

## DETEKSI DINI KANKER PAYUDARA MENGGUNAKAN METODE CONVOLUTION NEURAL NETWORK (CNN)

Merinda Lestandy<sup>1</sup>, Putri Khalifah Hanifah<sup>2</sup>, M. Irfan<sup>3</sup>  
Program Studi Teknik Elektro, Universitas Muhammadiyah Malang<sup>1</sup>  
Email : merindalestandy@umm.ac.id<sup>1</sup>

### ABSTRAK

Kanker payudara merupakan penyakit yang mematikan, tidak hanya terjadi pada wanita namun pria juga dapat terkena penyakit ini. Seseorang bisa rentan terkena penyakit kanker payudara seiring bertambahnya usia, baik disebabkan karena pola hidup yang kurang sehat atau bahkan karena keturunan. Oleh karenanya, perlu ada pencegahan sejak dini dengan cara mendeteksi adanya kanker. *Deep learning* telah menjadi salah satu topik hangat dalam dunia *machine learning* karena kapabilitasnya yang signifikan dalam memodelkan berbagai data kompleks seperti citra dan suara. *Deep learning* yang saat ini memiliki hasil paling signifikan dalam pengenalan citra adalah *Convolutional Neural Network* (CNN). Pada penelitian ini dilakukan deteksi kanker payudara menggunakan metode CNN. Data yang digunakan dalam penelitian ini adalah data citra histopatologi sebanyak 277.524 data dengan komposisi data training 80% dan data uji 20%. Dataset terdiri dari citra hispatologi kanker payudara normal dan tidak normal dimana data dilakukan preprocessing terlebih dahulu. Hasil penelitian ini mendapat akurasi sebesar 80%.

**Kata Kunci:** Citra Histopatologi; Convolution Neural Network (CNN); Kanker Payudara

### ABSTRACT

*Breast cancer is a deadly disease, not only in women but also in men. A person can be susceptible to breast cancer with age, either due to an unhealthy lifestyle or even because of heredity. Therefore, there needs to be early prevention by detecting the presence of cancer. Deep learning has become one of the hot topics in machine learning because of its significant capabilities in modeling complex data such as images and sounds. The deep learning that currently has the most significant results in image recognition is the Convolutional Neural Network (CNN). In this study, breast cancer was detected using the CNN method. The data used in this research is histopathological image data as much as 277,524 data with 80% training data composition and 20% test data. The dataset consists of hispathological images of normal and abnormal breast cancer where the data is preprocessed first. The results of this study obtained an accuracy of 80%.*

**Keywords:** Histopathological Image; Convolution Neural Network (CNN); Breast cancer.

### 1. PENDAHULUAN

Kanker merupakan salah satu penyakit yang banyak menyebabkan kematian di seluruh dunia. Dari data *International Agency for Research of Cancer* (IARC) pada tahun 2012 tercatat lebih dari 8,2 juta jiwa meninggal dunia akibat penyakit

kanker. Sedangkan keseluruhan kasus yang tercatat pada tahun tersebut mencapai 14.067.894 jiwa. penderita kanker payudara dalam kurun waktu lima tahun terakhir mendapatkan presentasi yaitu sebesar 19,2% dari keseluruhan kasus (Binabar & Ivandari, 2017).

Kanker payudara masih menjadi masalah yang sangat penting khususnya di Indonesia, karena 68,6% pasien berobat pada stadium lanjut, sedangkan yang berobat pada stadium dini hanya 22,4%, yang berarti masih banyak pasien maupun calon pasien yang tidak berobat sebelum kanker payudara tersebut semakin parah (Darsyah, 2013). Penyebab kanker payudara masih belum diketahui sampai saat ini. Meskipun demikian, beberapa riset mengidentifikasi sejumlah faktor yang dapat meningkatkan risiko pada individu tertentu, diantaranya keluarga yang memiliki riwayat penyakit serupa, usia yang makin bertambah, tidak memiliki anak, kehamilan pertama pada usia di atas 30 tahun, periode menstruasi yang lebih lama (menstruasi pertama lebih awal atau menopause lebih lambat), faktor hormonal (baik estrogen maupun androgen) (Darsyah, 2013).

Terdapat beberapa penelitian menggunakan citra salah satunya adalah Wayan Suartika E. P, Arya Yudhi Wijaya, dan Rully Soelaiman melakukan penelitian untuk klasifikasi citra umum untuk mengenali generalisasi pada citra tersebut, menggunakan metode *Convolutional Neural Network* (CNN). Hasil akurasi pada penelitian ini sebesar 20%-50% (Eka Putra, 2016). Metode berbasis *machine learning* sangat berpengaruh dibidang kedokteran mulai dari penemuan obat sampai pengambilan keputusan pada suatu penyakit (Lestandy et al., 2021; Ma et al., 2019; Puttagunta & Ravi, 2021; Syafa'ah et al., 2021; Syafa & Lestandy, 2021). Penelitian yang dilakukan oleh Riko Rochan (2019) mengenai deteksi kontur tulang dan mengidentifikasi terjadinya patah tulang tertutup dengan menggunakan

metode CNN pada citra ultasonik. Hasil penelitiannya yaitu mendapatkan akurasi 95,3%, sensitivitas 95%, dan *specificity* sebesar 96% (Rokhana et al., 2019). CNN merupakan metode yang paling direkomendasikan untuk mendeteksi suatu citra karena CNN dapat mendeteksi citra yang beresolusi tinggi maupun rendah dengan sangat baik (Dong et al., 2016; Ma et al., 2019; Sun et al., 2019). *Machine learning* yang lain memiliki akurasi dibawah CNN untuk mendeteksi suatu citra seperti algoritma ANN 74,48% (Janghel et al., 2010) dan SVM 73,7% (Azis et al., 2019).

Berdasarkan beberapa penelitian menunjukkan bahwa CNN memiliki tingkat akurasi yang lebih baik pada data citra. Hal ini dikarenakan CNN memiliki kelebihan dibandingkan *neural network*, yaitu CNN dapat mengklasifikasikan inputan baru yang posisinya berada diluar *data training* sedangkan pada *neural network* kurang maksimal dalam mengklasifikasikan data (Chauhan et al., 2018; Tobias et al., 2016). Maka dari itu, pada penelitian ini mengimplementasikan metode CNN sebagai pendeteksi kanker payudara. Metode CNN diharapkan dapat bermanfaat untuk mengetahui kanker pada payudara sehingga dapat mengurangi jumlah kasus kanker tersebut.

## 2. LANDASAN TEORI

### 2.1. Kanker Payudara

Kanker payudara adalah kanker yang berkembang dari jaringan payudara, tanda-tanda kanker payudara dapat berupa benjolan pada payudara, perubahan bentuk payudara, lesung kulit, cairan yang keluar dari puting, atau bercak merah bersisik pada kulit. Pada mereka dengan penyebaran

penyakit yang jauh, mungkin ada nyeri tulang, pembengkakan kelenjar getah bening, sesak napas, atau kulit menguning (Feng et al., 2019).

Faktor penyebab kanker payudara pada wanita adalah obesitas, kurang latihan fisik, minum alkohol, terapi penggantian hormon selama menopause, radiasi pengion, usia dini saat menstruasi pertama, memiliki anak terlambat atau tidak sama sekali, usia lebih tua, dan riwayat keluarga. Sekitar 5–10% kasus disebabkan oleh gen yang diwarisi dari orang tua seseorang, termasuk BRCA1 dan BRCA2. Kanker payudara paling sering berkembang dalam sel dari lapisan saluran susu dan lobulus yang memasok susu ke saluran tersebut (Amkrane et al., 2020; Khasana et al., 2020).

## 2.2. Citra Histopatologi

Citra histopatologi merupakan standart tertinggi pemeriksaan kanker pada pemeriksaan kanker, sebelum pasien tersebut di diagnosa. Pengambilan citra histopatologi sendiri menggunakan mikroskop digital (Hyperastuty, 2017). Citra histopatologi mempunyai beberapa komponen yaitu latar belakang, squamous epithelium (inti sel dan sitoplasma) dan stroma.

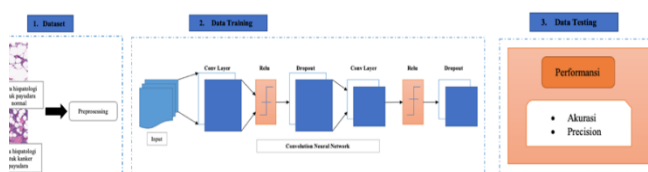
## 2.3. Convolutional Neural Network

Convolutional Neural Network (CNN) adalah sebuah metode pengembangan dari Jaringan Saraf Tiruan (JST) yang biasa digunakan untuk mengelola data citra. Metode CNN dapat digunakan untuk mendeteksi atau mengenali suatu objek pada sebuah citra. Metode ini terdiri dari neuron yang memiliki bobot, bias, dan fungsi aktivasi seperti halnya JST yang lain. Sehingga dapat dikatakan bahwa CNN tidak jauh berbeda dengan JST yang lain

seperti *perceptron* maupun *backpropagation*. Arsitektur jaringan pada CNN secara umum dibagi menjadi dua yaitu *feature extraction* dan *classification* (Alwanda et al., 2020; Tobias et al., 2016).

## 3. METODE PENELITIAN

Pada Gambar 1 terlihat bahwa metode penelitian dibagi menjadi dua bagian yaitu dataset, data training dan data testing. Pada bagian pertama yaitu dataset terdapat proses preprossesing data citra yang akan dikelola pada proses selajutnya. Kemudian untuk tahap training masuk pada tahapan metode CNN dimana ada beberapa bagian yaitu input, *convolution layer*, *pooling layer*, dan *performance*.



Gambar 1. Metode Penelitian

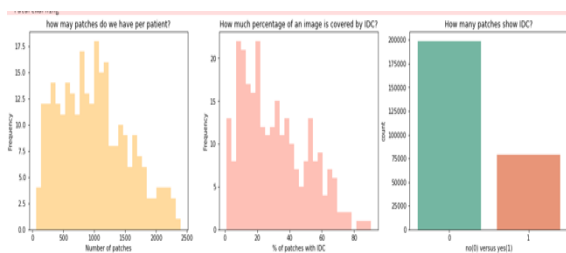
### 3.1. Dataset

Dataset yang digunakan pada penelitian ini adalah citra histopatologi yang diambil dari *Kaggle* <https://www.kaggle.com/paultimothymoon/breast-histopathology-images> yang berupa data *breast histopathology images*. Jumlah data sebanyak 277.524 dengan pengujian data uji sebesar 20% dan 80% sebagai data latih.

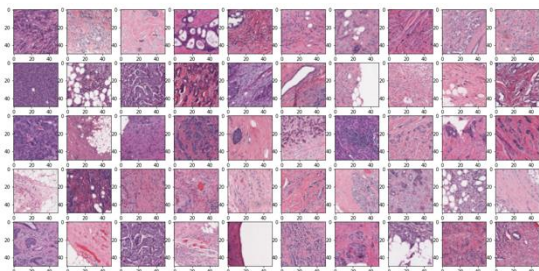
### 3.2. Preprossesing

Proses yang dilakukan pada tahapan preprossesing adalah pengecekan data yang di paparkan dalam bentuk grafik, dapat dilihat pada Gambar 2. Dimana pada grafik pertama dapat dilihat bahwa jumlah tambahan (jaringan baru) gambar per pasien bervariasi banyak. Pada grafik kedua dapat disimpulkan bahwa beberapa pasien

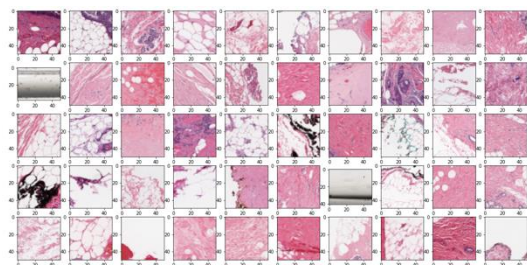
memiliki lebih dari 80% tambahan (jaringan baru) yang menunjukkan *Infiltrating Ductal Carcinomi* (IDC). Akibatnya jaringan tersebut penuh dengan kanker atau hanya sebagian payudara saja yang ditutupi oleh potongan jaringan yang difokuskan pada kanker IDC. Grafik ketiga menunjukkan bahwa kelas *Infiltrating Ductal Carcinomi* (IDC) dan kelas tanpa *Infiltrating Ductal Carcinomi* (IDC) tidak seimbang. Artinya data antara pasien kanker dan pasien tidak kanker tidak seimbang.



Gambar 2. Grafik perbandingan citra IDC dan nonIDC



Gambar 3. Citra Kanker Payudara IDC (+)



Gambar 4. Citra Non-Kanker Payudara IDC (-)

Selanjutnya pengecekan data kanker payudara dan tidak kanker (*non-kanker*) dengan menampilkan 50 data per kondisi. Dapat dilihat pada Gambar 3 dan Gambar 4

bahwa perbedaan warna pada citra kanker dan non kanker. Dimana citra histopologi kanker payudara lebih ungu dan pekat warnanya dibandingkan dengan citra bukan kanker. Citra histopologi payudara *non-kanker* berwarna ungu muda.

### 3.3. Performansi

Proses ini adalah proses menganalisis hasil inputan citra histopologi kanker payudara setelah di proses training, sehingga dapat diketahui apakah citra tersebut termasuk kanker payudara normal atau terdiagnosa kanker. Selain itu pada proses ini juga dilakukan pengukuran keakuratan hasil untuk mengetahui keakuratan deteksi kanker payudara dengan mengukur tingkat akurasi, sensitivitas, spesifitas, dan presisi. *Confusion matriks* merupakan tabel yang umum digunakan untuk mengukur hasil kerja klasifikator seperti yang ditunjukkan pada Tabel 1.

Tabel 1. *Confusion Matriks*

		Kelas Hasil Prediksi	
		Positif	Negatif
Data Asli	Positif	True Positive (TP)	False Negative (FN)
	Negatif	False Positive (FP)	True Negative (TN)

- **Akurasi**  
 Akurasi adalah kemampuan pengujian untuk mengidentifikasi jumlah data yang di deteksi secara benar.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- **Precision**  
 Precision adalah kemampuan pengujian untuk mengidentifikasi hasil yang positif dari

sejumlah data yang terdiagnosa positif oleh sistem.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

#### 4. HASIL DAN PEMBAHASAN

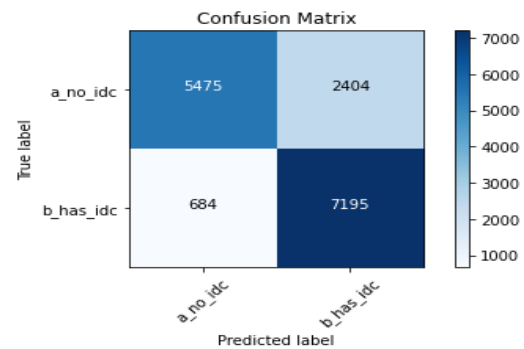
Model modifikasi terdiri dari input, 9 *convolution layer*, 1 *flatten*, dan lapisan *output*. Inputan yang digunakan adalah data *training* kemudian data tersebut di proses pada *convolution layer*. Pada *convolution layer* yang akan digunakan menggunakan fungsi aktivasi ReLU dengan menggunakan operasi *dot*. Sebelum melakukan *pooling layer* harus melewati aktivasi ReLU. *Pooling layer* yang digunakan memakai operasi *max pooling*.

Metode CNN pada tahapan klasifikasi menggunakan 1 *fully connected layer*, yaitu fungsi aktivasi yang digunakan adalah fungsi aktivasi *softmax*. Dimana sebelum masuk pada *fully connected layer* harus dilakukan *flatten* terlebih dahulu yaitu mengubah bentuk multidimensi array menjadi vektor. Pemodelan CNN pada citra kanker payudara ditunjukkan pada Tabel 2.

Tabel 2. Pemodelan CNN pada Citra Kanker Payudara

Layer	Type	Output shapes	Params
<i>conv2d</i>	<i>Conv2D</i>	None, 48x48x32	896
<i>conv2d_1</i>	<i>Conv2D</i>	None, 46x46x32	9248
<i>conv2d_2</i>	<i>Conv2D</i>	None, 44x44x32	9248
<i>max_pooli ng2d</i>	<i>MaxPoolin g2D</i>	None, 22x22x32	0
<i>dropout</i>	<i>Dropout</i>	None, 22x22x32	0

<i>conv2d_3</i>	<i>Conv2D</i>	None, 20x20x64	18496
<i>conv2d_4</i>	<i>Conv2D</i>	None, 18x18x64	36928
<i>conv2d_5</i>	<i>Conv2D</i>	None, 16x16x64	36928
<i>max_pooli ng2d_1</i>	<i>MaxPoolin g2D</i>	None, 8x8x64	0
<i>dropout_1</i>	<i>Dropout</i>	None, 8x8x64	0
<i>conv2d_6</i>	<i>Conv2D</i>	None, 6x6x128	73856
<i>conv2d_7</i>	<i>Conv2D</i>	None, 4x4x128	147584
<i>conv2d_8</i>	<i>Conv2D</i>	None, 2x2x128	147584
<i>max_pooli ng2d_2</i>	<i>MaxPoolin g2D</i>	None, 1x1x128	0
<i>dropout_2</i>	<i>Dropout</i>	None, 1x1x128	0
<i>flatten</i>	<i>Flatten</i>	None, 128	0
<i>dense</i>	<i>Dense</i>	None, 256	33024
<i>dropout_3</i>	<i>Dropout</i>	None, 256	0
<i>dense_1</i>	<i>Dense</i>	None, 2	514
<b>Total params</b>			<b>514,306</b>
<b>Trainable params</b>			<b>514,306</b>
<b>Nontrainable params</b>			<b>0</b>



Gambar 5. Hasil Confusion Matriks

Dari hasil pengujian yang didapatkan dapat kita hitung secara matematis seperti yang ditunjukkan Gambar 5. dengan menggunakan persamaan 1 dan persamaan 2, sehingga didapatkan presentase keakuratan sebagai berikut:

- Akurasi yang di dapatkan pada penelitian ini adalah 80%, artinya apabila pasien melakukan uji diagnosa untuk deteksi kanker payudara maupun normal akan memiliki peluang terdeteksi benar sebesar 80%.

$$Akurasi = \frac{5475 + 7195}{5475 + 7195 + 2404 + 684} \times 100\% = 80\%$$

- Pada penelitian ini mendapatkan nilai presisi dari hasil klasifikasi sebesar 75% untuk pasien yang menderita kanker payudara (IDC).

$$Precision = \frac{5475}{5475 + 2404} \times 100\% = 75\%$$

## 5. SIMPULAN DAN SARAN

Pada penelitian ini menggunakan metode CNN untuk menganalisa jaringan IDC pada citra histopathologi kanker payudara. Dengan menggunakan model modifikasi terbaik nilai akurasi yang didapatkan untuk mendeteksi kanker payudara adalah 80%. Arsitektur yang digunakan ditunjang oleh kumpulan data 277.524, dengan ukuran citra 50x50 piksel RGB. Hasil akurasi yang didapatkan pada penelitian ini termasuk dalam kategori baik, sehingga dapat membantu dalam mengurangi kesalahan proses diagnosa kanker payudara. Keterbatasan utama pada penelitian ini adalah dataset yang digunakan, dimana merupakan data sekunder dari *Kaggle*. Diharapkan untuk penelitian selanjutnya menggunakan data primer sehingga lebih akurat untuk mengidentifikasi adanya kanker pada payudara.

## 6. UCAPAN TERIMA KASIH

Tim peneliti mengucapkan terima kasih dan penghargaan kepada Fakultas Teknik, Universitas Muhammadiyah Malang atas dukungan terselenggaranya pekerjaan ini melalui skema Pusat Kajian dan Rekayasa Teknik PUSKAREKA 2021.

## DAFTAR PUSTAKA

- Alwanda, M. R., Ramadhan, R. P. K., & Alamsyah, D. (2020). Implementasi Metode Convolutional Neural Network Menggunakan Arsitektur LeNet-5 untuk Pengenalan Doodle. *Jurnal Algoritme*, 1(1), 45–56. <https://doi.org/10.35957/algoritme.v1i1.1434>
- Amkrane, Y., El Adoui, M., & Benjelloun, M. (2020). Towards Breast Cancer Response Prediction using Artificial Intelligence and Radiomics. *Proceedings of 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications, CloudTech 2020*. <https://doi.org/10.1109/CloudTech49835.2020.9365890>
- Azis, A. I. S., Irma Surya Kumala Idris, Budy Santoso, & Yasin Aril Mustofa. (2019). Pendekatan Machine Learning yang Efisien untuk Prediksi Kanker Payudara. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(3), 458–469. <https://doi.org/10.29207/resti.v3i3.1347>
- Binabar, S. W., & Ivandari. (2017). Optimasi Parameter K pada Algoritma KNN untuk Deteksi Penyakit Kanker Payudara. *IC-Tech, XII(2)*, 11–18.
- Chauhan, R., Ghanshala, K. K., & Joshi, R. C. (2018). Convolutional Neural Network (CNN) for Image Detection and Recognition. *ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications, December 2018*,

- 278–282.  
<https://doi.org/10.1109/ICSCCC.2018.8703316>
- Darsyah, M. Y. (2013). Menakar Tingkat Akurasi Support Vector Machine Study Kasus Kanker Payudara. *Statistika Universitas Muhammadiyah*, 1(1), 15–20.
- Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307.  
<https://doi.org/10.1109/TPAMI.2015.2439281>
- Eka Putra, W. S. (2016). Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101. *Jurnal Teknik ITS*, 5(1).  
<https://doi.org/10.12962/j23373539.v5i1.15696>
- Feng, X., Song, L., Wang, S., Song, H., Chen, H., Liu, Y., Lou, C., Zhao, J., Liu, Q., Liu, Y., Zhao, R., Xing, K., Li, S., Yu, Y., Liu, Z., Yin, C., Han, B., Du, Y., Xin, R., ... Zhou, F. (2019). Accurate prediction of neoadjuvant chemotherapy pathological complete remission (PCR) for the four sub-types of breast cancer. *IEEE Access*, 7, 134697–134706.  
<https://doi.org/10.1109/ACCESS.2019.2941543>
- Hyperastuty, A. S. (2017). Artificial Neural Network dalam Menentukan Grading Histopatologi Kanker Payudara. *Jurnal Biosains Pascasarjana*, 19(2), 176.  
<https://doi.org/10.20473/jbp.v19i2.2017.176-188>
- Janghel, R. R., Shukla, A., Tiwari, R., & Kala, R. (2010). Breast Cancer diagnosis using Artificial Neural Network models. *Proceedings - 3rd International Conference on Information Sciences and Interaction Sciences, ICIS 2010, CI*, 89–94.  
<https://doi.org/10.1109/ICICIS.2010.5534716>
- Khasana, U., Sigit, R., & Yuniarti, H. (2020). Segmentation of Breast Using Ultrasound Image for Detection Breast Cancer. *IES 2020 - International Electronics Symposium: The Role of Autonomous and Intelligent Systems for Human Life and Comfort*, 584–587.  
<https://doi.org/10.1109/IES50839.2020.9231629>
- Lestandy, M., Abdurrahim, A., & Syafa'ah, L. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naive Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 802–808.
- Ma, X., Yang, G., & Yang, Q. (2019). Application of Deep Convolution Neural Network in Automatic Classification of Land Use. *Journal of Physics: Conference Series*, 1187(4).  
<https://doi.org/10.1088/1742-6596/1187/4/042104>
- Puttagunta, M., & Ravi, S. (2021). Medical image analysis based on deep learning approach. *Multimedia Tools and Applications*, 80(16), 24365–24398.  
<https://doi.org/10.1007/s11042-021-10707-4>
- Rokhana, R., Priambodo, J., Karlita, T., Sunarya, I. M. G., Yuniarno, E. M., Purnama, I. K. E., & Purnomo, M. H. (2019). Convolutional Neural Network untuk Pendeteksian Patah Tulang Femur pada Citra Ultrasonik B-Mode. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(1), 59.  
<https://doi.org/10.22146/jnteti.v8i1.491>
- Sun, Y., Zhang, W., Gu, H., Liu, C., Hong, S., Xu, W., Yang, J., & Gui, G. (2019). Convolutional Neural Network Based Models for Improving Super-Resolution Imaging. *IEEE Access*, 7(1), 43042–43051.  
<https://doi.org/10.1109/ACCESS.2019.2941543>

9.2908501

- Syafa'ah, L., Zulfatman, Z., Pakaya, I., & Lestandy, M. (2021). Comparison of Machine Learning Classification Methods in Hepatitis C Virus. *Jurnal Online Informatika*, 6(1), 73. <https://doi.org/10.15575/join.v6i1.719>
- Syafa, L., & Lestandy, M. (2021). Penerapan Deep Learning untuk Prediksi Kasus Aktif Covid-19. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(1), 453–457.
- Tobias, L., Ducournau, A., Rousseau, F., Mercier, G., & Fablet, R. (2016).

Convolutional Neural Networks for object recognition on mobile devices: A case study. *Proceedings - International Conference on Pattern Recognition*, 0, 3530–3535. <https://doi.org/10.1109/ICPR.2016.7900181>