

## Penerapan Model CRISP-DM Untuk Klasifikasi Tumor Otak Menggunakan Algoritma CNN

Yuri Yuliani<sup>1</sup>, Hamdun Sulaiman<sup>2\*</sup>, Yanto<sup>3</sup>, Ibnu Al Farobi<sup>4</sup>, Kukuh Panggalih<sup>5</sup>

<sup>1,2</sup>Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

<sup>3,4</sup>Program Studi Informatika, Universitas Bina Sarana Informatika

<sup>5</sup>Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika

\*[hamdun.hsl@bsi.ac.id](mailto:hamdun.hsl@bsi.ac.id)

### Abstrak

Secara umum tumor adalah pertumbuhan jaringan yang tidak normal di dalam tubuh yang terjadi akibat pembelahan sel yang tidak terkendali. Tumor dapat bersifat jinak (non kanker) atau ganas (kanker). Tumor jinak cenderung tumbuh secara lokal tanpa menyebar ke jaringan lain, sementara tumor ganas dapat menyerang jaringan sekitarnya dan menyebar ke bagian tubuh lain melalui aliran darah. Tumor otak adalah massa atau pertumbuhan jaringan abnormal di dalam atau di sekitar otak. Secara spesifik tumor otak dapat bersifat primer (berasal dari sel-sel otak itu sendiri atau jaringan di sekitarnya, seperti meninges, kelenjar hipofisis, atau saraf kranial) atau sekunder (metastasis dari kanker yang berasal dari bagian tubuh lain). Masalah utama saat ini adalah tumor otak memiliki banyak jenisnya, maka dari itu penelitian tentang klasifikasi jenis tumor otak sangat berharga saat ini. Agar dapat meningkatkan akurasi dan juga mempercepat proses diagnosis. Hasil penelitian ini, telah dilakukan klasifikasi data citra tumor otak menggunakan metode Algoritma CNN dengan teknik data mining CRISP-DM. Dataset terdiri dari 3.554 dataset, namun dari hasil pengujian algoritma ini memiliki kinerja yang baik dengan akurasi sebesar 98% dengan Evaluasi dan validasi menggunakan parameter uji Confusion Matrix. dan berhasil mengklasifikasikan otak manusia yang terkena tumor dan tidak terkena tumor, selain itu dapat mengklasifikasi 3 jenis tumor yang berbeda yaitu glioma, meningioma, dan pituitary.

Kata kunci : Algoritma CNN, Klasifikasi, Tumor Otak, CRISP-DM

### Abstract

*In general, a tumor is an abnormal tissue growth in the body that occurs due to uncontrolled cell division. Tumors can be benign (non-cancerous) or malignant (cancerous). Benign tumors tend to grow locally without spreading to other tissues, while malignant tumors can invade surrounding tissues and spread to other parts of the body through the bloodstream. A brain tumor is a mass or growth of abnormal tissue in or around the brain. Specifically, brain tumors can be primary (originating from the brain cells themselves or surrounding tissues, such as the meninges, pituitary gland, or cranial nerves) or secondary (metastasis from cancer originating from other parts of the body). The main problem today is that brain tumors have many types, therefore research on the classification of brain tumor types is very valuable today. In order to improve accuracy and also speed up the diagnosis process. The results of this study, brain tumor image data classification has been carried out using the CNN Algorithm method with the CRISP-DM data mining technique. The dataset consists of 3,554 datasets, but from the results of testing this algorithm has good performance with an accuracy of 98% with Evaluation and validation using the Confusion Matrix test parameters. and successfully classified human brains affected by tumors and those not affected by tumors, in addition to being able to classify 3 different types of tumors, namely glioma, meningioma, and pituitary.*

Keywords : CNN Algorithm, Classification, Brain Tumor, CRISP-DM.

### 1. Pendahuluan

Tumor otak merupakan salah satu kondisi kesehatan kritis yang memiliki dampak signifikan

pada kualitas hidup pasien. Identifikasi dan klasifikasi dini tumor otak menjadi kunci dalam menentukan strategi pengobatan yang efektif.

Teknologi pencitraan medis seperti Magnetic Resonance Imaging (MRI) telah memainkan peran penting dalam mendeteksi dan menganalisis jenis tumor otak secara non-invasif. Namun, kompleksitas struktur otak dan sifat heterogen tumor otak menantang akurasi klasifikasi berbasis MRI secara tradisional [1].

Dalam skala global, menurut laporan Globocan 2020 yang diterbitkan oleh WHO, insiden tumor otak ganas primer dilaporkan mencapai sekitar 14,8 kasus per 100.000 orang per tahun, dengan tingkat prevalensi yang lebih tinggi pada pria dibandingkan wanita untuk jenis tumor ganas[2]. Di Indonesia, jumlah keseluruhan kasus kanker baru pada tahun 2020 tercatat sebanyak 396.914 kasus, dengan 234.511 kematian akibat kanker. Meskipun data spesifik prevalensi tumor otak di Indonesia belum dirinci, kanker otak termasuk salah satu penyakit dengan tingkat morbiditas signifikan [3].

Risiko kanker otak di Indonesia paling banyak terjadi pada kelompok usia anak-anak dan dewasa muda (20–40 tahun), dengan prognosis yang buruk apabila tidak segera mendapatkan penanganan. Tantangan dalam diagnosis dan klasifikasi tumor otak ini mendorong pengembangan teknologi berbasis pembelajaran mesin (machine learning) dan jaringan saraf dalam (deep learning). Teknologi ini memungkinkan pengolahan data MRI dalam volume besar untuk mengklasifikasikan berbagai

jenis tumor, seperti glioma, meningioma, dan tumor pituitary, dengan tingkat akurasi yang semakin meningkat [2]

Penelitian ini bertujuan untuk evaluasi model CNN dan mengklasifikasikan tumor otak berdasarkan data citra MRI. Model ini diharapkan mampu membedakan antara beberapa jenis tumor otak yang umum, dengan tingkat akurasi yang tinggi dan waktu pemrosesan yang cepat

## **2. Tinjauan Pustaka**

### **2.1. Penelitian Terkait**

Penulis melakukan analisa terhadap penelitian sebelumnya, melihat metode apa saja yang telah digunakan, dan melihat bagaimana penelitian sebelumnya pernah dilakukan yang berhubungan dengan klasifikasi tumor otak, terdapat 3 penelitian terkait:

- Penelitian pertama yang dilakukan oleh (Prasetyo & Nabiilah, 2023) dengan judul “Perbandingan Model Machine Learning Pada Klasifikasi Tumor Otak Menggunakan Fitur Discrete Cosine Transform”, dengan hasil “Tingkat akurasi model yang tinggi dicapai dengan menguji berbagai arsitektur model machine learning yang ada, diikuti dengan memasukkan feature map yang diekstraksi dari Discrete Cosine Transform (DCT). Klasifikasi gambar MRI mencapai akurasi pada data test tertinggi sebesar 93% dengan

- menggunakan model Support Vector Machine (SVM) [4].
- Penelitian kedua yang diteliti oleh (Adilah & Azizah, 2022), dengan judul “Klasifikasi Tumor Otak Menggunakan Ekstraksi Fitur HOG dan Support Vector Machine”, dengan hasil penelitian ini, proses klasifikasi dilakukan dalam tiga langkah, Pre processing yang mengubah ukuran citra, ekstraksi fitur yang mengambil informasi (fitur), kemudian diikuti dengan melatih data dan dilakukan pengujian pada SVM. Hasil dari penelitian ini adalah untuk mengklasifikasikan jenis Tumor Otak. Dari hasil pengujian klasifikasi dengan SVM, didapatkan tingkat akurasi yang cukup baik sebesar yaitu 91% [5].
  - Penelitian ketiga yang dilakukan oleh (Saiful, et.al, 2021), dengan judul “Sistem Deteksi Infeksi COVID-19 Pada Hasil X-Ray Rontgen menggunakan Algoritma Convolutional Neural Network (CNN)” tujuan dari penelitiannya adalah mengetahui bagaimana perancangan sistem, arsitektur jaringan yang digunakan untuk pendeteksian infeksi COVID-19. Sistem ini tidak dapat melakukan pendeteksian objek yang lain. Hasil deteksi infeksi COVID-19 dengan algoritma convolutional neural network menunjukkan nilai akurasi yang unlimited yaitu berkisar di angka 60-99% [6].
  - Penelitian ke empat oleh (Laksono, et.al, 2023), dengan judul penelitian “Deteksi Tumor Otak Melalui Penerapan GLCM dan NaïveBayes Classification”. Tujuan dari penelitian ini adalah mengetahui nilai akurasi yang dihasilkan oleh GLCM dan Naïve Bayes Classification dalam mempreiksi citra tumor otak yang didapatkan. Berdasarkan hasil uji yang telah dilakukan terhadap 253 dataset pasien diketahui bahwa nilai akurasi yang dihasilkan oleh Naïve Bayes dalam melakukan klasifikasi ternyata cukup tinggi yaitu 80% dengan perbandingan data testing dan data trainingnya adalah 20% :80%, selain itu rata-rata dari nilai presisi dan recall yang dihasilkan pun sama yaitu 85% [7].
  - Penelitian kelima yang dilakukan oleh (Wasil, et.al, 2022), dengan judul “Pengaruh Epoch pada Akurasi menggunakan Convolutional Neural Network untuk Klasifikasi fashion dan Furniture”. Bahwa Metode Convolutional Neural Network (CNN) adalah metode yang sangat tepat untuk klasifikasi gambar karena terdapat banyak arsitektur yang bisa digunakan. Hasil penelitian ini menunjukkan bahwa akurasi terbaik pada furniture=94.18% dengan penggunaan epoch=500, dan fashion =99.15% dengan penggunaan epoch=1500 [8].
  - Penelitian selanjutnya oleh (Winnarto et al., 2024), dengan judul “Ekstraksi Fitur Dan

Implementasi Machine Learning Untuk Klasifikasi Jenis Tumor Otak”, dengan hasil “Hasil pengujian menggunakan empat metode ekstraksi fitur dan tujuh metode klasifikasi, performa terbaik untuk klasifikasi jenis tumor otak adalah menggunakan Global Feature Extraction dan klasifikasi Random Forest dengan tingkat akurasi yang baik yaitu sebesar 88% [9].

## 2.2. Landasan Teori

### 1. Tumor Otak

Tumor otak merupakan sel-sel abnormal yang berkembang pada ruang intrakranial dengan tidak wajar dan berbahaya. Tumor otak dapat mempengaruhi Self-efficacy dan kualitas hidup serta dapat mengancam jiwa. Tumor otak merupakan hasil dari proliferasi sel yang tidak terkontrol yang berasal dari jaringan saraf atau jaringan suportif di dalam otak [3].

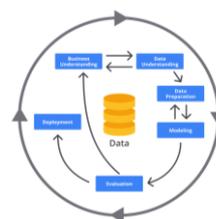
### 2. Data Mining

Data mining adalah serangkaian teknik ekstraksi pengetahuan yang berguna untuk menemukan pola khusus dari kumpulan data sehingga menghasilkan informasi nilai tambah yang sebelumnya tidak diketahui [5].

### 3. Model CRISP-DM

CRISP-DM (Cross Industry Standard for Data mining) merupakan metode yang banyak digunakan oleh para ahli dengan menggunakan proses pemodelan data di dalamnya, Tujuan dari

metode CRISP-DM ini yaitu untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan. CRISP-DM memiliki tahapan dan kerangka kerja yang terstruktur sehingga pengguna metode ini akan lebih terarah dan mengetahui langkah yang harus dikerjakan dalam penelitian [6].



Gambar 1. Model CRISP-DM

### 4. Klasifikasi

Dalam klasifikasi, terdapat dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai prototype untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpannya [10].

### 5. Deep Learning

Deep Learning adalah suatu teknik pada machine learning yang dapat mengolah informasi nonlinier dengan menggunakan banyak lapisan untuk menjalankan identifikasi pola, ekstraksi fitur dan klasifikasi yang merupakan metode studi representasi yang memungkinkan model perhitungan komputasi terdiri dari banyak layer

pengolahan dengan mempelajari data dari banyak tingkat abstraksi [11],[12]

## 6. Algoritma CNN

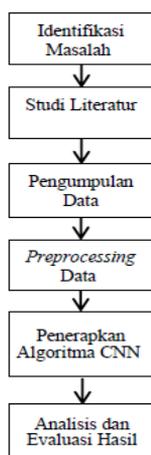
CNN merupakan neural networks yang digunakan dalam pengolahan data berbentuk citra yang dapat bekerja menggunakan kernel yang dimana kernel mengekstrak fitur pada data masukan melalui operasi konvolusi [12],[13],[14]

## 3. Metode Penelitian

Penelitian ini bertujuan untuk evaluasi model CNN dan mengklasifikasikan tumor otak berdasarkan data citra MRI.

### 3.1. Tahap Penelitian

Berikut adalah tahapan penelitian untuk klasifikasi tumor otak menggunakan algoritma CNN :



Gambar 2. Tahap Penelitian<sup>[15]</sup>

### 3.2. Sumber Data

Data yang digunakan dalam penelitian ini mencakup data tumor otak yang diambil dari <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>.

## 4. Hasil dan Pembahasan

Pada bagian hasil dan pembahasan ini, akan diuraikan hasil dari setiap tahapan dalam penerapan model CRISP-DM untuk mengklasifikasikan tumor otak dengan implementasi algoritma CNN.

### 1. Business Understanding

Penerapan data mining pada penelitian ini berhubungan langsung dengan data citra tumor otak untuk mengklasifikasikan data tumor otak apa yang masuk kedalam kategori tumor glioma, tumor meningioma, tumor pituitary dan no tumor. Serta untuk melihat hasil akurasi dari penerapan Algoritma CNN.

### 2. Data Understanding

Dataset penelitian merupakan data publik yang diunduh dari kaggle.com tentang citra tumor otak yang bersumber dari <https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>, dengan jenis citra berdasarkan kategori tumor otak tersebut.

### 3. Data Preparation

Jumlah dataset yang digunakan dalam penelitian ini sebanyak 3.554, proses pengolahan dengan menerapkan prinsip data mining, dimana pembagiannya 80% data training sebanyak 2.843

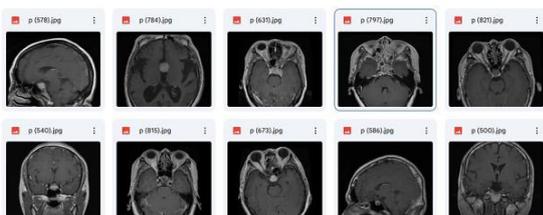
dan 20% data testing sebanyak 711 citra tumor otak. Berikut beberap citra tumor otak :



Gambar 3. Tumor Glioma



Gambar 4. Tumor Meningioma



Gambar 5. Tumor Pituitary



Gambar 6. No Tumor

#### 4. Modelling

Pada tahap ini, proses dimana model dengan teknik klasifikasi yang dipilih menghasilkan suatu pola informasi yang dapat memudahkan. Pola klasifikasi yang dihasilkan teknik data mining ini digunakan untuk mengklasifikasikan citra tumor otak. Adapun tools atau alat bantu dalam proses

pengolahan data mining yaitu menggunakan Visual Code Studio, dengan menambahkan extensions python serta menginstal beberapa library untuk proses pemodelan dan pengolahannya. Tahap ini menampilkan serta memberikan informasi terhadap kinerja Algoritma CNN dalam metode klasifikasi.

```
# Calculate the bar plot for each label
label_counts = {label: np.sum(y_train == label) for label in labels}

# Plot the bar plot and sample images in one chart
plt.figure(figsize=(8, 6))

colors = ["C0", "C1", "C2", "C3"]

# Plot the histogram
plt.subplot(2, 1, 1)
bars = plt.bar(label_counts.keys(), label_counts.values(), color=colors)
mplyberpump.add_bar_gradient(bars=bars)
# plt.xlabel('Labels')
plt.ylabel('Count')
plt.title('Distribution of Labels')

# Plot sample images from each label
k = 0
for i in labels:
    j = 0
    while True:
        if y_train[j] == i:
            plt.subplot(2, 4, k + 5)
            plt.imshow(X_train[j])
            plt.axis('off')
            k += 1
            break
        j += 1

plt.tight_layout()
plt.show()
```

Gambar 7. Proses Klasifikasi Citra Tumor Otak

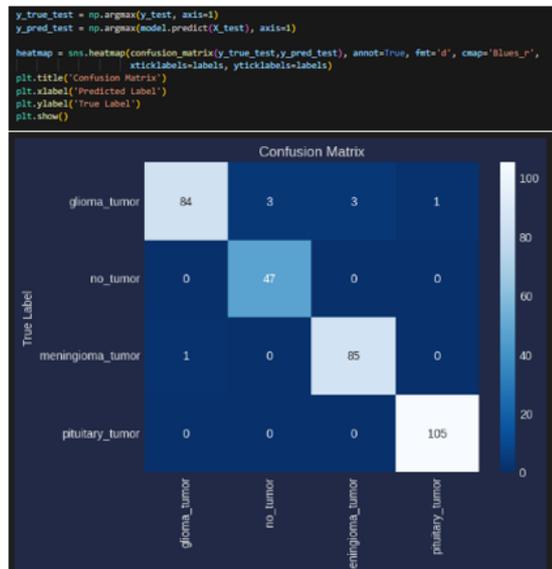


Gambar 8. Hasil Klasifikasi Citra Tumor Otak

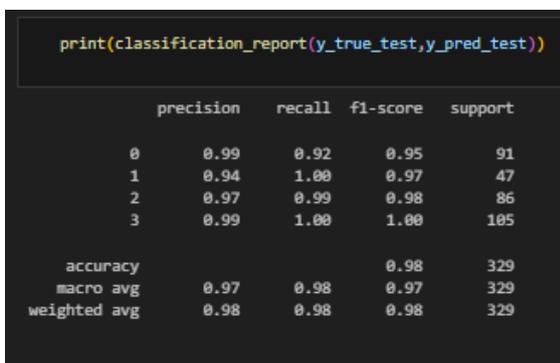
#### 5. Evaluasi

Setelah pola klasifikasi didapatkan pada algoritma CNN, selanjutnya dilakukan tahap evaluasi komparasi algoritma dengan parameter yang digunakan ialah Confusion Matrix yang pada dasarnya untuk memberikan informasi hasil yang telah dilakukan oleh model dengan hasil klasifikasi

sebenarnya dengan melihat nilai akurasi, presisi dan recall. Berikut hasil evaluasi pada model yang telah dibuat dengan algoritma ini.



Gambar 9. Confusion Matrix Pada Algoritma CNN



Gambar 10. Hasil Precision, Recall, Accuracy

## 6. Deployment

Setelah tahap evaluasi dimana menilai secara detail hasil dari sebuah model maka dilakukan pengimplementasian dari keseluruhan model yang telah dibangun. Selain itu juga dilakukan penyesuaian terhadap model sehingga dapat

menghasilkan suatu hasil yang sesuai dengan target awal tahap CRISP-DM ini

## 5. Kesimpulan

Pada penelitian ini, telah dilakukan klasifikasi data citra tumor otak menggunakan metode Algoritma CNN dengan teknik data mining CRISP-DM. Dataset terdiri dari 3.554 dataset, namun dari hasil pengujian algoritma ini memiliki kinerja yang baik dengan akurasi sebesar 98% dengan Evaluasi dan validasi menggunakan parameter uji Confusion Matrix. dan berhasil mengklasifikasikan otak manusia yang terkena tumor dan tidak terkena tumor, selain itu dapat mengklasifikasi 3 jenis tumor yang berbeda yaitu glioma, meningioma, dan pituitary..

## 6. Daftar Pustaka

- [1]. Husen, D. (2024). Klasifikasi Citra MRI Tumor Otak Menggunakan Metode Convolutional Neural Network. Bit-Tech, 7(1),143–152. <https://doi.org/10.32877/bt.v7i1.1576>
- [2]. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. <https://doi.org/10.3322/caac.21660>
- [3]. Iqhrammullah, M., Refin, R. Y., Rasmji, R. I., Andika, F. F., Hajjah, H., Marlina, M., & Ningsih, R. (2023). Cancer in Indonesia: A bibliometric surveillance. Narra X, 1(2). <https://doi.org/10.52225/narrax.v1i2.86>
- [4]. Badža, M. M., & Barjaktarović, M. C. (2020). Classification of brain tumors from mri images using a convolutional neural

- network. Applied Sciences (Switzerland), 10(6).<https://doi.org/10.3390/app10061999>
- [5]. Prasetyo, S. Y., & Nabiilah, G. Z. (2023). PERBANDINGAN MODEL MACHINE LEARNING PADA KLASIFIKASI TUMOR OTAK MENGGUNAKAN FITUR DISCRETE COSINE TRANSFORM. Jurnal Teknologi Terpadu, 9(1), 29–34. <https://journal.nurulfikri.ac.id/index.php/jtt>
- [6]. Laksono, P., Herliana., Prabowo, T., (2023). Deteksi Tumor Otak Melalui Penerapan GLCM dan Naïve Bayes Classification. Jurnal Ilmiah Intech, Vol. 41, No. 48, Mei 2023, pp. 41-48.
- [7]. Wasil, M., Harianto., Fathurrahman (2022). Pengaruh Epoch pada Akurasi menggunakan Convolutional Neural Network untuk Klasifikasi fashion dan Furniture. Infotek : Jurnal Informatika Dan Teknologi, Vol. 5 No. 1, 53-61. <https://dx.doi.org/10.29408/jit.v5i1.4393>
- [8]. Winnarto, M. N., Mailasari, M., & Purnamawati, A. (2024). Ekstraksi Fitur Dan Implementasi Machine Learning Untuk Klasifikasi Jenis Tumor Otak. In Jurnal (Vol. 6, Issue 1). <http://ejournal.bsi.ac.id/ejournal/index.php/infortech65>
- [9]. Bahrain, S., Simbolon, E., & Terusanti, P. (2024). Penyuluhan Gejala Tumor Otak Pada Siswi Smk Negeri 1 Balige
- [10]. Hasim Azari, Dwi Hartanti, & Aprilisa Arum Sari. (2024). Pengelompokan Produksi Padi dan Beras Provinsi Jawa Timur dengan Metode Agglomerative Hierarchical Clustering. Infotek: Jurnal Informatika Dan Teknologi, 7(2), 379–389. <https://doi.org/10.29408/jit.v7i2.26016>
- [11]. Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., Jajuli, M., HSRonggo Waluyo, J., Telukjambe Timur, K., Karawang, K., & Barat, J. (2022). Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. Jurnal Teknologi Dan Informasi.<https://doi.org/10.34010/jati.v12i1>
- [12]. Nurhidayati, Yahya, Fathurrahman, L.M Samsu, Wajizatul Amnia. (2023). Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi). Infotek : Jurnal Informatika Dan Teknologi, 6(1), 177-188. <https://dx.doi.org/10.29408/jit.v6i1.7529>
- [13]. Iskandar Mulyana, D., Ainur Rofik, M., & Ohan Zoharuddin Zakaria, M. (2022). Klasifikasi Kendaraan pada Jalan Raya menggunakan Algoritma Convolutional Neural Network ( CNN ).
- [14]. Al Adilah, T., & Azizah, Q. N. (2022). Klasifikasi Tumor Otak Menggunakan Ekstraksi Fitur HOG dan Support Vector Machine. In Jurnal (Vol. 4, Issue 1). <http://ejournal.bsi.ac.id/ejournal/index.php/infortech45>
- [15]. Saiful, M., Samsu, L.M., Fathurrahman (2021). Sistem Deteksi Infeksi COVID-19 Pada Hasil X-Ray Rontgen menggunakan Algoritma Convolutional Neural Network (CNN). Infotek : Jurnal Informatika Dan Teknologi, Vol. 4 No. 2, 217-227. <https://dx.doi.org/10.29408/jit.v4i2.3582>