasukkan ke model. Hasil dari *pra-process* menghasilkan perubahan pada ukuran citra pada setiap kelas data. Praproses data citra *benign zoom scale 40X* mengubah ukuran citra dari 3120 x 3120 piksel menjadi 64 x 64 piksel. Praproses data citra *benight zoom scale 100X* mengubah ukuran citra dari 3120 x 3120 piksel menjadi 64 x 64 piksel. Praproses data citra *benight zoom scale 200X* mengubah ukuran citra dari 3120

x 3120 piksel menjadi 64 x 64 piksel. Praproses data citra *benight zoom scale 400X* mengubah ukuran citra dari 3120 x 3120 piksel menjadi 64 x 64 piksel. Hasil praproses data zoom scale 40X pada kelas data *benign* ditunjukkan pada Gambar 6. Hasil praproses data zoom scale 40X pada kelas data *malignant* ditunjukkan pada Gambar 7.



Gambar 6. Hasil praproses data zoom scale 40x pada kelas data *benign*



Gambar 7. Hasil praproses data zoom scale 40x pada kelas data *malignant*

Pre-training Model Ekstraksi Feature

Proses *transfer learning* menggunakan metode *Inception-V3* untuk mendapatkan *ekstraksi feature* setiap ukuran data citra breakhis pada zoom scale 40x, 100x, 200x dan 400x. Proses *transfer learning* dilakukan menggunakan modul *Conv2D* yang tersedia di *package Keras* dalam perangkat lunak Python. Matriks komposisi *feature* setiap ukuran data ditunjukkan pada Tabel 2.

No	Zoom scale	Jumlah Feature Data
1.	40x	1995, 2048
2.	100x	2081, 2048
3.	200x	2013, 2048
4.	400x	1820, 2048

Pemodelan dan Pengujian

Pada penelitian ini pemodelan klasifikasi kanker payudara masing-masing dilakukan sebanyak 4 kali yaitu implementasi *Inception-V3* dengan RBF *Neural Network*, *Inception-V3* dengan *Logistic Regression*, *Inception-V3* dengan *Feedforward*

Network dan Inception-V3 dengan Fine Tunning. Proses pelatihan untuk setiap model memanfaatkan package Keras dan Sickit-Learn. Sebelum proses training di jalankan. Tahapan pertama dilakukan pembagian data training dan data testing yang digunakan sebanyak 7.909 citra. Pembagian data memanfaatkan package train test split pada Python dengan nilai probabilitas pembagian data 0.2 dengan 80% data latih dan 20% data uji. Setiap zoom scale data citra telah dilakukan pembagian data training dan data testing sesuai dengan nilai probabilitas secara random. Tahapan selanjutnya penentuan parameter dan tuning parameter kebutuhan train data setiap pemodelan yang akan diamati berdasarkan learning rate, epoch dan parameter pendukung algoritme Logistic Regression, Feedforward Neural Network, RBF Neural Network. Detail parameter algoritme klasifikasi ditunjukkan pada Tabel 3.

Tabel 3. Parameter Pelatihan

Model	Package	Tunning Parameter	Nilai
Inception-V3 dengan Logistic Regresion	Menggunakan modul Logistic Regression yang tersedia di package Sickit-Learn.	Pinalty C	12 11.28
Inception-V3 dengan RBF Neural Network	Menggunakan modul yang tersedia di package Sickit-Learn RBF Layer dan sklearn. cluster import kmeans dan random	Cluster RBF Layer Beta Fungsi activation	1024 512, 0.003 Sigmoid Biner
Inception-V3 dengan Feed Forward Neural Network	Menggunakan modul yang tersedia pada Keras	Fungsi activation	Sigmoid Biner
Inception-V3 dengan Fine Tunning		Iterasi Convolution Layer	50, 100 : 249, :164

Setelah parameter pelatihan di tentukan selanjutnya dilakukan proses pelatihan model. Pelatihan dan pengujian model Inception-V3 dengan Logistic Regression, Feedforward Neural Network, RBF Neural Network dan Inception-V3 dengan fine tuning diimplementasikan menggunakan packaging Keras pada bahasa pemograman python 3.7. Hasil rata-rata evaluasi model terbaik setiap algorime ditunjukkan pada Tabel 4. Hasil rata-rata evaluasi model Inception-V3 dengan fine tuning ditunjukkan pada Tabel 4 sebagai berikut.

Tabel 4. Evaluasi model Inception-V3 – Algoritme machine learning

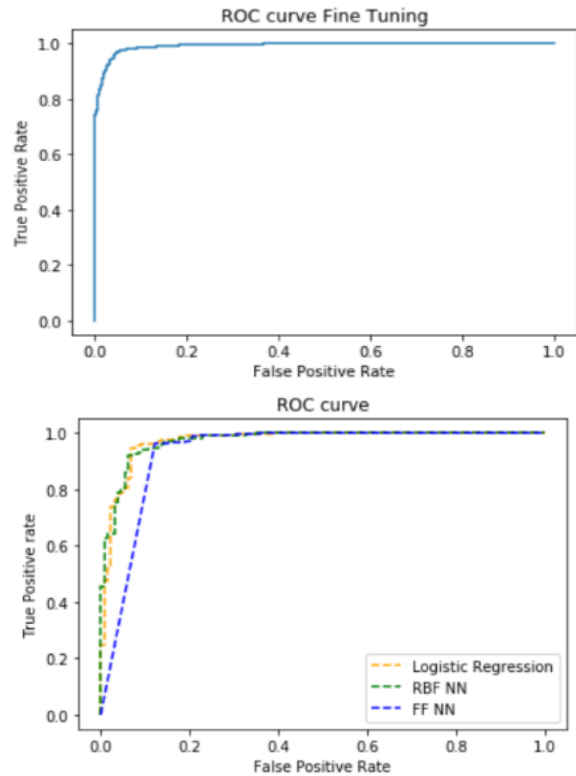
Zoom scale	Model	Evaluasi (%)			
		Akura si	Presi si	Reca ll	F1-Scor e
40X	Incepti on-V3 + Logisti c Regres sion	93	94	91	92
100X	Incepti on-V3 + RBF Neural Netwo rk	93	91	91	91
200X	Incepti on-V3 + Logisti c Regres sion	91	90	88	89
400X	Incepti on-V3 + Logisti c Regres sion	87	85	84	85

Berdasarkan Tabel 4 Akurasi terbaik pada dataset zoom scale 40X adalah model inception-V3 dengan logistic regression yaitu 93% dengan nilai presisi 94%, recall 91%, F1-score 92%.

Tabel 5 Evaluasi model *Inception-V3* dengan *fine tuning*

literasi	Model	Evaluasi (%)			
		Akura si	Presi si	Reca ll	F1- Scor e
50	Fine Tunning di layer ke 249	87	74	87	78
	Fine Tunning di layer ke 164	88	76	90	80
100	Fine Tunning di layer ke 249	88	77	88	81
	Fine Tunning di layer ke 164	90	78	91	82

Berdasarkan Tabel 5 Parameter model *Inception-V3* dengan Fine Tuning terbaik yang didapatkan pada epoch 50 dan convolution layer ke 164. Pengujian ini menghasilkan akurasi sebesar 97% dengan nilai presisi 94%, nilai *recall* 85% dan F1-Score 90%. Informasi kinerja algoritma klasifikasi dalam bentuk grafik digunakan Receiver Operating Characteristic (ROC) ditunjukkan pada Gambar 8.



Gambar 8. Receiver Operating Characteristic (ROC)

Berdasarkan grafik Gambar 8 disimpulkan bahwa kinerja *Inception-v3* dengan *logistic regression* dan fine tuning lebih bagus dibandingkan RBF Neural Network dan *feedforward Neural Network* karena lebih mendekati titik 0,1.

KESIMPULAN

Berdasarkan skenario pelatihan data zoom scale 40X, 100X, 200X dan 400X dengan metode ekstraksi fitur *Inception-V3* dan model klasifikasi *logistic regression*, *Radial Basic Function NN* dan *Feedforward NN* didapatkan metode terbaik didapatkan pada data zoom scale 40X dengan ekstraksi fitur *Inception-V3* dan model klasifikasi *logistic regression* yaitu dengan akurasi 93,00%, presisi 94,00%, *recall* 91,00% dan *F1-score* 92,00%. Skenario pelatihan *Inception-V3* dengan fine tuning dilakukan perubahan hyperparameter layer : 164 dan epoch 100 menghasilkan akurasi terbaik yaitu dengan akurasi 90,00%, presisi 78,00%, *recall* 91,00% dan *F1-score* 82,00%.

DAFTAR PUSTAKA

- [1] S. Sugiharto, O. D. Wahyuni, C. Song, "Edukasi Kanker Payudara Dan Deteksi Dini Pada Kader Wanita Kelurahan Tomang," *Jurnal Bakti Masyarakat Indonesia*, vol. Vol. 4, No. 2, pp. 351-359, Agustus 2021.
- [2] Kementrian Kesehatan RI. "The Global cancer observatory". 2020, [Online]. <https://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf>
- [3] Kementerian Kesehatan RI, "Panduan program nasional gerakan pencegahan dan deteksi dini kanker leher Rahim dan kanker payudara", 2015, <http://p2ptm.kemkes.go.id/uploads/2016/10/Panduan-Program-Nasional-Gerakan-Pencegahan-dan-Deteksi-Dini-Kanker-Kanker-Leher-Rahim-dan-Kanker-Payudara-21-April-2015.pdf>
- [4] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *Journal of IEEE Transactions on Biomedical Engineering (TBME)*, vol. 63(7), pp. 1455-1462, 2016.
- [5] G. M. Sasmita, D. P. Githa, K. S. Wibawa, D. P. Singgih, "Mammogram Menggunakan Metode K-Means, Gcm, Dan Support Vector Machine (Svm)," *jurnal merpati*, Vol. 6, No. 1, pp. 2252-3006, 2018.
- [6] F. S. Nugraha, M. J. Shidiq, Sri Rahayu, "Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara," *Journal of Computing and Information System Nusa Mandiri*, vol. 15, 2019.
- [7] I. Cholissodin, A. A. Soebroto, "Machine Learning & Deep Learning" (Teori & Implementasi)., 2021.
- [8] F. Rochman, H. Junaedi, "All-transfer learning for deep neural networks and its application to sepsis classification," *Journal of Frontiers in Artificial Intelligence and Applications*, vol. 285, p. 1586–1587, 2016.
- [9] R. Erwandi, S. Suyanto, "Klasifikasi Kanker Payudara Menggunakan Residual Neural Network," *Indonesia Journal on Computing*, vol. 5, pp. 45-52, 2020.
- [10] J. Elisha, & S. Budi, "Pengaruh Preprocessing Terhadap Klasifikasi Diabetic Retinopathy dengan Pendekatan Transfer learning Convolutional Neural network," *Jurnal Teknik Informatika Dan Sistem Informasi*, vol. 7(1), p. 110–124, 2021.
- [11] S. Liu, X. Liang, L. Liu, X. Shen, J. Yang, "Matching-CNN meets KNN: Quasi-parametric human parsing," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, p. 1419–1427, 2015.
- [12] S. Dewi, "Komparasi 5 Metode Algoritme Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *Journal of Computing and Information Technology*, vol. Vol. XIII, No. 1, pp. 1978-2136, 2016.
- [13] E. Y. Puspaningrum, B. Nugroho, H. A. Manggala, "Penerapan Radial Basis Function Untuk Klasifikasi Jenis Tanah," vol. Vol. Xv Nomor 1, pp. 2686-6099, 2020.