PROPOSAL SKRIPSI

DETEKSI COVID-19 BERDASARKAN HASIL RONTGEN DADA(CHEST X-RAY) MENGGUNAKAN PYTHON



DISUSUN OLEH:

PEBRI ANDHI HERRY PRATAMA

DBC 116 039

PROGRAM STUDI/JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS PALANGKA RAYA

2020

LEMBAR PERSETUJUAN

DAFTAR ISI

DDODOG	SAL SKRIPSI	:						
	R TABEL							
	R GAMBAR							
		_						
PENDAF	HULUAN	6						
1.1.	LATAR BELAKANG	6						
1.3.	BATASAN MASALAH	8						
1.4.	TUJUAN	9						
1.5.	MANFAAT	9						
1.6.	SISTEMATIKA PENULISAN	9						
1.7	JADWAL SKRIPSI	11						
BAB II		12						
LANDAS	SAN TEORI	12						
2.1.	TINJAUAN PUSTAKA	12						
2.2	TEORI PENDUKUNG	21						
2.2.	1 Computer Vision	21						
2.2.2	2 Machine Learning	21						
2.2.	3 Deep Learning	24						
2.2.4	4 Convolutional Neural Network (CNN)	27						
2.2.3	5 Citra X-Ray	31						
2.2.	6 Thorax	32						
BAB III .		34						
METOD	OLOGI PENELITIAN	34						
3.1	Populasi dan Sampel	34						
3.2	Variabel dan Definisi Operasional	34						
3.3	Jenis dan Sumber Data	34						
3.4	Metode Analisis Data	34						
3.5	.5 Tahapan penelitian35							
3.6								
3.7	Perangkat Pengembangan dan Pengujian	36						
DAFTAF	R PUSTAKA	38						

DAFTAR TABEL

Tabel 1. 2. Rancangan jadwal penelitian	. 11
Tabel 3. 1. Definisi Operasional Variabel	. 34

DAFTAR GAMBAR

Gambar 2. 1. Perbandingan Pemrograman Tradisional dengan Machi.	ne Learning
	22
Gambar 2. 2. Tipe-tipe machine learning	24
Gambar 2. 3. Layer-layer pada <i>Deep Learning</i>	25
Gambar 2. 4. Visual object recognition pada deep learning	26
Gambar 2. 5. Contoh CNN dengan banyak lapisan	28
Gambar 2. 6. Operasi pada konvolusi	29
Gambar 2. 7. Operasi pada max pooling	30
Gambar 2. 8. Gambar citra x-ray thorax	32

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Saat ini sistem dengan bantuan komputer sudah banyak membantu manusia di berbagai bidang, salah satunya di bidang kesehatan yaitu medical informatics. Health informatics disebut juga healthcare informatics, medical informatics, nursing informatics, clinical informatics, atau biomedical informatics adalah teknik informasi yang diterapkan dalam bidang perawatan kesehatan, pada dasarnya pengelolaan dan penggunaan informasi perawatan kesehatan pasien. Salah satu teknologi yang kini mulai banyak di aplikasikan pada medical informatics di beberapa belahan dunia adalah teknologi Deep Learning.

Deep learning merupakan cabang ilmu dari Machine Learning sebagai pengembangan dari Multilayer Percepton (MLP) yang dirancang untuk pengolahan data dalam bentuk dua dimesi. CNN merupakan salah satu algoritma dari Deep learning yang saat ini menjadi salah satu topik hangat sebagai alat yang kuat untuk analisis visual dan telah menunjukkan kemampuannya di berbagai bidang seperti analisis visual dan pengolahan suara. CNN mampu mengklasifikasikan citra dengan tingkat akurasi tinggi karena dapat mengurangi sejumlah parameter bebas serta dapat menangani deformasi gambar input seperti translasi, rotasi dan skala. CNN telah banyak digunakan oleh peneliti dalam berbagai bidang salah satunya yaitu mengenai pengenalan pola ataupun klasifikasi pada data citra sebuah gambar.

Sinar-X atau sinar Röntgen adalah salah satu bentuk dari radiasi elektromagnetik dengan panjang gelombang berkisar antara 10 nanometer ke 100 pikometer (sama dengan frekuensi dalam rentang 30 petahertz – 30 exahertz) dan memiliki energi dalam rentang 100 eV – 100 Kev, hal inilah nantinya yang membuat hasil foto x-ray menampakkan perbedaan warna dari putih, abu-abu, hingga hitam yang dicetak pada pelat film untuk menghasilkan gambar berupa citra tubuh manusia (foto rontgen). Sinar X dapat dimanfaatkan untuk mengetahui kondisi tubuh bagian dalam (tulang) yakni dalam dunia kedokteran disebut sinar

rontgen. X-rays memiliki fungsi penting dalam tiga bidang yaitu diagnosis, perawatan dan pemeriksaan ulang. Sebuah rontgen pada dada (Chest X-ray/Thorax X-ray), mengambil gambar dari jantung, paru, pembuluh darah, saluran pernafasan dan kelenjar getah bening di daerah ini. Tulang dada, tulang rusuk, dan tulang selangka juga masuk dalam gambar rontgen dada. Gambar ini juga menunjukkan bagian atas tulang belakang. Rontgen dada digunakan untuk mendiagnosis masalah kesehatan yang menimbulkan gejala di daerah dada seperti masalah paru dan pernafasan serta gangguan jantung. Citra dari sinar-x yang dihasilkan dapat menjadi rujukan utama dalam menentukan jenis penyakit yang dilakukan oleh dokter.

Studi baru dari Cina menyarankan *chest radiograph* (X-rays) dan *chest computed tomography* (CT) scans dapat membantu mendiagnosa COVID-19. Hal ini dinyatakan bahwa *chest radiographs* (X-rays) dan chest computed tomography (CT) scans adalah metode yang tepat untuk mendeteksi infeksi paru-paru akibat dari COVID-19. Hal merupakan sebuah tantangan bagi di bidang kesehatan untuk menyediakan ahli klinisi ke setiap rumah sakit karena keterbatasan jumlah ahli Radiologi.

Pemeriksaan chest radiograph adalah salah satu pemeriksaan pencitraan medis yang paling sering digunakan karena lebih terjangkau. Pembacaan foto chest radiograph memiliki kekurangan, yaitu sulit terdeteksinya penyakit, sehingga memerlukan waku lama sebelum tenaga medis atau dokter mendiagnosis penyakit yang diderita oleh pasien. Salah satu metode untuk mengatasi masalah tersebut adalah dengan mengklasifikasikan citra rontgen dada ke dalam kelas tertentu dengan menggunakan Deep Learning. Data yang berupa citra chest radiograph, merupakan data gambar dua dimensi yang dapat diproses menggunakan deep learning dengan metode Convolutional Neural Network (CNN). Kemampuan CNN di klaim sebagai model terbaik untuk memecahkan permasalahan object detection dan object recognition pada data citra sebuah gambar hasil rontgen dada.

Berdasarkan permasalahan tersebut penulis mengusulkan sebuah model yang dapat membaca hasil rontgen dada untuk diagnosis COVID-19 dengan

menggunakan tiga model CNN, yaitu VGG19, MobileNet, dan Xception. Dari ketiga model CNN tersebut akan dibandingkan model yang mana yang cocok digunakan untuk diagnosis COVID-19 dari hasil rontgen dada. Oleh karena itu, peneliti membuat penelitian yang berjudul "Deteksi Covid-19 Berdasarkan Hasil Rontgen Dada (Chest X-Ray) Menggunakan Python". Sehingga teknologi AI dalam Medical Informatics dapat menjadi alat bantu seorang radiologis ataupun dokter untuk mendapatkan diagnosis yang akurat. Metode yang di usulkan ini hanya memerlukan gambar hasil rontgen dada untuk digunakan sebagai dataset pelatihan dan pengujian.

1.2. RUMUSAN MASALAH

Berdasarkan latar belakang di atas maka dirumuskan permasalahan sebagai berikut

- a. Bagaimana diagnosa penyakit berdasarkan klasifikasi data foto ronsen normal, positif terinfeksi COVID-19, Tuberculosis (TBC), dan phenomonial menggunakan bahasa pemrograman Python?
- b. Bagaimana implementasi metode *Deep Learning* dengan tiga model CNN berdasarkan gambar hasil Rontgen (X-ray) pada dada manusia?
- c. Bagaimana tingkat akurasi yang didapatkan dari hasil deteksi menggunakan tida model CNN?
- d. Apa hasil output dari model aplikasi yang telah dibuat?

1.3. BATASAN MASALAH

Permasalahan dalam diagnosis penyakit Gigi dan Mulut cukup beragam sehingga dalam pengerjaan tugas akhir ini terdapat beberapa batasan sebagai berikut:

- a. Dataset yang digunakan untuk data latih berupa gambar hasil rontgen dada normal, pneumonia, dan positif terinfeksi COVID-19.
- b. Model CNN yang digunakan adalah yaitu VGG19, MobileNet, dan Xception.

c. Bahasa Pemrograman yang digunakan adalah Python.

1.4. TUJUAN

Tujuan dari penelitian ini adalah sebagai berikut:

- a. Mengetahui bagaimana diagnosa penyakit berdasarkan klasifikasi data foto rontgen dada berdasarkan positif COVID-19, Tuberculosis (TBC), atau phenomonial menggunakan bahasa pemrograman Python.
- b. Mengetahui bagaimana implementasi tiga model model *CNN* dalam rontgen dada.
- c. Mengetahui perbandingan akurasi dari tiga model CNN yang cocok digunakan untuk deteksi COVID-19?
- d. Output menampilkan gambar hasil rontgen dada dengan label penyakit yang diderita beserta berapa tingkat akurasinya.

1.5. MANFAAT

Manfaat dari penelitian ini adalah sebagai berikut:

- a. Dapat memudahkan suatu pihak khusunya pihak kesehatan dalam mendeteksi penyakit dari hasil Rontgen dada pasien dengan bantuan teknologi melalui model hasil pengujian penelitian ini.
- b. Hasil penelitian ini dapat dijadikan acuan untuk penelitian lebih lanjut yang berbasis pada pendeteksian objek pada suatu gambar lain.

1.6. SISTEMATIKA PENULISAN

Sistematika dalam penulisan proposal skripsi ini adalah sebagai berikut :

- a. Cover mencakup judul proposal skripsi, nama mahasiswa, NIM, lambang Universitas Palangka Raya. Ukuran lambang Universitas Palangka Raya adalah lebar dan tingginya 5cm.
- b. Halaman persetujuan dari Dosen Pembimbing I.
- c. Daftar isi
- d. BAB I PENDAHULUAN

Bab ini berisikan Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan, Manfaat, Sistematika Penulisan, Jadwal skripsi yang mencakup kegiatan, rincian kegiatan dan waktu yang diperlukan untuk menyelesaikan setiap kegiatan tersebut. Jadwal dapat disajikan dalam bentuk matriks.

e. BAB II LANDASAN TEORI

Bab ini berisikan Tinjauan Pustaka 10 (sepuluh) tahun terakhir minimal 3 (tiga) diluar judul skripsi, dan Teori-teori pendukung.

f. BAB III METODOLOGI PENELITIAN

Bab ini berisikan Daftar Pustaka 10 (sepuluh) tahun terakhir dan Lampiran.

1.7 JADWAL SKRIPSI

Perkiraan waktu yang dibutuhkan untuk menyelesaikan penelitian diperlihatkan oleh tabel 1.1 di bawah.

Tabel 1. 1. Rancangan jadwal penelitian

No.	Kegiatan	Bulan																			
			Agu	stus		S	epte	mbe	er		Okto	ber		N	love	mbe	r	I	Dese	mbe	r
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Studi Pustaka																				
2	Pengumpulan dan																				
	Pengolahan Data																				
3	Implementasi																				
	sistem																				
4	Interpretasi Hasil																				
5	Pengujian																				
6	Penulisan laporan																				

BAB II

LANDASAN TEORI

2.1. TINJAUAN PUSTAKA

Tinjauan Pustaka (*Literature Review*) merupakan ulasan tentang penelitian yang telah dilakukan sebelumnya yang berkaitan dengan permasalahan dalam penelitian dan bertujuan untuk mendapatkan landasan teori yang dapat mendukung pemecahan masalah yang sedang diteliti. Adapun 4 (Tiga) *literature review* yang memiliki hubungan searah dengan penelitian ini adalah sebagai berikut:

Penelitian Pertama adalah dari Muhamad Fajar Suryana, Fauziah, Ratih Titi Komala Sari pada tahun 2020 mahasiswa Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional yang berjudul "Implementasi Sistem Pakar Menggunakan Metode Certainty Factor Untuk Mendiagnosa Dini Corona Virus Desease (COVID-19)". Dalam penelitiannya menghasilkan aplikasi Untuk membantu para tenaga medis dalam mendeteksi dini gejala-gejala yang dialami oleh pasien dan memudahkan pencatatan administrasi oleh pihak rumah sakit, maka salah satunya dibuat system pakar yang dapat mendeteksi dini COVID-19 ini dengan metode Certainty Factor (CF). sistem pakar ini meniru gejala-gejala yang mirip yang dialami oleh pasien COVID-19 dan akan dikelompokan menjadi beberapa status pasien. Pasien yang mengalami gejala serius akan dikelompokan menjadi Pasien Dalam Pengawasan (PDP) dan pasien yang dianggap memiliki gejala lebih ringan akan dikelompokan kedalam status Orang Dalam Pengawasan (ODP) sedangkan yang mengalami gejala yang diluar dari gejala utama akan dikelompokan menjadi status Non Suspect (NON). Dari 152 data pasien yang diinput pada penelitian ini mendapatkan hasil 114 ODP dengan rata-rata nilai CF 91,38%, 36 PDP deangan rata-rata nilai CF 98,25% dan 2 NON dengan rata-rata nilai CF 40%. CF dengan percobaan perhitungan sistem yaitu data yang

mewakili pasien mendapatkan nilai CF 0.998848 atau 99.88% menjadi PDP. Sistem pakar ini dapat digunakan pengambilan keputusan yang dapat membantu tenaga medis melakukan tindakan dan pengadministrasian dengan lebih baik sebelum dilakukan tes secara menyeluruh di laboratorium untuk memastikan pasien positif atau negatif COVID-19.

Penelitian kedua adalah dari Robby Rizky, Sukisno, Mohammad Ridwan, dan Zaenal Hakim mahasiswa dari Fakultas Tehnologi Dan Informatika Universitas Mathla'ul Anwar Banten, dan Fakultas Tehnik, Universitas Islam Syekh Yusuf pada tahun 2020 dengan judul "Implementasi Metode Forward Chaining Untuk Diagnosa Penyakit COVID-19 Di Rsud Berkah Pandeglang Banten". penelitian ini ditujukan untuk dapat membantu pengecekan pasien yang terjangkit positif Covid-19. Adapun metode yang digunakan adalah metode forward chaining yang dianggap sederhana namun mudah untuk dipahami. Metode ini menggunakan system runut maju yang akan mendeteksi setiap gejala yang dialami pasien. Data yang digunakan yaitu berupa kuesioner hasil wawancara dengan para pakar dibidangnya. Penelitian ini diharapkan dapat memberikan kontribusi terhadap Kabupaten Pandeglang dan memutus rantai penyebaran Covid-19 di daerah tertinggal yang sulit mendapatkan tenaga medis dan peralatan yang memadai. Kesimpulan dari penelitian ini metode dapat memecahkan forward chaining permasalahan tersebut. Berdasarkan hasil penelitian ini metode forward chaining sangat cocok untuk penelitian diagnosis Virus COVID 19 dikarnakan algoritma forward chaining menggunakan algoritma runut maju yang setiap langkah mendiagnosis gejala dan langkah selanjutnya yaitu membuat system aplikasi agar penelitian ini dapat berkembang menjadi system pakar deteksi Virus COVID 19 dengan metode forward chaining kedepanya.

Penelitian terakhir yang digunakan sebagai study literature adalah dari Jimmy Pujoseno mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia dengan judul "Implementasi Deep Learning Menggunakan Convolutional Neural Network Untuk Klasifikasi Alat Tulis". Pada penelitian ini dilakukan pengenalan objek ballpoint, penghapus, dan penggaris menggunakan software Rstudio dan package keras dengan back-end Tensorflow. Data sample yang digunakan sebanyak 300 gambar merupakan hasil crawling dari google image. Untuk melakukan klasifikasi perlu dilakukan training data yang akan membentuk sebuah model. Model tersebut digunakan untuk klasifikasi gambar train dan test untuk 3 kategori. Akurasi yang dihasilkan model untuk data train adalah 100% dan untuk data test sebesar 95%. Kemudian dilakukan percobaan untuk 3 data baru, didapatkan 1 gambar masuk kategori yang salah.

Tabel 2.1. Kajian Teori

No	Penulis dan Tahun	Topik Penelitian	Metode yang digunakan	Hasil
1.	Muhamad Fajar	IMPLEMENTASI	Certainty Factor	Dalam penelitiannya menghasilkan
	Suryana , Fauziah,	SISTEM PAKAR		aplikasi Untuk membantu para tenaga
	Ratih Titi Komala Sari	MENGGUNAKAN		medis dalam mendeteksi dini gejala-
	(2020)	METODE CERTAINTY		gejala yang dialami oleh pasien dan
		FACTOR UNTUK		memudahkan pencatatan administrasi
		MENDIAGNOSA DINI		oleh pihak rumah sakit, maka salah
		CORONA VIRUS		satunya dibuat system pakar yang
		DESEASE (COVID-19)		dapat mendeteksi dini COVID-19 ini
				dengan metode Certainty Factor (CF).
				sistem pakar ini meniru gejala-gejala
				yang mirip yang dialami oleh pasien
				COVID-19 dan akan dikelompokan
				menjadi beberapa status pasien. Pasien
				yang mengalami gejala serius akan
				dikelompokan menjadi Pasien Dalam
				Pengawasan (PDP) dan pasien yang
				dianggap memiliki gejala lebih ringan

akan dikelompokan kedalam status Orang Dalam Pengawasan (ODP) sedangkan yang mengalami gejala yang diluar dari gejala utama akan dikelompokan menjadi status Non Suspect (NON). Dari 152 data pasien yang diinput pada penelitian ini mendapatkan hasil 114 ODP dengan rata-rata nilai CF 91,38%, 36 PDP deangan rata-rata nilai CF 98,25% dan 2 NON dengan rata-rata nilai CF 40%. CF dengan percobaan perhitungan sistem yaitu data yang mewakili pasien mendapatkan nilai CF 0.998848 atau 99.88% menjadi PDP. Sistem ini pakar dapat digunakan pengambilan keputusan yang dapat membantu tenaga medis melakukan tindakan dan pengadministrasian

				dengan lebih baik sebelum dilakukan tes secara menyeluruh di laboratorium untuk memastikan pasien positif atau negatif COVID-19.
2.	Robby Rizky,	IMPLEMENTASI	Forward chaining	Adapun metode yang digunakan
	Sukisno, Mohammad	METODE FORWARD		adalah metode forward chaining yang
	Ridwan, dan Zaenal	CHAINING UNTUK		dianggap sederhana namun mudah
	(2020)	DIAGNOSA PENYAKIT		untuk dipahami. Metode ini
		COVID-19 DI RSUD		menggunakan system runut maju yang
		BERKAH		akan mendeteksi setiap gejala yang
		PANDEGLANG		dialami pasien. Data yang digunakan
		BANTEN		yaitu berupa kuesioner hasil
				wawancara dengan para pakar
				dibidangnya. Penelitian ini diharapkan

dapat memberikan kontribusi terhadap Kabupaten Pandeglang dan memutus rantai penyebaran Covid-19 di daerah tertinggal yang sulit mendapatkan tenaga medis dan peralatan yang memadai. Kesimpulan dari penelitian ini metode forward chaining dapat memecahkan permasalahan tersebut. Berdasarkan hasil penelitian ini metode forward chaining sangat cocok untuk penelitian diagnosis Virus COVID 19 dikarnakan algoritma forward chaining menggunakan algoritma runut maju yang setiap langkah mendiagnosis gejala dan langkah selanjutnya yaitu membuat system aplikasi agar penelitian ini dapat berkembang menjadi system pakar deteksi Virus COVID 19 dengan

				metode forward chaining kedepanya.
3.	Jimmy Pujoseno	IMPLEMENTASI DEEP	Convolution Neural Network	Dengan melakukan klasifikasi objeck
	(2018)	LEARNING		diharapkan mampu mempermudah
		MENGGUNAKAN		toko dalam menjual alat tulis. Pada
		CONVOLUTIONAL		penelitian ini dilakukan pengenalan
		NEURAL NETWORK		objek ballpoint, penghapus, dan
		UNTUK KLASIFIKASI		penggaris menggunakan software
		ALAT TULIS		Rstudio dan package keras dengan
				back-end Tensorflow. Data sample
				yang digunakan sebanyak 300 gambar
				merupakan hasil crawling dari google
				image. Untuk melakukan klasifikasi
				perlu dilakukan training data yang
				akan membentuk sebuah model.
				Model tersebut digunakan untuk

				klasifikasi gambar train dan test untuk 3 kategori. Akurasi yang dihasilkan model untuk data train adalah 100% dan untuk data test sebesar 95%.
				Kemudian dilakukan percobaan untuk
				3 data baru , didapatkan 1 gambar
				masuk kategori yang salah.
		ITIAN SEKARANG		
4.	Pebri Andhi Herry	DETEKSI COVID-19	Convolution Neural Network	Dengan adanya hasil penelitian ini
	Pratama (2020)	BERDASARKAN HASIL		dapat menjadi alat bantu seorang
		RONTGEN		radiologis untuk mendapatkan
		DADA(CHEST X-RAY)		diagnosis yang akurat. Metode yang
		MENGGUNAKAN		di usulkan ini hanya memerlukan
		PYTHON		gambar hasil X-Ray Dada untuk
				digunakan sebagai dataset pelatihan
				dan pengujian.

2.2 TEORI PENDUKUNG

2.2.1 Computer Vision

Computer Vision merupakan salah satu cabang ilmu pengetahuan yang bertujuan untuk membuat suatu keputusan yang berguna untuk mengenali objek fisik nyata dan keadaan berdasarkan sebuah gambar atau citra (Shapiro & Stockman, 2001). Computer Vision menjadikan komputer "acts like human sight", sehingga mendekati kemampuan manusia dalam menangkap informasi visual. Kemampuan itu diantaranya adalah:

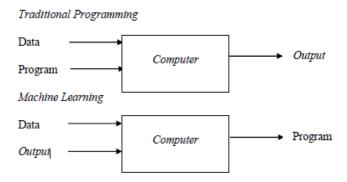
- *Object Detection*: Mengenali sebuah objek ada pada *scene* dan mengetahui dimana batasanya.
- Recognition: Menempatkan label pada objek.
- Description: Menugaskan properti kepada objek.
- 3D Inference: Menafsirkan adegan 3D dari 2D yang dilihat.
- Interpreting motion: Menafsirkan gerakan.

Bidang yang berkaitan erat dengan computer vision adalah image processing (pengolahan citra) dan machine vision (visi mesin). Ada tumpang tindih yang signifikan dalam berbagai teknik dan aplikasi yang mencakup tiga bidang tersebut. Hal ini menunjukkan teknik dasar yang digunakan dan dikembangkan kurang lebih sama (identik). Computer vision mencakup teknologi utama untuk mengalisis citra (visual) secara otomatis yang digunakan dalam bidang lain. Sedangkan machine vision biasanya mengacu pada proses menggabungkan analisis citra otomatis dengan metode dan/atau teknologi lain baik berupa software maupun hardware untuk mencapai tujuan tertentu. Secara luas computer vision berhubungan dan dapat diterapkan/dikombinasikan dengan bidang lain seperti artificial intelligence (kecerdasan buatan), robotika, otomasi industri, pengolahan sinyal, optic fisik, neurobiology, dan bidang lainnya.

2.2.2 Machine Learning

Machine Learning merupakan rangkaian teknik yang mampu membantu untuk menangani dan memprediksi data yang sangat besar dengan cara

mempresentasikan data besar tersebut menggunakan algoritma pembelajaran. *Machine Learning* dapat membuat komputer memprogram diri mereka sendiri. Pada dasarnya, *Machine Learning* membiarkan data untuk melakukan pekerjaannya sendiri. Berikut merupakan gamnbaran umum untuk *Machine Learning* dibandingkan dengan pemrograman yang lainnya (tradisional) tanpa metode ini (Danukusumo, 2017).



Gambar 2. 1. Perbandingan Pemrograman Tradisional dengan Machine Learning

Dari **gambar 2.1** dapat diketahui bahwa pemrograman dengan cara tradisional data dan program komputer untuk menghasilkan suatu *output*. Sedangkan, dengan menggunakan teknik *Machine Learning* data dan *output* dijalankan dengan komputer untuk membuat sebuah program.

Banyak algoritma menggunakan *Machine Learning* yang dikembangkan pada setiap tahunnya. Setiap algoritma *Machine Learning* memiliki tiga komponen yang penting, yaitu :

a. Representasi

Bagaimana mempresentasikan pengetahuan. Contohnya: *Decision tree, Neural Network, Support Vector Machine* dan lain-lain.

b. Evaluasi

Cara mengevaluasi prediksi dan hipotesis. Contohnya: *Mean Squared Error* (MSE), *Cost Function* dan lain-lain.

c. Optimasi

Cara program dari model dihasilkan dan proses pencarian parameter terbaik. Misalnya *Convex Optimization* dan *Gradient Descent*.

Machine learning sendiri dikemukakan oleh Arthur Samuel pada tahun 1959 (Provost, 1998). Machine learning sendiri terdiri dari 3 (tiga) tipe yaitu:

• Supervised learning

Supervised learning biasanya digunakan untuk predictive modeling. Predictive modeling adalah model dari machine learning yang dapat memprediksi nilai dari nilai lain yang ada di input data. Supervised learning mempunyai dua tipe yaitu:

- a. Classification
- b. Regression

• *Unsupervised learning*

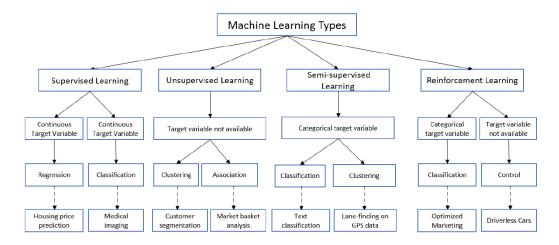
Unsupervised learning biasanya digunakan untuk pattern detection dan descriptive modeling. Algoritma ini mengubah input data untuk mendeteksi pola atau mengmumpulkan data. Unsupervised learning mempunyai dua tipe yaitu:

- a. Clustering
- b. Association rule learning
- Semi supervised learning

Semi supervised learning biasanya digunakan untuk descriptive modeling, karena pada dasarnya descriptive modeling itu mahal maka semi-supervised learning menjadi salah satu cara untuk menangani hal tersebut.

• Reinforcement learning

Reinforcement learning menggunakan interaksi pada lingkungan untuk memaksimalkan hasil atau meminimalisir kegagalan. Reinforcement learning mempelajari pola-pola dari data dengan cara pengulangan, maka dari proses tersebut reinforcement learning akan mendapatkan suatu intelligent yang diharapkan dapat menyelesaikan masalah lain.



Gambar 2. 2. Tipe-tipe machine learning

Terdapat 5 langkah dasar yang digunakan untuk melakukan tugas *Machine Learning* dan tugas ini sangat penting dalam mempersiapkan solusi untuk segala bentuk permasalahan dalam *Machine Learning* maupun *Deep Learning*:

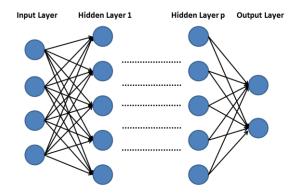
- 1. Mengumpulkan data : terbentuk dari simpanan pada beberapa file yang berisikan data yang dapat diperlajari komputer, dan setelah itu dipisah antara fitur masukan (*input*) dan keluaran (*output*).
- 2. Mempersiapkan data : penentuan kualitas data pada *preprocessing data* sehingga hasil yang didapat juga optimal/baik.
- 3. Melatih sebuah model : langkah ini dilakukan dengan pemilihan algoritma dan representasi data yang tepat dalam bentuk model.
- 4. Mengevaluasi model : menguji keakuratan hasil berdasarkan bagian *test* dataset.
- 5. Meningkatkan kinerja : langkah ini melibatkan pemilihan *hyperparameter* untuk meningkatkan efisiensi dan biasanya menggunakan *cross-validation*.

2.2.3 Deep Learning

Deep Learning adalah cabang ilmu dari Machine Learning berbasis jaringan saraf tiruan (JST) atau bisa dikatakan perkembangan dari JST yang mengajarkan komputer untuk melakukan suatu tindakan yang dianggap alami oleh manusia, misalnya belajar dari contoh. Dalam deep learning, sebuah komputer belajar

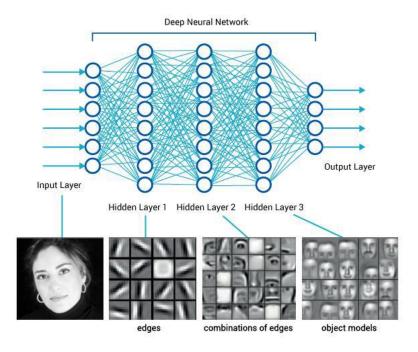
mengklasifikasi secara langsung dari gambar, teks, atau suara. Sebagaimana sebuah komputer dilatih menggunakan data set berlabel dalam jumlah besar lalu kemudian mengubah nilai piksel dari sebuah gambar menjadi representasi internal atau *feature vector* dimana pengklasifikasi dapat mendeteksi atau mengklasifikasi pola pada *input* (LeCun et al, 2015).

Metode deep learning merupakan metode pembelajaran representasi dengan beberapa tingkat representasi, dimana representasi membentuk medan arsitektur jaringan syaraf yang berisi banyak *layer* (lapisan). Lapisan pada *deep learning* terdiri atas tiga bagian, yaitu *input layer*, *hidden layer*, dan *output layer*. Pada *hidden layer* dapat dibuat berlapis-lapis untuk menemukan komposisi algoritma yang tepat agar meminimalisir *error* pada *output*.



Gambar 2. 3. Layer-layer pada *Deep Learning*

Gambar 2.3 mengilustrasikan *layer-layer* pada *deep learning* yang memiliki p + 2 *layer* (p hidden *layer*, 1 *input* dan 1 *output layer*). Bulatan berwarna biru menggambarkan *neuron*. Disetiap lapis *hidden layer* terdapat satu atau lebih neuron. Neuron-neuron tersebut akan terhubung langsung dengan neuron lain pada *layer* selanjutnya. Koneksi antar neuron hanya terjadi di antara 2 buah *layer* (*input* dan *output*) tidak ada koneksi pada *layer* yang sama walaupun secara teknis bisa saja dibuat dan juga *fully connected*.



Gambar 2. 4. Visual object recognition pada deep learning

Deep learning memungkinkan model komputasi yang terdiri dari beberapa processing layer untuk mempelajari representasi data dengan berbagai tingkat abstraksi. Metode ini telah secara dramatis memperbaiki state-of-the-art dalam pengenalan suara (speech recognition), pengenalan objek visual (visual object recognition), deteksi objek (object detection) dan banyak domain lainnya seperti penemuan obat dan genomik. Deep learning menemukan struktur yang rumit besar kumpulan dengan dalam data yang menggunakan algoritma backpropagation untuk menunjukkan bagaimana sebuah mesin harus mengubah parameter internalnya yang digunakan untuk menghitung representasi pada setiap lapisan dari representasi pada lapisan sebelumnya (LeCun, Bengio, & Hinton, 2015).

Beberapa algoritma yang menerapkan konsep *deep learning* antara lain *Convolutional Neural Networks* (CNN) untuk klasifikasi gambar, *Deep Belief Network – Deep Neural Network* (DBN-DNN) untuk pengenalan suara, *Recurrent Neural Network* (RNN) untuk penerjemahan bahasa, *Query-Oriented Deep Extraction* (QODE) yang berbasis *Restricted Boltzmann Machine* (RBM) untuk peringkasan multi dokumen, *Conditional Restricted Boltzmann Machine* (RBM)

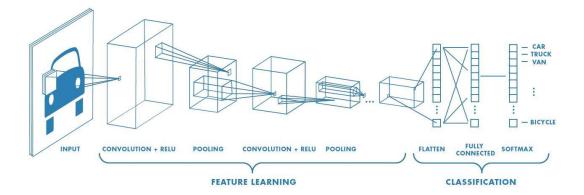
untuk memprediksi *Drug-Target Interaction* (DTI), dan *Deep Belief Network* (DBN) untuk prediksi data sesuai waktu.

2.2.4 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN/ConvNet) adalah salah satu algoritma dari deep learning yang merupakan pengembangan dari Multilayer Percepton (MLP) yang dirancang untuk mengolah data dalam bentuk dua dimensi, misalnya gambar atau suara. CNN digunakan untuk mengklasifikasi data yang terlabel dengan menggunakan metode supervised learning, yang mana cara kerja dari supervised learning adalah terdapat data yang dilatih dan terdapat variabel yang ditargetkan sehingga tujuan dari metode ini adalah mengkelompokan suatu data ke data yang sudah ada.

CNN sering digunakan untuk mengenali benda atau pemandangan, dan melakukan deteksi dan segmentasi objek. CNN belajar langsung dari data citra, sehingga menghilangkan ekstraksi ciri secara manual. Penelitian awal yang mendasari penemuan ini pertama kali dilakukan oleh Hubel dan Wiesel yang melakukan penelitian *visual cortex* pada indera penglihatan kucing. *Visual cortex* pada hewan sangat *powerful* dalam sistem pemrosesan visual yang pernah ada. Hingga banyak penelitian yang terinspirasi dari cara kerjanya dan menghasilkan model-model baru diantaranya seperti Neocognitron, HMAX, LeNet-5, dan AlexNet.

Lapisan-lapisan CNN memiliki susunan neuron 3 dimensi (lebar, tinggi, kedalaman). Lebar dan tinggi merupakan ukuran lapisan sedangkan kedalaman mengacu pada jumlah lapisan. Sebuah CNN dapat memiliki puluhan hingga ratusan lapisan yang masing-masing belajar mendeteksi berbagai gambar. Pengolahan citra diterapkan pada setiap citra latih pada resolusi yang berbeda, dan *output* dari masing-masing gambar yang diolah dan digunakan sebagai *input* ke lapisan berikutnya (gambar 2.6). Pengolahan citra dapat dimulai sebagai fitur yang sangat sederhana, seperti kecerahan dan tepi atau meningkatkan kompleksitas pada fitur yang secara unik menentukan objek sesuai ketebalan lapisan (Mathworks, 2017).



Gambar 2. 5. Contoh layer CNN

Secara umum tipe lapisan pada CNN dibagi menjadi dua yaitu:

- a) Layer ekstraksi fitur (*feature extraction layer*) gambar, letaknya berada pada awal arsitektur tersusun atas beberapa lapisan dan setiap lapisan tersusun atas *neuron* yang terkoneksi pada daerah lokal (*local region*) dari lapisan sebelumnya. Lapisan jenis pertama adalah *convolutional layer* dan lapisan kedua adalah *pooling layer*. Setiap lapisan diberlakukan fungsi aktivasi dengan posisinya yang berselang-seling antara jenis pertama dengan jenis kedua. Lapisan ini menerima *input* gambar secara langsung dan memprosesnya hingga menghasilkan *output* berupa vektor untuk diolah pada lapisan berikutnya.
- b) Layer klasifikasi (*classfication layer*), tersusun atas beberapa lapisan dan setiap lapisan tersusun atas *neuron* yang terkoneksi secara penuh (*fully connected*) dengan lapisan lainnya. Layer ini menerima *input* dari hasil keluaran layer ekstrasi fitur gambar berupa vektor kemudian ditransformasikan seperti *Multi Neural Networks* dengan tambahan beberapa *hidden layer*. Hasil keluaran berupa akurasi kelas untuk klasifikasi.

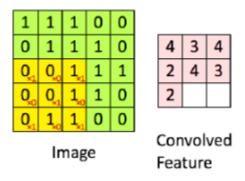
Dengan demikian CNN merupakan metode untuk mentransformasikan gambar asli lapisan per lapisan dari nilai piksel gambar kedalam nilai skoring

kelas untuk klasifikasi. Dan setiap lapisan ada yang memiliki *hyperparameter* dan ada yang tidak memiliki parameter (bobot dan bias pada neuron).

Pada penelitian ini terdapat empat macam lapisan utama antara lain:

1) Convolution Layer

Convolution Layer melakukan operasi konvolusi pada output dari lapisan sebelumnya. Layer tersebut adalah proses utama yang mendasari sebuah CNN. Konvolusi adalah suatu istilah matematis yang berati mengaplikasikan sebuah fungsi pada output fungsi lain secara berulang. Dalam pengolahan citra, konvolusi berati mengaplikasikan sebuah kernel (kotak kuning) pada citra disemua offset yang memungkinkan seperti yang ditunjukkan pada Gambar 2.7.



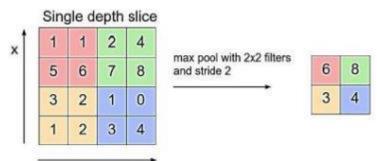
Gambar 2. 6. Operasi pada konvolusi

Kotak hijau secara keseluruhan adalah citra yang akan dikonvolusi. *Kernel* bergerak dari sudut kiri atas ke kanan bawah. Sehingga hasil konvolusi dari citra tersebut dapat dilihat pada gambar disebelah kanannya. Tujuan dilakukannya konvolusi pada data citra adalah untuk mengekstraksi fitur dari citra *input*. Konvolusi akan menghasilkan transformasi *linear* dari data *input* sesuai informasi spasial pada data. Bobot pada lapisan tersebut menspesifikasikan *kernel* konvolusi yang digunakan, sehingga kernel konvolusi dapat dilatih berdasarkan *input* pada CNN.

2) Subsampling Layer

Subsampling adalah proses mereduksi ukuran sebuah data citra. Dalam pengolahan citra, subsampling juga bertujuan untuk meningkatkan invariansi posisi dari fitur. Dalam sebagian besar CNN, metode

subsampling yang digunakan adalah max pooling. Max pooling membagi output dari convolution layer menjadi beberapa grid kecil lalu mengambil nilai maksimal dari setiap grid untuk menyusun matriks citra yang telah direduksi seperti yang ditunjukkan pada Gambar 2.8.



Gambar 2. 7. Operasi pada max pooling

Grid yang berwarna merah, hijau, kuning dan biru merupakan kelompok grid yang akan dipilih nilai maksimumnya. Sehingga hasil dari proses tersebut dapat dilihat pada kumpulan grid disebelah kanannya. Proses tersebut memastikan fitur yang didapatkan akan sama meskipun objek citra mengalami translasi (pergeseran).

Penggunaan *pooling layer* pada CNN hanya bertujuan untuk mereduksi ukuran citra sehingga dapat dengan mudah digantikan dengan sebuah *convolution layer* dengan *stride* yang sama dengan *pooling layer* yang bersangkutan (Springenberg et al, 2014).

3) Fully Connected Layer

Layer tersebut adalah lapisan yang biasanya digunakan dalam penerapan MLP dan bertujuan untuk melakukan transformasi pada dimensi data agar data dapat diklasifikasikan secara linear. Setiap *neuron* pada *convolution layer* perlu ditransformasi menjadi data satu dimensi terlebih dahulu sebelum dapat dimasukkan ke dalam sebuah *fully connected layer*. Karena hal tersebut menyebabkan data kehilangan informasi spasialnya dan tidak reversibel, *fully connected layer* hanya dapat diimplementasikan di akhir jaringan.

4) Fungsi Aktivasi

Pada penelitian ini fungsi aktivasi yang digunakan adalah softmax layer. Softmax layer digunakan apabila pada permasalahan multiclass classification, output layer biasanya memiliki lebih dari satu neuron. Softmax layer pada dasarnya adalah probabilitas eksponensial yang dinormalisasi dari pengamatan kelas yang diwakili sebagai aktivasi neuron. Fungsi eksponensial akan meningkatkan probabilitas nilai maksimum lapisan sebelumnya dibandingkan dengan nilai lainnya.

2.2.5 Citra *X-Ray*

Wilhelm Conrad Rontgen adalah seorang ahli fisika dari Universitas Wurzburg, Jerman, yang pertama kali menemukan sinar rontgen pada tahun 1895 sewaktu melakukan percobaan dengan sinar katoda. Pada saat itu beliau melihat timbulnya sinar flourensensi yang berasal dari Kristal barium platiosianida dalam tabung crookes hittorf yang dialiri listrik. Beliau segera menyadari bahwa fenomena ini merupakan suatu penemuan baru sehingga dengan gigih beliau terus menerus melanjutkan penyelidikannya dalam minggu-minggu berikutnya. Tidak lama kemudian ditemukanlah sinar yang disebutnya sinar sinar baru atau x-ray, dikemudian hari masyarakat menamakan sinar tersebut sebagai sinar *rontgen* sebagai penghargaan kepada Wilhelm Conrad Rontgen. (Gabriel, 1996 & Ainatul, 2011).

X-ray merupakan suatu bentuk radiasi seperti cahaya atau gelombang bunyi. *X-ray* dapat melewati banyak objek termasuk tubuh. Mesin *x-ray* menghasilkan pancaran radiasi kecil yang melewati tubuh, merekam citra pada film atau plat khusus untuk merekam gambar digital. Masing-masing bagian tubuh menyerap *x-ray* dengan dosis bervariasi. Tulang yang padat menerima radiasi yang lebih besar dibanding bagian yang lebih lunak seperti otot, lemak ataupun organ tubuh. Sebagian hasilnya, tulang tergambar putih pada *x-ray*, bagian yang lunak tergambar abu-abu dan udara hitam. Pada *x-ray* dada, tulang rusuk dan tulang belakang akan menyerap banyak radiasi dan tampak abu-abu putih atau terang pada gambar. Jaringan paru-paru menyerap radiasi kecil dan akan terlihat gelap

pada gambar. Pada sistem pencitraan *x-ray* diperlukan tegangan tinggi, dengan tujuan agar dapat dihasilkan berkas *x-ray*. Untuk itu rangkaian listriknya dirancang sedemikian rupa sehingga tegangan tinggi (*kV*) dengan rentang yang besar. Jika *kV*-nya rendah maka sinar-x memiliki gelombang yang panjang sehingga akan mudah diserap oleh atom dari target (anoda), kemudian disebut sebagai soft *x-ray*. Radiasi yang dihasilkan dengan pengaturan tegangan yang cukup tinggi maka akan dihasilkan x- ray dengan daya tembus yang besar dan panjang gelombang yang pendek (Ferry Suyanto, 2008).

Manfaat x-ray dalam ilmu kedoteran, yaitu x-ray dapat digunakan untuk melihat kondisi tulang, gigi, paru-paru serta organ tubuh yang lain tanpa melakukan pembedahan langsung pada tubuh pasien. Selain bermanfaat, x-ray mempunyai efek atau dampak yang sangat berbahaya bagi tubuh kita yaitu apabila digunakan berlebihan akan menimbulkan penyakit yang berbahaya, seperti kanker dan penyakit lainnya. Oleh sebab itu, para dokter tidak menganjurkan terlalu sering memakai "foto *rontgen*" secara berlebihan (Gabriel, 1996).



Gambar 2. 8. Gambar citra x-ray thorax

2.2.6 *Thorax*

Thorax (rongga dada) adalah daerah tubuh yang terletak diantara leher dan abdomen. Thorax rata dibagian depan dan belakang tetapi melengkung di bagian samping. Rangka dinding thorax yang dinamakan cavea thoracis dibentuk oleh bagian columna vertebralis dibelakang, costae dan spatium intercostale disamping, serta sternum dan cartilage costalis didepan. Dibagian atas, thorax

berhubungan dengan leher dan dibagian bawah dipisahkan dari *abdomen* oleh *diaphragma*. *Cavea thoracis* melindungi paru dan jantung dan merupakan tempat perlekatan otot-otot *thorax*, *extremitas superior*, *abdomen* dan punggung. *Cavitas thoracis* (rongga *thorax*) dapat dibagi menjadi: bagian tengah yang disebut *mediastinum* dan bagian lateral yang ditempati *pleura* dan paru. Paru diliputi oleh selapis *membrane* tipis yang disebut *pleura viscelaris*, yang beralih di hilus *pulmonalis* (tempat saluran udara utama dan pembuluh darah masuk ke paru-paru) menjadi *pleura parietalis* dan menuju ke permukaan dalam dinding *thorax*. Dengan cara ini terbentuk dua kantong *membranosa* yang dinamakan civitas pleuralis pada setiap sisi *thorax*, diantara paru-paru dan dinding *thorax* (Richard S. Snell, 2006).

BAB III

METODOLOGI PENELITIAN

3.1 Populasi dan Sampel

Populasi dalam penelitian ini adalah citra hasil rontgen dada. Sedangkan sampel yang digunakan dalam penelitian ini adalah dataset citra rontgen dada yaitu normal, pneumonia, dan COVID-19. Jumlah citra yang digunakan untuk pengujian dilakukan secara bertahap dimulai dari 5, 10, dan seterusnya menysuaikan batasan computer. Hasil dari tiap jumlah yang telah di uji coba akan di analisis bagaiamana tingkat akurasinya menggunakan tiga model CNN.

3.2 Variabel dan Definisi Operasional

Variabel yang digunakan dalam penelitian ini ditampilkan dalam Tabel 3.1 tentang penjelasan dan definisi operasional penelitian:

Tabel 3. 1. Definisi Operasional Variabel

No	Variabel	Definisi
1	Normal	Dataset citra rontgen dada paru-paru normal
		yang tidak terkena infeksi pneumonia ataupun
		COVID-19.
2	Pneumonia	Dataset citra rontgen dada infeksi paru-paru.
3	COVID-19	Dataset citra rontgen dada terinfeksi positif
		COVID-19.

3.3 Jenis dan Sumber Data

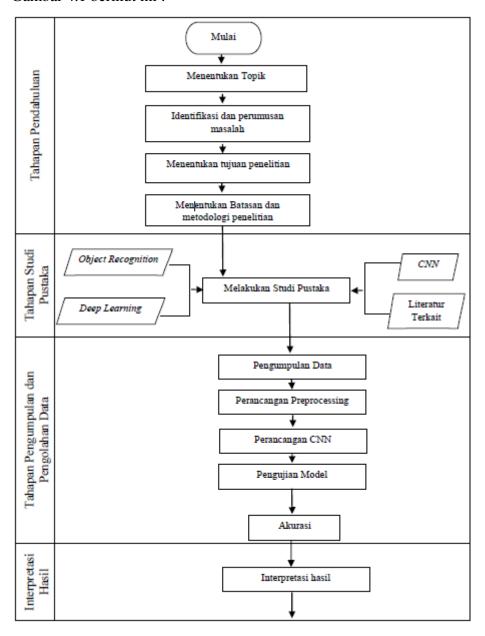
Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut diperoleh dari website kaggle.com.

3.4 Metode Analisis Data

Bahasa Pemrograman yang digunakan pada penelitian ini adalah Python. Sedangkan Metode analisis data yang digunakan dalam penelitian ini adalah metode *Convolutional Neural Network* (CNN) yang bertujuan untuk mengklasifikasikan citra rontgen dada yaitu normal, pneumonia, dan positif COVID-19.

3.5 Tahapan penelitian

Langkah atau tahapan yang dilakukan pada penelitian ini digambarkan melalui Gambar 4.1 berikut ini :

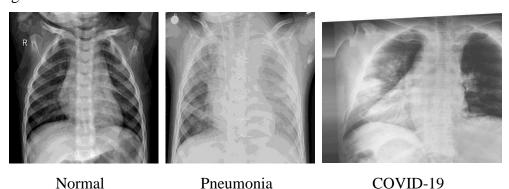


Gambar 3. 1. Tahapan Penelitian

3.6 Rancangan Dataset

Penggunaan dataset pada metode CNN yaitu berupa data gambar. Model CNN akan berjalan dengan baik ketika menggunakan data train gambar yang banyak. Sehingga sebuah model dapat belajar mengenali gambar tersebut. Dataset yang digunakan dalam penelitian ini berupa gambar yang dikumpulkan melalui website *kaggle.com* yang menyediakan banyak dataset hasil penelitian. Data gambar yang digunakan adalah gambar hasil rontgen dada normal, pneumonia, dan COVID-19.

Kaggle adalah situs dan platform untuk berlomba membuat model terbaik untuk menganalisa dan memprediksi suatu dataset. Dataset ini diberikan oleh suatu perusahaan, dengan suatu deskripsi masalah tertentu. Sumber data ini didapatkan dari COVID-19 RADIOGRAPHY DATABASE (Winner of the COVID-19 Dataset Award) yang dipublikasi oleh Tawsifur Rahman bersama para kolaboratornya yang dipublikasi di Kaggle.com. Dataset dibagi berdasarkan 4 *subclass* yang dapat dilihat contohnya pada Gambar 3.2 yaitu gambar hasil rontgen dada normal, positif COVID-19, dan pneumonia gambar.



Gambar 3. 2. Contoh Gambar Subclass Dataset Kaggle

3.7 Perangkat Pengembangan dan Pengujian

- A. Perangkat Keras (*Hardware*) yang digunakan adalah Laptop Asus A407U dengan speksifikasi sebagai berikut:
 - 1) Intel® Core i3 6006U CPU @2000 GHz

- 2) CPU Intel Core i3 6006U
- 3) Intel® HD Graphics 520
- 4) Memory/RAM 10GB
- 5) Harddisk 1 TB
- B. Perangkat Lunak (Software) yang digunakan adalah sebagai berikut:
 - 1) Microsoft Windows 10
 - 2) Edge
 - 3) Visual Studio Code
 - 4) Jupyter Notebook
 - 5) PyCharm

DAFTAR PUSTAKA

- Tim Dosen, 2020, "Panduan Skripsi Jurusan Teknik Informatika 2020", Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangkaraya.
- Today.line.me, "Apa itu virus Corona dan COVID-19?", LINE TODAY, 13 Maret 2020, https://today.line.me/id/article/Apa+itu+Corona+Virus+dan+Covid+19-pwzorq [Diakses Pada 10 Agustus 2020]
- id.techinasia.com, "Apa itu Machine Learning dan perbedaannya dengan Deep Learning", TECHINASIA, 6 Juli 2018, https://id.techinasia.com/apa-itu-machine-learning> [Diakses Pada 10 Agustus 2020]
- Pujoseno, J. (2018). Implementasi *Deep Learning* menggunakan *Convolutional Neural Network* untuk klasifikasi Alat Tulis. Jurusan Statistika, Fakultas Matematika Dan Ilmu Pengetahuan Alam , Universitas Islam Indonesia Yogyakarta.
- Gunawan, C. I., (2020). ANOMALI COVID-19: DAMPAK POSITIF VIRUS CORONA UNTUK DUNIA. CV IRDH.
- LiceFrense, e D. (2013). Hak Cipta Dilindungi. *Python "Belajar Pemrograman Python Dasar*." http://www.ask-jansen.com/wp-content/uploads/2014/04/Kontroversi-Kalori-ebook.pdf
- Rizky, R., Ridwan, M., & Hakim, Z. (2020). Implementasi Metode Forward Chaining Untuk Diagnosa Penyakit Covid 19 Di Rsud Berkah Pandeglang Banten. *Jurnal Teknologi Informasi*, 4(1), 1–4.
- Suryana, M. F., Fauziah, F., & Sari, R. T. K. (2020). Implementasi Sistem Pakar Menggunakan Metode Certainty Factor Untuk Mendiagnosa Dini Corona Virus Desease (COVID-19). *Jurnal Media Informatika Budidarma*, *4*(3), 559. https://doi.org/10.30865/mib.v4i3.2132