

PENERAPAN *VISUAL QUESTION ANSWERING* DALAM MENAFSIRKAN CITRA MEDIS MENGUNAKAN *DEEP LEARNING*

TUGAS AKHIR

Diajukan untuk melengkapi tugas-tugas dan
memenuhi syarat-syarat guna memperoleh gelar Sarjana Komputer

Oleh:

ABDUL HAFIDH
2008107010056



**PROGRAM STUDI INFORMATIKA JURUSAN INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SYIAH KUALA, BANDA ACEH
JANUARI, 2024**

PENGESAHAN

PENERAPAN VISUAL QUESTION ANSWERING DALAM MENAFSIRKAN CITRA MEDIS MENGGUNAKAN DEEP LEARNING

APPLICATION OF VISUAL QUESTION ANSWERING IN INTERPRETING MEDICAL IMAGES USING DEEP LEARNING

Oleh:

Nama : Abdul Hafidh
NPM : 2008107010056
Jurusan : Informatika

Menyetujui:

Pembimbing I

Pembimbing II

Alim Misbullah, S.Si., M.S.
NIP. 198806032019031011

Laina Farsiah, S.Si., M.S.
NIP. 198902032022032004

Mengetahui:

Dekan Fakultas MIPA
Universitas Syiah Kuala,

Koordinator Program Studi Informatika
Universitas Syiah Kuala,

Prof. Dr. Teuku M. Iqbalsyah, S.Si, M.Sc.
NIP. 197110101997031003

Viska Mutiawani, B.IT, M.IT.
NIP. 198008312009122003

Lulus Sidang Sarjana pada hari Kamis, 14 Mei 2024

PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini,

Nama lengkap : Abdul Hafidh
Tempat/tanggal lahir : Banda Aceh/ 29 Maret 2002
NPM : 2008107010056
Program Studi : Informatika
Fakultas : MIPA
Judul Tugas Akhir : PENERAPAN *VISUAL QUESTION ANSWERING* DALAM
MENAFSIRKAN CITRA MEDIS MENGGUNAKAN
DEEP LEARNING

Menyatakan dengan sesungguhnya bahwa Laporan Tugas Akhir saya dengan judul di atas adalah **hasil karya saya sendiri** bersama dosen pembimbing dan **bebas plagiasi**.

Jika ternyata di kemudian hari terbukti bahwa Laporan Tugas Akhir merupakan hasil plagiasi, saya bersedia menerima sanksi yang berlaku di Universitas Syiah Kuala.

Banda Aceh, 14 Januari 2024

Abdul Hafidh
NPM. 2008107010056

SURAT PERNYATAAN

Yang bertanda tangan di bawah ini,

1. Nama : Abdul Hafidh
NPM : 2008107010056
Jurusan/Prodi : Informatika
Status : Mahasiswa
2. Nama : Alim Misbullah, S.Si., M.S.
NIP : 198806032019031011
Jurusan/Prodi : Informatika
Status : Pembimbing I
3. Nama : Laina Farsiah, S.Si., M.S.
NIP : 198902032022032004
Jurusan/Prodi : Informatika
Status : Pembimbing II

Dengan ini menyatakan hasil penelitian Tugas Akhir yang berjudul **“PENERAPAN VISUAL QUESTION ANSWERING DALAM MENAFSIRKAN CITRA MEDIS MENGGUNAKAN DEEP LEARNING”** tidak dipublikasikan secara *full-text* di sistem ETD (*Electronic Theses and Dissertations*) Universitas Syiah Kuala hingga batas waktu 5 tahun dari tanggal kelulusan.

Demikian surat pernyataan ini dibuat dengan sebenarnya untuk dapat dipergunakan seperlunya.

Darussalam, 14 Mei 2022
Yang membuat pernyataan,

Pembimbing I,

Pembimbing II,

Mahasiswa,

Alim Misbullah, S.Si., M.S.
NIP. 198806032019031011

Laina Farsiah, S.Si., M.S.
NIP. 198902032022032004

Abdul Hafidh
NPM. 2008107010056

Mengetahui:

Koordinator Program Studi Informatika
Universitas Syiah Kuala,

Koordinator TA,

Viska Mutiawani, B.IT, M.IT.
NIP. 198008312009122003

Alim Misbullah, S.Si., M.S.
NIP. 198806032019031011

KATA PENGANTAR

Segala puji dan syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya kepada kita semua, sehingga penulis dapat menyelesaikan penulisan Tugas Akhir yang berjudul **“PENERAPAN *VISUAL QUESTION ANSWERING* DALAM MENAFSIRKAN CITRA MEDIS MENGGUNAKAN *DEEP LEARNING*”** yang telah dapat diselesaikan sesuai rencana. Penulis banyak mendapatkan berbagai pengarahan, bimbingan, dan bantuan dari berbagai pihak. Oleh karena itu, melalui tulisan ini penulis mengucapkan rasa terima kasih kepada:

1. Ayah dan Ibu sebagai kedua orang tua penulis yang senantiasa selalu mendukung aktivitas dan kegiatan yang penulis lakukan baik secara moral maupun material serta menjadi motivasi terbesar bagi penulis untuk menyelesaikan Tugas Akhir ini.
2. Ibu Viska Mutiawani, B.IT., M.IT., selaku Ketua Jurusan Informatika Fakultas MIPA Universitas Syiah Kuala.
3. Bapak Alim Misbullah, S.Si., M.S. selaku Dosen Pembimbing I yang telah banyak memberikan bimbingan dan arahan kepada penulis, dan selaku koordinator tugas akhir yang membantu penulis sehingga bisa menyelesaikan Tugas Akhir ini.
4. Ibu Laina Farsiah, S.Si., M.S. selaku Dosen Pembimbing II yang telah banyak memberikan bimbingan dan arahan kepada penulis, sehingga penulis dapat menyelesaikan Tugas Akhir ini.
5. Ibu Dalila Husna Yunardi, B.Sc., M.Sc., selaku Dosen Wali.
6. Seluruh Dosen dan Staff di Jurusan Informatika Fakultas MIPA atas ilmu dan didikannya selama perkuliahan.
7. Sahabat dan teman-teman seperjuangan Jurusan Informatika USK 2020 lainnya.

Penulis juga menyadari segala ketidaksempurnaan yang terdapat didalamnya baik dari segi materi, cara, ataupun bahasa yang disajikan. Seiring dengan ini penulis mengharapkan kritik dan saran dari pembaca yang sifatnya dapat berguna untuk kesempurnaan Tugas Akhir ini. Harapan penulis semoga tulisan ini dapat bermanfaat bagi banyak pihak dan untuk perkembangan ilmu pengetahuan.

Banda Aceh, 14 Januari 2024

Abdul Hafidh
NPM. 2008107010056

ABSTRAK

lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed euismod, nisl quis lacinia ultricies, nunc nisl ultricies diam, quis aliquam nisl nisl quis nisl.

Kata kunci : lorem ipsum dolor sit amet

ABSTRACT

lorem ipsum dolot sit amet, consectetur adipiscing elit. Sed euismod, nisl quis lacinia ultricies, nunc nisl ultricies diam, quis aliquam nisl nisl quis nisl.

Kata kunci : lorem, ipsum, dolor, sit, amet

DAFTAR ISI

Halaman Judul	i
Halaman Pengesahan	ii
Pernyataan Bebas Plagiasi	iii
Surat Pernyataan	iv
Kata Pengantar	v
Abstrak	vi
Abstract	vii
Daftar Isi	viii
Daftar Tabel	x
Daftar Gambar	xi
Daftar Program	xii
Daftar Lampiran.....	xiii
 BAB I PENDAHULUAN	 1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	1
1.3. Tujuan Penelitian.....	2
1.4. Manfaat Penelitian	2
 BAB II TINJAUAN KEPUSTAKAAN	 3
2.1. Pengertian Citra	3
2.2. <i>Deep Learning</i>	4
2.2.1. <i>Artificial Neural Network</i>	5
2.2.2. <i>Fungsi Aktivasi</i>	5
2.2.3. <i>Loss Function</i>	6
2.2.4. <i>Optimization Function</i>	7
2.3. <i>Convolutional Neural Network</i>	7
2.4. <i>Long Short Term Memory</i>	9
2.5. <i>Transfer Learning</i>	9
2.6. <i>Performance Metrics</i>	12
2.6.1. <i>Accuracy Score</i>	12
2.6.2. <i>BLEU Score</i>	13
2.7. <i>Visual Question Answering</i>	13
2.8. <i>Medical Visual Question Answering Dataset</i>	14
2.8.1. <i>PathVQA Dataset</i>	15
2.8.2. <i>VQA-RAD Dataset</i>	15
2.9. Penelitian Terkait.....	15
 BAB III METODOLOGI PENELITIAN	 17
3.1. Waktu dan Lokasi Penelitian	17
3.2. Jadwal Pelaksanaan.....	17
3.3. Alat dan Bahan	18
3.3.1. Perangkat Keras.....	19
3.3.2. Perangkat Lunak.....	19

3.4. Metode Penelitian.....	19
3.4.1. Studi Literatur	20
3.4.2. Pengumpulan Data	20
3.4.3. <i>Exploratory Data Analysis</i>	21
3.4.4. Pemrosesan Data	21
3.4.5. Membangun Model	21
3.4.6. Melatih Model.....	21
3.4.7. Perbandingan Hasil	21
3.4.8. Membangun Sistem Medical Cerdas Berbasis Web.....	21
BAB IV HASIL DAN PEMBAHASAN	22
4.1. Pengumpulan Data.....	22
4.2. Pengembangan Model	22
4.3. Perbandingan Model	22
4.4. Pengembangan Sistem Medical Cerdas Berbasis Web.....	22
BAB V KESIMPULAN DAN SARAN	23
5.1. Kesimpulan	23
5.2. Saran	23
DAFTAR PUSTAKA	24
LAMPIRAN.....	27

DAFTAR TABEL

Tabel 3.1	Jadwal pelaksanaan bulan agustus hingga Oktober 2023	17
Tabel 3.2	Jadwal pelaksanaan November 2023 hingga Januari 2024	17
Tabel 3.3	Jadwal pelaksanaan Februari hingga April 2024	18

DAFTAR GAMBAR

Gambar 2.1	Kategori Pendekatan <i>Deep Learning</i> (Alom et al., 2019)	4
Gambar 2.2	Perbandingan teknik optimisasi Adam dengan teknik optimisasi lainnya (Kingma and Ba, 2014)	7
Gambar 2.3	Arsitektur CNN (Monedero et al., 2021)	8
Gambar 2.4	Diagram untuk LSTM (Alom et al., 2019)	9
Gambar 2.5	Arsitektur VGG19 (Nguyen et al., 2022)	11
Gambar 2.6	Arsitektur ViLT (Kim et al., 2021)	12
Gambar 2.7	Vanilla VQA <i>Network</i> Model (Srivastava et al., 2021)	14
Gambar 3.1	Sampel data PathVQA	18
Gambar 3.2	Sampel data VQA-RAD	19
Gambar 3.3	Diagram Alir Penelitian	20

DAFTAR PROGRAM

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Citra medis adalah subjek yang sering menjadi fokus penelitian dan merupakan gambar dua dimensi yang mewakili bagian dalam tubuh manusia. Citra ini digunakan oleh profesional medis untuk mendeteksi serta menganalisis penyakit pada pasien. Ada berbagai teknik yang digunakan untuk menghasilkan citra medis, yaitu *Magnetic Resonance Imaging* (MRI), *X-Ray*, *Ultrasonography* (USG), *Endoscopy*, *Computed Tomography* (CT-Scan), dan *Nuclear Medicine* (Kusuma and Kusumadewi, 2020). Namun, dalam beberapa kasus seorang dokter juga bisa membuat kesalahan dalam menganalisis penyakit pasien, karena pada dasarnya mereka adalah manusia dan rentan terhadap kesalahan (Mauli, 2018).

Oleh sebab itu, dibutuhkan sebuah sistem yang dapat membantu dokter dalam menjawab permasalahan yang terdapat pada citra medis. Salah satu solusi yang dapat digunakan adalah dengan pendekatan *visual question answering* (VQA). Dalam konteks medis VQA dapat memberikan manfaat bagi dokter dan pasien yang mana dokter bisa memperoleh jawaban yang diperoleh dari sistem VQA sebagai bentuk dalam pengambilan keputusan. Sedangkan pasien bisa mengajukan pertanyaan ke sistem VQA dengan gambar medis yang ada pada dirinya untuk mengetahui kondisi kesehatannya (Nguyen et al., 2019).

Sistem VQA medis ini juga dapat berperan sebagai asisten yang memiliki pengetahuan yang luas. Sebagai contoh, "pendapat tambahan" dari sistem VQA dapat membantu dokter dalam menafsirkan citra medis dan mengurangi potensi kesalahan dalam diagnosis pada saat bersamaan (Tschandl et al., 2020).

Pada penelitian ini, akan dibuat sebuah sistem VQA yang dapat menjawab pertanyaan berdasarkan citra medis. Sistem ini akan dibuat dengan menggunakan *deep learning* yang merupakan salah satu cabang dari *machine learning* yang menggunakan jaringan saraf tiruan untuk memodelkan masalah yang kompleks. Dalam penelitian ini, akan digunakan *deep learning* untuk memodelkan masalah VQA dalam citra medis.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang di atas, permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana membangun model *visual question answering* (VQA) untuk citra medis dengan mengintegrasikan metode *deep learning*??

2. Model *deep learning* seperti apa yang paling cocok untuk ekstraksi fitur citra medis dalam konteks *medical visual question answering*?
3. Bagaimana cara mengevaluasi model VQA untuk citra medis?

1.3 TUJUAN PENELITIAN

Adapun maksud dan tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan model VQA berdasarkan metode *deep learning* untuk citra medis.
2. Membangun model *deep learning* yang efektif untuk ekstraksi fitur citra medis dalam konteks *medical visual question answering*.
3. Mengevaluasi model VQA untuk citra medis.

1.4 MANFAAT PENELITIAN

Manfaat yang diinginkan dari penelitian ini adalah untuk memberikan gambaran mengenai sistem VQA untuk citra medis dengan *deep learning*. Lalu, memberikan gambaran mengenai model *deep learning* yang dapat digunakan untuk membangun sistem VQA untuk citra medis. Dan yang terakhir, memberikan gambaran mengenai cara mengevaluasi sistem VQA untuk citra medis.

BAB II

TINJAUAN KEPUSTAKAAN

2.1 PENGERTIAN CITRA

Citra adalah gambaran dari suatu objek yang bisa kita lihat. Citra analog tidak bisa disimpan secara langsung dalam komputer. Oleh karena itu, kita perlu mengubah citra analog menjadi citra digital agar bisa diolah oleh komputer. Citra digital adalah citra yang bisa diolah menggunakan perangkat komputer. Alasan utama citra analog tidak bisa diolah oleh komputer adalah karena citra analog tidak memiliki konsep sampling dan kuantisasi. Konsep sampling adalah suatu metode yang mengubah citra analog menjadi grid berbentuk M baris dan N kolom, sehingga citra menjadi lebih tersegmentasi. Semakin besar nilai M dan N, semakin halus citra digital yang dihasilkan. Sedangkan konsep kuantisasi adalah suatu metode yang mengubah intensitas dari citra analog menjadi intensitas diskrit. Baik sampling maupun kuantisasi memegang peranan penting dalam mengambil potongan citra menjadi bentuk M baris dan N kolom (proses sampling) serta menentukan nilai intensitas yang terdapat pada setiap titik tersebut (proses kuantisasi). Hasil akhir dari kedua konsep ini adalah citra yang memiliki resolusi sesuai dengan yang kita inginkan (Andono et al., 2018).

Dalam ilmu pengolahan citra, beragam jenis citra ada, dan salah satunya adalah citra RGB, juga disebut sebagai citra *true color*. Citra ini memiliki matriks data berukuran $m \times n \times 3$, yang menggambarkan warna merah, hijau, dan biru pada setiap piksel. Rentang nilai untuk setiap warna adalah antara 0 (nilai minimum) hingga 255 (nilai maksimum) dalam monitor komputer. Skala 256 ini dipilih berdasarkan representasi delapan digit dalam bilangan biner yang digunakan oleh komputer. Dengan demikian, lebih dari 16 juta warna dapat dihasilkan secara total. Penentuan warna pada setiap piksel didasarkan pada tingkat kecerahan merah, hijau, dan biru (Ardiansyah, 2013).

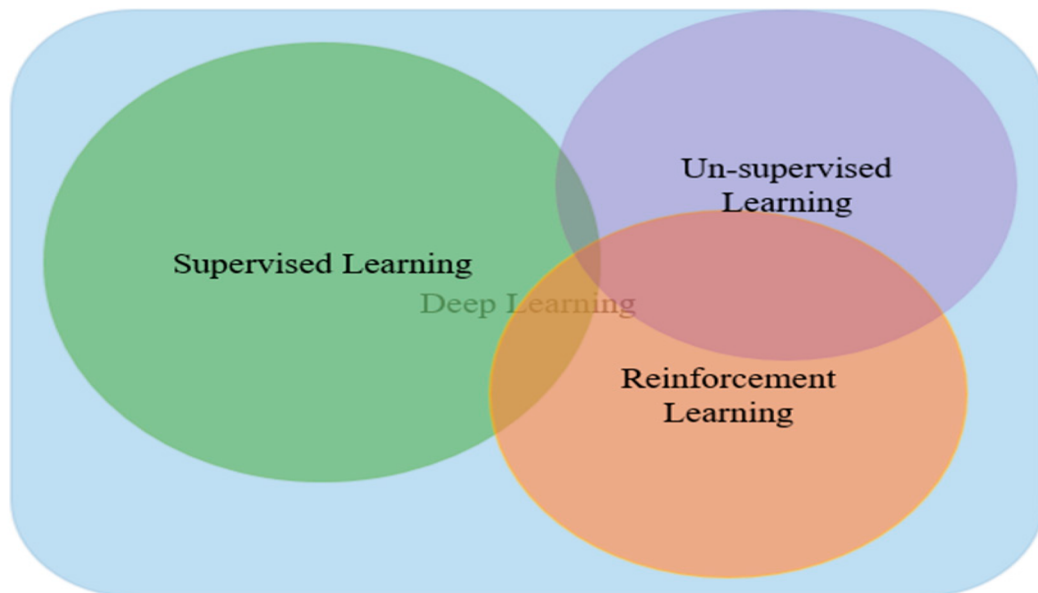
Di bidang kesehatan, pencitraan medis memegang peranan penting dalam berbagai bidang klinis seperti prosedur medis yang digunakan untuk deteksi dini, pemantauan, diagnosis, dan evaluasi pengobatan berbagai kondisi medis (Puttagunta and Ravi, 2021). Citra medis merujuk pada gambar dua dimensi yang menggambarkan struktur internal tubuh manusia, dimanfaatkan oleh profesional kesehatan untuk diagnosis penyakit. Pengolahan citra medis memiliki aplikasi luas, termasuk deteksi tumor atau kanker pada organ reproduksi wanita, identifikasi patologi pada organ-organ seperti paru-paru, hati, dan tulang, segmentasi struktur tulang dari jaringan otot, klasifikasi gigi, serta analisis gambar dari mikroskop. Berbagai teknologi digunakan dalam pencitraan medis ini, antara lain *Magnetic Resonance Imaging* (MRI), *X-Ray*,

Ultrasonography (USG), endoskopi, *Computed Tomography* (CT-Scan), dan *Nuclear Medicine* (Kusuma and Ellyana, 2018). Untuk memahami citra medis diperlukan bidang keilmuan yang mempelajari suatu asal usul penyakit yang disebut patologi (Cohnheim, 1889), dan bidang keilmuan yang mempelajari yang mempelajari cara mendiagnosa penyakit dengan menggunakan citra medis yang disebut radiologi (Wang and Summers, 2012).

2.2 DEEP LEARNING

Deep Learning adalah cabang ilmu *machine learning* yang berbasis pada *artificial neural network* sebagai pengembangannya. Metode dalam *machine learning* biasanya hanya mengandalkan alat seperti CPU (*Central Processing Unit*) dan RAM (*Random Access Memory*) dalam menentukan kecepatan komputasi. Sedangkan metode *deep learning* selain menggunakan CPU dan RAM, Metode ini juga menggunakan GPU (*Graphics Processing Unit*) sehingga proses komputasi data yang besar dapat dilakukan dengan cepat (Ilahiyah and Nilogiri, 2018).

Deep Learning memiliki pendekatan yang dapat dikategorikan seperti *supervised learning*, *unsupervised learning*, *semi-supervised learning* dan *reinforcement learning* (Alom et al., 2019). Gambar 2.1 menunjukkan diagram dari pendekatan pada *deep learning*.



Gambar 2.1. Kategori Pendekatan *Deep Learning* (Alom et al., 2019)

Pada bidang kesehatan *deep learning* digunakan untuk memproses citra medis seperti X-rays, CT, dan scan MRI *Magnetic Resonance Imaging* dan mendiagnosa kondisi kesehatan (Kelleher, 2019).

2.2.1. *Artificial Neural Network*

Artificial Neural network atau jaringan saraf tiruan adalah jaringan saraf yang memproses informasi dengan cara yang mirip dengan otak manusia (Kristiyanti and Saputra, 2023). *Neural network* terdiri dari elemen pemrosesan sederhana yang disebut *node* yang saling terhubung, mirip dengan cara kerja neuron dalam otak manusia. Kemampuan untuk melakukan pemrosesan dalam jaringan ini disimpan dalam koneksi antara *node*, yang biasanya disebut sebagai *weight*. Nilai-nilai *weight* ini diperoleh melalui proses pembelajaran atau adaptasi yang berdasarkan pada pola data yang dipelajari oleh *artificial neural network* (Gurney, 1997)

2.2.2. *Fungsi Aktivasi*

Fungsi aktivasi dalam konteks jaringan saraf dapat diibaratkan dengan cara tubuh manusia merespon rangsangan dari lingkungan. Ketika seseorang menerima rangsangan eksternal, tubuhnya secara otomatis meresponsnya. Sebagai contoh, ketika tangan kita digigit, tubuh kita akan merespons dengan menolak atau melepaskan gigitan tersebut. Respons tubuh ini akan semakin intens jika rangsangan yang diterima semakin kuat. Dalam konteks algoritma jaringan saraf, respons tubuh ini analoginya digantikan oleh nilai bobot dan tingkat aktivasi yang tinggi. Dengan pengaruh dari fungsi aktivasi, persamaannya akan berubah menjadi seperti berikut ini:

$$z = Act(w_1x_1 + w_2x_2 + w_3x_3 + b) \quad (2.1)$$

Persamaan 2.1 ini menunjukkan bahwa nilai keluaran z akan bergantung pada fungsi aktivasi terhadap nilai prediksi yang dihasilkan dari perkalian nilai *input* dan bobot (Kristiyanti and Saputra, 2023).

Fungsi aktivasi adalah komponen penting dalam jaringan saraf tiruan yang mengatur bagaimana nilai hasil penjumlahan terbobot dari *input* data diubah menjadi *output* yang dikeluarkan oleh neuron dalam jaringan saraf. Fungsi aktivasi ini digunakan untuk menentukan apakah suatu neuron akan meneruskan nilai kalkulasinya ke neuron berikutnya, berdasarkan suatu nilai ambang tertentu. Fungsi ini juga sering disebut sebagai fungsi transfer karena memiliki kemampuan untuk mengubah data yang dihasilkan dari hasil penjumlahan terbobot pada suatu lapisan, yang kemudian akan diteruskan ke lapisan selanjutnya. Fungsi aktivasi bisa berupa fungsi linear atau fungsi non-linear tergantung pada tugas yang ingin diselesaikan, dan fungsi aktivasi ini dapat digunakan dalam berbagai hal seperti pengenalan objek dan klasifikasi (Nwankpa et al., 2018).

Fungsi aktivasi perlu memiliki sifat diskriminatif, yang merupakan aspek yang penting karena memungkinkan penggunaan proses propagasi balik kesalahan dalam

pelatihan jaringan. Salah satu fungsi aktivasi yang umum digunakan dalam konteks CNN adalah fungsi ReLU (*Rectified Linear Unit*). Fungsi ini merupakan fungsi aktivasi yang mengubah seluruh isi nilai *input* menjadi angka positif (Alzubaidi et al., 2021). Persamaan fungsi aktivasi ini dapat dilihat pada persamaan 2.2.

$$f(x)_{\text{ReLU}} = \max(0, x) \quad (2.2)$$

Salah satu fungsi aktivasi yang dapat digunakan untuk mengklasifikasi lebih dari dua kelas adalah fungsi aktivasi *softmax*. Persamaan 2.3 menunjukkan persamaan dari fungsi aktivasi *softmax*.

$$f_j(Z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \quad (2.3)$$

Pada persamaan 2.3, Notasi f_j menunjukkan hasil fungsi pada setiap elemen ke- j pada vektor keluaran kelas. Argumen z merupakan hipotesis yang diberikan oleh model pelatihan agar dapat diklasifikasi oleh fungsi *softmax* (Ilahiyah and Nilogiri, 2018).

2.2.3. Loss Function

Loss function adalah salah satu fungsi pada *artificial neural network* untuk melakukan perhitungan nilai *error* atau kesalahan dari hasil prediksi dari suatu *output artificial neural network* (Zhang, 2016).

Loss function yang menghukum kesalahan probabilitas *false negative* daripada *false positive* adalah *categorical cross entropy* (Ho and Wookey, 2019). Persamaan 2.4 menunjukkan persamaan dari *categorical cross-entropy*.

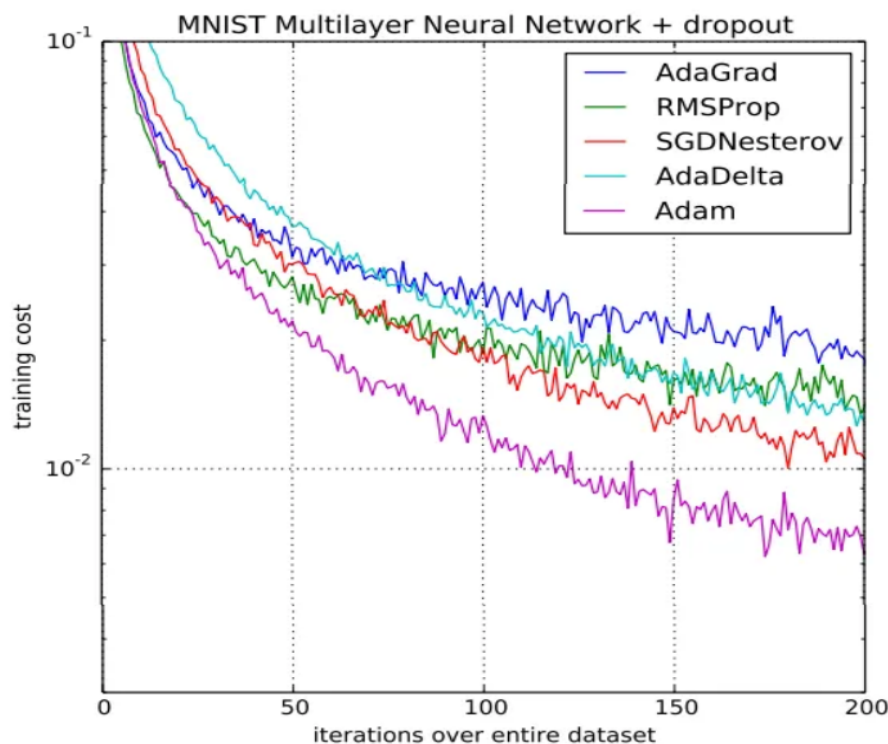
$$J_{cce} = -\frac{1}{M} \sum_{k=1}^K \sum_{m=1}^M y_m^k \log_{(\theta)}(x_m, k) \quad (2.4)$$

- M adalah jumlah sampel data yang digunakan untuk pelatihan.
- K adalah jumlah kelas yang ada pada data.
- y_m^k adalah nilai target dari sampel data ke- m pada kelas ke- k .
- x adalah *input* untuk contoh pelatihan ke- m .
- H_{θ} adalah bobot model *neural network* θ .

2.2.4. Optimization Function

Optimization function atau fungsi optimisasi dapat diartikan sebagai suatu fungsi yang berperan sebagai *black box*, dimana fungsi ini menerima kesalahan sebagai input dan menghasilkan nilai bobot yang optimal untuk suatu model *artificial neural network* (Li and Malik, 2017).

Salah satu dari fungsi optimisasi yang dapat digunakan dalam pengembangan model *deep learning* adalah Adam (*Adaptive Moment Estimation*). Adam adalah teknik optimisasi untuk *gradient descent*. Metode ini sangat efisien saat bekerja dengan masalah yang melibatkan banyak data atau parameter. Algoritma ini membutuhkan memori yang lebih sedikit dan efisien. Secara intuitif, Adam merupakan gabungan antara algoritma *stochastic gradient descent momentum* dan *RMSProp*. Secara eksperimen Adam adalah teknik optimisasi terbaik dengan *training cost* yang rendah menurut (Kingma and Ba, 2014). Gambar 2.2 menunjukkan perbandingan teknik optimisasi Adam dengan teknik optimisasi lainnya.

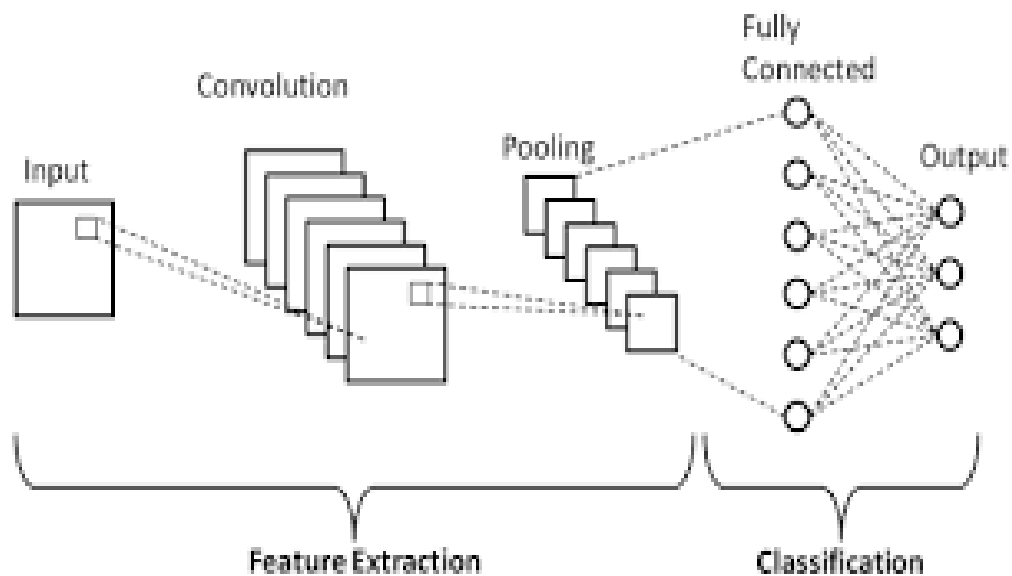


Gambar 2.2. Perbandingan teknik optimisasi Adam dengan teknik optimisasi lainnya (Kingma and Ba, 2014)

2.3 CONVOLUTIONAL NEURAL NETWORK

CNN (*Convolutional Neural Network*) adalah turunan dari algoritma *neural network* yang dikhususkan untuk memproses data yang berupa gambar. CNN adalah

algoritma yang meniru proses pengolahan visual yang terjadi pada manusia. Seperti mata manusia yang berfungsi sebagai alat input, CNN menggunakan lapisan konvolusi yang terdiri dari miliaran neuron untuk memproses informasi visual dan menghasilkan prediksi terhadap objek yang diamati (Kristiyanti and Saputra, 2023). Algoritma CNN dirancang dengan neuron yang berfungsi mirip dengan cara area penglihatan pada otak manusia dan hewan bekerja, seperti yang dijelaskan oleh (Henningsen-Schomers and Pulvermüller, 2022). Arsitektur ini terdiri dari sejumlah lapisan, yang biasanya disebut sebagai blok-blok multi-bangunan.



Gambar 2.3. Arsitektur CNN (Monedero et al., 2021)

Pada gambar 2.3 dapat dilihat bahwa konvolusi merupakan langkah awal dalam pengolahan gambar yang digunakan untuk mengekstraksi fitur penting dari gambar input. Dalam konvolusi, hubungan antar piksel dipertahankan dengan cara mengoperasikan kotak kecil pada masukan untuk memahami fitur-fitur gambar. Konvolusi melibatkan operasi matematika linear yang mencakup perkalian antara matriks gambar dan filter (kernel) yang merupakan matriks bobot dua dimensi

Fungsi dari lapisan *pooling* adalah untuk secara bertahap mengurangi ukuran representasi spasial gambar, sehingga mengurangi jumlah parameter yang dibutuhkan dalam jaringan. Lapisan *pooling* biasanya ditempatkan di antara lapisan-lapisan konvolusi. Lapisan ini beroperasi secara independen pada setiap peta fitur (Monedero et al., 2021).

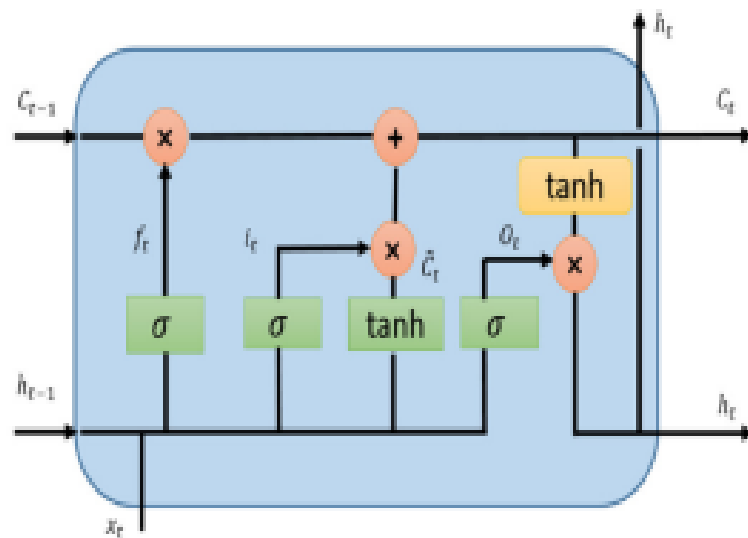
CNN merupakan terobosan besar dalam pengenalan gambar. Mereka belajar langsung dari data gambar, menggunakan pola untuk mengklasifikasikan gambar dan menghilangkan kebutuhan ekstraksi fitur manual. Saat ini, CNN merupakan topik yang

menarik dalam *machine learning*, dan telah menunjukkan kinerja yang sangat baik dalam klasifikasi (Khan et al., 2020).

2.4 LONG SHORT TERM MEMORY

LSTM (*long short term memory*) adalah tipe dari RNN (*recurrent neural network*) yang diciptakan untuk menangani data yang bersifat *sequential* seperti data *time series*, *speech*, dan *text*. LSTM ini dikembangkan untuk mengatasi masalah *vanishing gradient* yang ada dalam RNN tradisional, yang membuatnya sulit bagi jaringan untuk mempelajari ketergantungan jangka panjang (Brownlee, 2017).

Menurut (Alom et al., 2019) LSTM adalah model jaringan saraf yang menggunakan *cell state* untuk menyimpan informasi dari input sebelumnya. *Cell state* memiliki tiga gerbang: *input gate* (i_t), *forget gate* (f_t), dan *output gate* (o_t) yang bisa dilihat pada gambar 2.4:



Gambar 2.4. Diagram untuk LSTM (Alom et al., 2019)

Input gate (i_t) digunakan untuk mengontrol pengaruh data yang masuk saat ini terhadap bobot unit tersebut, *forget gate* (f_t) bertujuan untuk mengendalikan pengaruh riwayat informasi pada bobot unit saat ini, *output gate* (o_t) bertujuan dalam mengendalikan ekspor nilai keadaan unit memori (Huang et al., 2021)

2.5 TRANSFER LEARNING

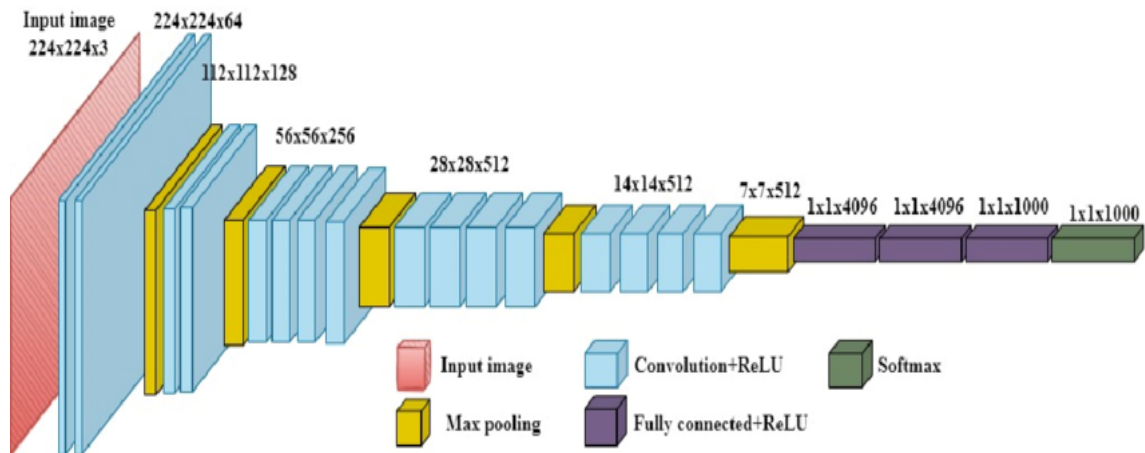
Transfer learning digunakan untuk meningkatkan pembelajaran dari satu domain dengan mentransfer informasi dari domain terkait. Kita dapat mengambil pengetahuan dunia nyata non-teknis untuk memahami mengapa *transfer learning* memungkinkan.

Pertimbangkan contoh dua orang yang ingin belajar bermain piano. Satu orang tidak memiliki pengalaman sebelumnya dalam bermain musik, dan orang lain memiliki pengetahuan musik yang luas melalui bermain gitar. Orang dengan latar belakang musik yang luas akan dapat belajar piano dengan lebih efisien dengan mentransfer pengetahuan musik yang sudah dipelajari sebelumnya ke tugas belajar bermain piano. Satu orang dapat mengambil informasi dari tugas yang sudah dipelajari sebelumnya dan menggunakannya secara bermanfaat untuk belajar tugas yang terkait (Pan and Yang, 2009).

Transfer learning pada *artificial neural network* bisa dibayangkan seperti seseorang yang belajar menjadi lebih mudah, cepat, dan akurat dalam memahami tugas dan konsep baru jika mereka sudah memiliki pengalaman belajar yang serupa dengan konsep baru yang ingin dipelajari. Ini mirip dengan bagaimana seorang individu dapat lebih mudah memahami fisika setelah belajar matematika atau bagaimana seseorang dapat lebih lancar mengendarai truk setelah menguasai kemampuan mengemudi mobil. Pada dasarnya, transfer learning terjadi ketika pemahaman tentang suatu konteks dipengaruhi oleh pemahaman sebelumnya tentang konteks yang mirip. (Cireşan et al., 2012).

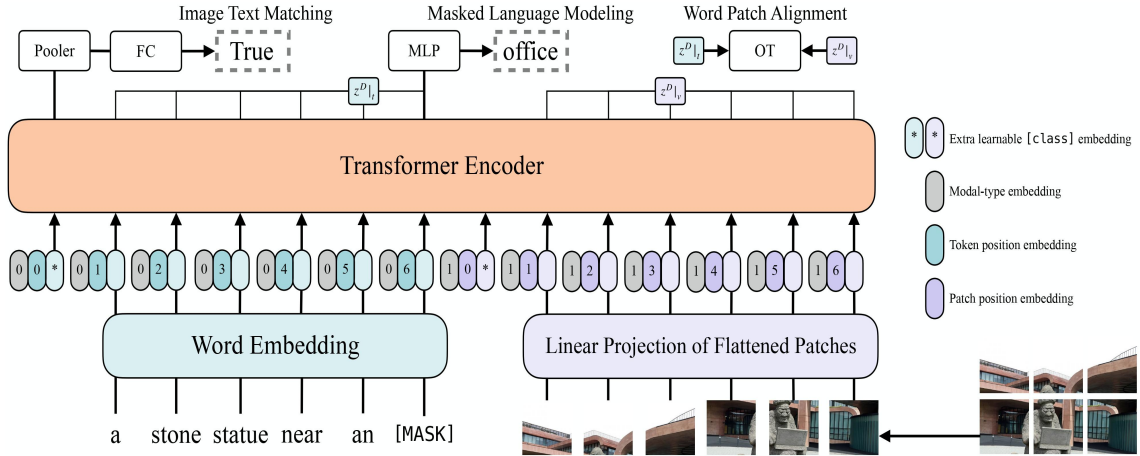
Model dasar yang digunakan dalam *transfer learning* adalah *pre-trained* model, dimana bobot di seluruh jaringan saraf tiruan sudah disesuaikan untuk data tertentu. Ini memungkinkan model untuk memiliki pemahaman yang lebih mendalam tentang fitur dasar dan fitur tingkat tinggi yang dapat mempercepat proses pelatihan. Secara konsisten, model jaringan saraf yang telah dilatih sebelumnya memberikan hasil prediksi yang lebih tepat dibandingkan dengan jaringan saraf yang dimulai dengan bobot-bobot yang dinisiasi secara acak dalam konteks masalah yang melibatkan data yang sudah memiliki label (Cireşan et al., 2012).

VGG19 merupakan salah satu model *pre-trained* yang digunakan dalam penelitian ini. VGG19 adalah model CNN yang memiliki 19 lapisan. Model ini dilatih pada dataset ImageNet yang memiliki 1000 kelas dan 1,2 juta gambar. Model ini memiliki 16 lapisan konvolusi dan 3 lapisan *fully connected*. Model ini memiliki 138 juta parameter dan 20 miliar operasi (Simonyan and Zisserman, 2014). Arsitektur VGG19 dapat dilihat pada gambar 2.5.



Gambar 2.5. Arsitektur VGG19 (Nguyen et al., 2022)

ViLT (*Vision-and-Language Transformer*) adalah suatu struktur model yang menggabungkan informasi dari unsur visual dan teks untuk berbagai tugas yang melibatkan unsur visual dan bahasa, seperti penggantian keterangan gambar (*image captioning*), menjawab pertanyaan visual (*visual question answering*), dan lain-lain. ViLT merupakan perkembangan lebih lanjut dari model VLP (*Vision and Language Pre-training*) dan diciptakan dengan tujuan untuk mengatasi beberapa masalah yang ada pada model VPL sebelumnya. Keunggulan dari model ViLT terletak pada kemampuannya yang lebih baik dibandingkan dengan model VLP, yang disebabkan oleh pendekatan yang lebih sederhana dan efisien dalam memproses informasi visual masukan, berbeda dengan model VPL yang memerlukan ekstraksi fitur gambar yang kompleks, seperti deteksi objek dan arsitektur konvolusi. Arsitektur ViLT mengurangi ketergantungan pada ekstraksi fitur gambar yang rumit dan mengusulkan suatu pendekatan yang lebih langsung dalam memproses informasi visual masukan, sehingga menghasilkan kecepatan yang lebih tinggi, tetapi tetap mempertahankan atau meningkatkan kinerja dalam tugas-tugas turunan yang melibatkan unsur visual dan bahasa (Kim et al., 2021). Struktur arsitektur ViLT dapat dilihat pada gambar 2.6.



Gambar 2.6. Arsitektur ViLT (Kim et al., 2021)

2.6 PERFORMANCE METRICS

Performance metrics (metrik kinerja) yang digunakan dalam menilai kinerja model *medical VQA* dapat dikategorikan menjadi dua jenis, yaitu *classification-based metrics* dan *language-based metrics*. Pada umumnya metrik yang digunakan pada kasus klasifikasi, seperti *accuracy* dan *F1-score*. Metrik ini memperlakukan jawaban sebagai hasil dari klasifikasi dan menghitung pencocokan yang tepat untuk *accuracy*, *precision*, *recall* dan lain-lain. Sedangkan metrik yang digunakan pada kasus *language-based* bertujuan untuk mengevaluasi tugas seperti penerjemah, *image captioning*, dan lain-lain. *dataset* seperti VQA-Med-2018, VQA-Med-2019, PathVQA, VQA-Med-2020, dan VQA-Med-2021 menggunakan metrik *language-based* seperti *BLEU*, untuk mengevaluasi kinerjanya (Lin et al., 2023).

2.6.1. Accuracy Score

Pada umumnya *accuracy* adalah metrik yang mengukur perbandingan antara jumlah prediksi yang benar dengan jumlah total kasus yang dievaluasi (Hossin and Sulaiman, 2015).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.5)$$

- TP (*True Positive*) adalah jumlah prediksi yang benar positif.
- TN (*True Negative*) adalah jumlah prediksi yang benar negatif.
- FP (*False Positive*) adalah jumlah prediksi yang salah positif.
- FN (*False Negative*) adalah jumlah prediksi yang salah negatif.

2.6.2. BLEU Score

BLEU (*bilingual evaluation understudy*) adalah metrik yang digunakan untuk mengukur kesamaan pada frase (n-grams) antara dua kalimat. BLEU adalah metrik original untuk penerjemah mesin dan juga bisa digunakan untuk tugas seperti pembuatan laporan medis (Li et al., 2021).

BLEU adalah metrik yang digunakan untuk mengevaluasi kualitas sistem terjemahan mesin (Papineni et al., 2002). Ini mengukur kemiripan antara terjemahan yang dihasilkan oleh mesin dan satu atau lebih referensi terjemahan menggunakan metrik presisi yang dimodifikasi dan hukuman singkat yang dimodifikasi dan *brevity penalty* yang dimodifikasi.

$$\text{BLEU} = \text{BP} \times \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (2.6)$$

- BP adalah *brevity penalty*.
- N adalah jumlah maksimal n-gram.
- w_n adalah bobot untuk n-gram.
- p_n adalah presisi n-gram.

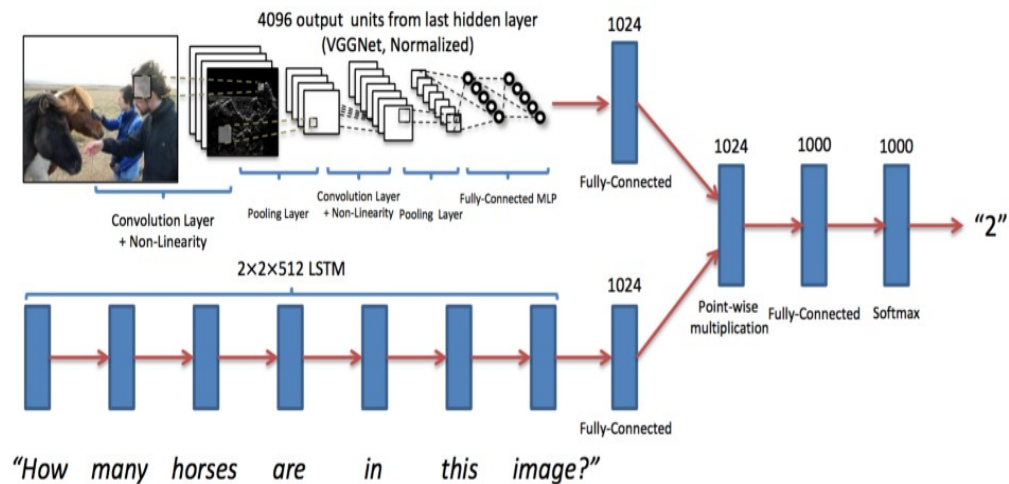
Berdasarkan persamaan 2.6 BP (*brevity penalty*) untuk menghukum jawaban singkat, w_n adalah bobot antara 0 dan 1 untuk $\log p_n$ dan $\sum_{n=1}^N w_n = 1$, p_n adalah nilai rata-rata geometris dari presisi n-gram yang dimodifikasi, dan N adalah panjang maksimum n-gram. N-gram disini mencakup n-gram hingga panjang 4.

2.7 VISUAL QUESTION ANSWERING

VQA (*visual question answering*) adalah suatu tugas tentang bagaimana menjawab pertanyaan bebas pada sebuah gambar. Karena membutuhkan pemahaman bahasa yang mendalam tentang pertanyaan dan kemampuan untuk mengaitkannya dengan berbagai objek yang ada dalam gambar. Ini adalah tugas yang ambisius dan memerlukan teknik dari baik *computer vision* dan *natural language processing* (Hildebrandt et al., 2020).

Salah satu contoh percobaan yang dilakukan oleh (Antol et al., 2015), mereka menerapkan *baseline Vanilla VQA* model yang dimana sebagai tolak ukur untuk metode *deep learning*. Model ini menggunakan CNN untuk ekstraksi fitur dan jaringan LSTM atau *recurrent network* untuk pemrosesan bahasa. Fitur-fitur ini digabungkan menggunakan operasi berdasarkan elemen ke dalam fitur bersama, yang digunakan

untuk mengklasifikasikan salah satu jawaban seperti yang ditunjukkan dalam gambar 2.7.



Gambar 2.7. Vanilla VQA Network Model (Srivastava et al., 2021)

VQA telah menarik minat besar dan mengalami perkembangan oleh para peneliti dan ilmuwan dari seluruh dunia. Tren terbaru yang diamati adalah dalam pengembangan *dataset* yang terlihat semakin mirip dengan dunia nyata dengan menggabungkan pertanyaan dan jawaban seputar kehidupan nyata. Tren terkini juga terlihat dalam pengembangan model *deep learning* yang lebih canggih dengan lebih baik memanfaatkan petunjuk visual dan petunjuk teks melalui berbagai cara. Kinerja model terbaik saat ini masih tertinggal dan hanya sekitar 60-70% saja. Oleh karena itu, masih merupakan masalah yang terbuka untuk mengembangkan model *deep learning* yang baik serta *dataset* yang lebih menantang untuk *visual question answering* (Srivastava et al., 2021).

2.8 MEDICAL VISUAL QUESTION ANSWERING DATASET

Medical visual question answering dataset adalah *dataset* yang digunakan untuk melakukan penelitian terkait tugas yang berkaitan dengan *medical visual question answering* (Lin et al., 2023). *Dataset* ini berisi gambar medis dan pertanyaan yang terkait dengan gambar tersebut. Pada penelitian ini akan digunakan *dataset* PathVQA (Xuehai et al., 2020) dan VQA-RAD (Lau et al., 2018). Berikut adalah penjelasan dari kedua *dataset* tersebut.

2.8.1. PathVQA Dataset

Dataset PathVQA (Xuehai et al., 2020) berisi 32,799 pertanyaan dan jawaban dari 1,670 gambar patologi yang dikumpulkan dari dua buku patologi yaitu "*Textbook of Pathology*", dan "*Basic Pathology*". Sedangkan 3.328 gambar lainnya dikumpulkan dari perpustakaan digital PEIR. Rata-rata setiap satu gambar memiliki 6.6 pertanyaan. Jumlah pertanyaan maksimal dan minimal adalah 14 dan 1 berturut-turut. Rata-rata jumlah kata per pertanyaan dan per jawaban adalah 9.5 dan 2.5 berturut-turut. Pada *dataset* ini terdapat 7 kategori pertanyaan yaitu: *what, where, when, whose, how, how much/ how many*, dan *yes/no*. Pertanyaan dari enam kategori pertama bersifat *open-ended*, sedangkan pertanyaan dari kategori terakhir bersifat *close-ended* "*yes/no*". Jumlah jawaban "*yes*" dan "*no*" adalah 8.145 dan 8.189 berturut-turut. Pertanyaan pada *dataset* ini mencakup banyak aspek konten visual, termasuk warna, lokasi, penampilan, bentuk dan lain-lain. Keragaman klinis ini menimbulkan tantangan yang besar bagi model AI dalam menjawab permasalahan pada gambar patologi.

2.8.2. VQA-RAD Dataset

Dataset VQA-RAD (Lau et al., 2018) adalah *dataset* yang dibuat secara manual untuk penelitian terkait *Medical Visual Question Answering* (VQA). *Dataset* ini berisi 3.515 pasangan pertanyaan dan jawaban yang dihasilkan oleh para klinisi, serta 315 gambar radiologi yang terbagi merata antara kepala, dada, dan perut. Setiap gambar terkait dengan beberapa pertanyaan. Pertanyaan-pertanyaan ini dikelompokkan ke dalam 11 kategori: kelainan, atribut, modalitas, sistem organ, warna, penghitungan, keberadaan objek/kondisi, ukuran, bidang, penalaran posisional, dan lain-lain. Setengah dari jawaban pada *dataset* ini bersifat *close-ended* (*yes/no*), sedangkan setengahnya lagi bersifat *open-ended*.

2.9 PENELITIAN TERKAIT

Salah satu penelitian dari (Xuehai et al., 2020), Dalam riset mereka berhasil menghadapi tantangan khusus yang muncul dalam analisis patologi dengan mengintegrasikan metode pemrosesan gambar dan bahasa alami yang canggih. Mereka menggunakan berbagai teknik termasuk pra-pemrosesan data, penggunaan arsitektur CNN seperti Faster R-CNN, ResNet-152, dan VGGNet, serta representasi kata menggunakan vektor GloVe yang telah di-latih sebelumnya. Pendekatan ini membantu mereka mengembangkan pemahaman yang lebih mendalam tentang bagaimana AI (*Artificial Intelligence*) dapat digunakan untuk mendukung diagnosa dan analisis medis dengan menggabungkan informasi visual dan teks. Hasil penelitian mereka memiliki implikasi penting untuk meningkatkan akurasi dan efisiensi dalam penanganan

kasus-kasus terkait gambar patologi melalui bantuan AI (*Artificial Intelligence*).

Penelitian lain juga dilakukan dengan mengembangkan pendekatan berbasis arsitektur encoder-decoder transformer oleh (Bazi et al., 2023). Secara khusus, mereka mengambil fitur dari gambar menggunakan model vision transformer (ViT) dan memasukkan pertanyaan ke dalam transformer encoder teks. Kemudian, representasi visual dan teks digabungkan dan dimasukkan ke dalam decoder multi-modal untuk menghasilkan jawaban secara otomatis. Dalam uji coba, model yang diusulkan diuji pada dua *dataset* VQA yaitu *dataset* PathVQA dan *dataset* VQA-RAD yang berfokus pada gambar radiologi dan patologi.

Penelitian yang dilakukan oleh (Nguyen et al., 2019), mereka mengemukakan sebuah konsep baru dalam bidang VQA medis yang memanfaatkan teknologi MAML (*model-agnostic meta learning*) dan CDAE (*convolutional denoising autoencoder*) untuk mengekstraksi fitur gambar. Hal ini bertujuan untuk mengatasi keterbatasan data pelatihan yang berlabel. Secara lebih rinci, CDAE membantu dalam mengambil manfaat dari informasi yang terdapat di dalam gambar yang tidak memiliki label dalam jumlah besar, Sementara MAML membantu dalam mempelajari bobot meta yang dapat dengan cepat diadaptasi untuk permasalahan *visual question answering*. Hasil eksperimen ini menunjukkan pencapaian baru yang sangat baik pada *dataset* VQA-RAD, baik dalam hal pertanyaan yang bersifat *open-ended* maupun *close-ended*.

Penelitian yang dilakukan oleh (Do et al., 2021), mereka mengusulkan metode baru untuk mengukur beberapa meta-model secara efektif dalam rangka memanfaatkan meta-anotasi dan mengatasi label yang tidak akurat dalam tugas VQA medis. Hasil eksperimen yang melibatkan sejumlah besar data menunjukkan bahwa metode yang dihasilkan dari penelitian ini jauh lebih unggul dibandingkan dengan metode berbasis meta-learning terkini dalam kedua *dataset* PathVQA dan VQA-RAD.

BAB III METODOLOGI PENELITIAN

3.1 WAKTU DAN LOKASI PENELITIAN

Penelitian ini akan bertempat di Ruang Lab *Database* dan *Data Mining*. Waktu yang dibutuhkan agar penelitian ini dapat di implementasikan adalah 4 bulan terhitung dari bulan Januari 2024 hingga Mei 2024.

3.2 JADWAL PELAKSANAAN

Untuk lebih memahami alur waktu dari peneltian ini, penulis telah menyusun sebuah jadwal yang rinci dari masing-masing periode tertentu. Untuk pelaksanaan bulan Agustus hingga Oktober 2023 dapat dilihat pada Tabel 3.1.

Tabel 3.1. Jadwal pelaksanaan bulan agustus hingga Oktober 2023

No	Kegiatan	Agustus				September				Oktober			
		Minggu Ke -				Minggu Ke -				Minggu Ke -			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Pengajuan Judul												
2	Studi Literatur												
3	Penyusunan BAB I												
4	Penyusunan BAB II												
5	Penyusunan BAB III												

Sedangkan pelaksanaan bulan November 2023 hingga Januari 2024 dapat dilihat pada Tabel 3.2 berikut ini.

Tabel 3.2. Jadwal pelaksanaan November 2023 hingga Januari 2024

No	Kegiatan	November				Desember				Januari			
		Minggu Ke -				Minggu Ke -				Minggu Ke -			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Mengecek kembali proposal												
2	Proposal ACC												

Tabel 3.3. Jadwal pelaksanaan Februari hingga April 2024

No	Kegiatan	Februari				Maret				April			
		Minggu Ke -				Minggu Ke -				Minggu Ke -			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Seminar Proposal												
2	Revisi Proposal												
3	Pelaksanaan Penelitian												
4	Penyusunan BAB IV												
5	Penyusunan BAB V												
6	Ujian Hasil Skripsi												
7	Revisi Hasil Skripsi												

3.3 ALAT DAN BAHAN

Alat dan Bahan yang akan digunakan pada penelitian ini terdiri dari beberapa perangkat keras (*hardware*) dan perangkat lunak. Lalu, data yang digunakan adalah data dari *dataset* PathVQA dan VQA-RAD, sampel kedua data ini bisa dilihat pada Gambar 3.1 dan Gambar 3.2.

image image	question string · lengths 8 255	answer string · lengths 1 220
	where are liver stem cells (oval cells) located?	in the canals of hering
	what are stained here with an immunohistochemical stain for cytokeratin 7?	bile duct cells and canals of hering
	what do the areas of white chalky deposits represent?	foci of fat necrosis
	is embolus derived from a lower-extremity deep venous thrombus lodged in a pulmonary artery branch?	yes
	how is hyperplasia without atypia characterized?	by nests of closely packed glands
	is normal palmar creases present?	no

Gambar 3.1. Sampel data PathVQA

image image	question string · lengths 	answer string · lengths 
	are regions of the brain infarcted?	yes
	are the lungs normal appearing?	no
	which organ system is abnormal in this image?	cardiovascular
	is the lesion causing significant brainstem herniation?	no
	how was this image taken?	mri

Gambar 3.2. Sampel data VQA-RAD

3.3.1. Perangkat Keras

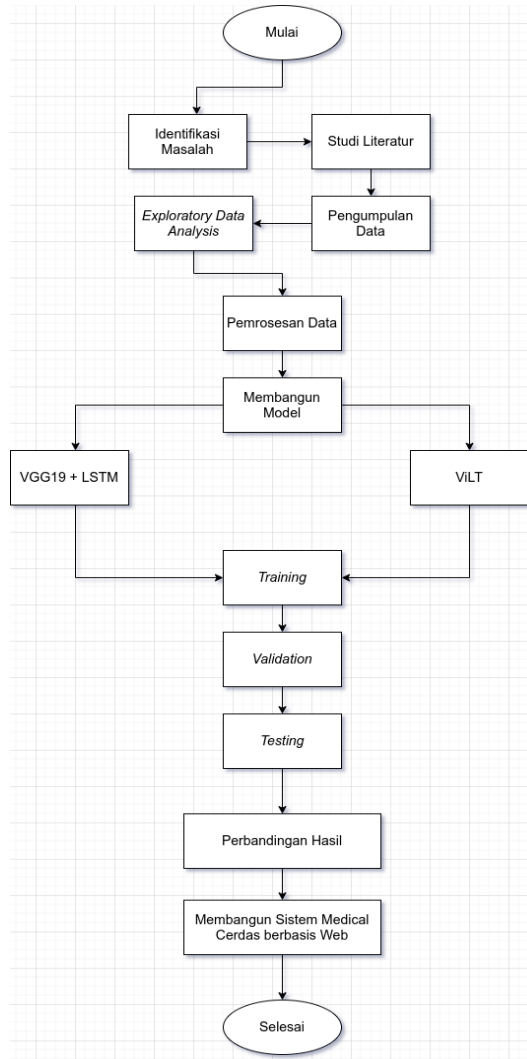
- Laptop Lenovo Yoga C740 dengan RAM 16GB DDR4, Intel® Core™ i7-10710U. 1.10 - 4.70 GHz, *Solid State Drive* (SSD) 1TB.
- Server spesifikasi prosesor Intel Xeon Gold 5218, CPU 2.30GHz 64 inti, RAM 128GB, VGA 4 x NVIDIA GeForce RTX 2080 Ti GPU VRAM 12GB, dan memory 8TB.

3.3.2. Perangkat Lunak

- Linux Debian Ubuntu versi 22.04 LTS
- Visual Studio Code versi 1.71.0
- Python versi 3.8.17
- Jupyter Lab
- Postman versi 9.31.0

3.4 METODE PENELITIAN

Metode penelitian yang dilakukan akan terdiri dari beberapa tahapan penelitian. Skema dari alur tahapan tersebut dapat dilihat pada Gambar 3.3.



Gambar 3.3. Diagram Alir Penelitian

3.4.1. Studi Literatur

Studi literatur dilakukan untuk memperoleh informasi mengenai penelitian yang sudah dilakukan sebelumnya. Studi literatur ini terdiri dari banyak kegiatan yaitu membaca referensi yang berkaitan dengan penelitian ini, membaca jurnal atau buku untuk menambah wawasan, sehingga dapat memperoleh informasi yang dibutuhkan untuk penelitian ini.

3.4.2. Pengumpulan Data

Penelitian tugas akhir ini menggunakan data dari *dataset* PathVQA dan VQA-RAD. Pengumpulan *dataset* dilakukan dengan mengunduh data dari *platform* Hugging Face.

3.4.3. *Exploratory Data Analysis*

Setelah data yang digunakan sudah terkumpul, maka dilakukan *exploratory data analysis* yang bertujuan untuk mengetahui karakteristik dari data yang digunakan pada penelitian.

3.4.4. Pemrosesan Data

Pemrosesan data dilakukan dengan tujuan untuk mengubah data yang digunakan menjadi data yang dapat digunakan untuk melatih model.

3.4.5. Membangun Model

Proses pembangunan model VQA ini menggunakan dua metode yaitu dengan arsitektur VGG19 digabungkan dengan LSTM, dan arsitektur ViLT (*Vision and Language Transformer*). Kedua metode tersebut akan dibandingkan hasilnya untuk mengetahui metode mana yang lebih baik.

3.4.6. Melatih Model

Setelah model berhasil dibangun, maka dilakukan pelatihan model. Pelatihan model dilakukan dengan menggunakan data latih yang telah diproses sebelumnya. Data latih tersebut akan digunakan untuk melatih model dan menghasilkan model yang dapat digunakan untuk menguji model.

3.4.7. Perbandingan Hasil

Setelah model berhasil dibangun, maka dilakukan perbandingan hasil dari kedua model yang telah dibangun. Perbandingan hasil dilakukan dengan menguji model menggunakan data uji. Data uji tersebut akan digunakan untuk menguji model dan menghitung akurasi dari model yang telah dibangun. Setelah itu, dilakukan perbandingan hasil dari kedua model yang telah dibangun untuk mengetahui model mana yang lebih baik.

3.4.8. Membangun Sistem Medical Cerdas Berbasis Web

Model terbaik akan di *deploy* ke dalam sistem medical cerdas berbasis web. *Website* ini bertujuan untuk menjawab hal-hal terkait dengan citra medis.

BAB IV

HASIL DAN PEMBAHASAN

4.1 PENGUMPULAN DATA

4.2 PENGEMBANGAN MODEL

4.3 PERBANDINGAN MODEL

4.4 PENGEMBANGAN SISTEM MEDICAL CERDAS BERBASIS WEB

BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

5.2 SARAN

DAFTAR PUSTAKA

- Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A., and Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *electronics*, 8(3):292.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., and Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8:1–74.
- Andono, P. N., Sutojo, T., et al. (2018). *Pengolahan citra digital*. Penerbit Andi.
- Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., and Parikh, D. (2015). Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*, pages 2425–2433.
- Ardiansyah, R. F. (2013). Pengenalan pola tanda tangan dengan menggunakan metode principal component analysis (pca). *Fakultas Ilmu Komputer Universitas Dian Nuswantoro*, 2:14.
- Bazi, Y., Rahhal, M. M. A., Bashmal, L., and Zuair, M. (2023). Vision–language model for visual question answering in medical imagery. *Bioengineering*, 10(3):380.
- Brownlee, J. (2017). A gentle introduction to long short-term memory networks for the experts. *Machine Learning Mastery*.
- Cireşan, D. C., Meier, U., and Schmidhuber, J. (2012). Transfer learning for latin and chinese characters with deep neural networks. In *The 2012 international joint conference on neural networks (IJCNN)*, pages 1–6. IEEE.
- Cohnheim, J. (1889). *Lectures on general pathology*, volume 1. New Sydenham Society.
- Do, T., Nguyen, B. X., Tjiputra, E., Tran, M., Tran, Q. D., and Nguyen, A. (2021). Multiple meta-model quantifying for medical visual question answering. In *MICCAI*.
- Gurney, K. (1997). *An Introduction to Neural Networks*. Taylor & Francis, Inc., USA.
- Henningsen-Schomers, M. R. and Pulvermüller, F. (2022). Modelling concrete and abstract concepts using brain-constrained deep neural networks. *Psychological research*, 86(8):2533–2559.
- Hildebrandt, M., Li, H., Koner, R., Tresp, V., and Günnemann, S. (2020). Scene graph reasoning for visual question answering. *arXiv preprint arXiv:2007.01072*.
- Ho, Y. and Wookey, S. (2019). The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE access*, 8:4806–4813.

- Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2):1.
- Huang, D., Fu, Y., Qin, N., and Gao, S. (2021). Fault diagnosis of high-speed train bogie based on lstm neural network. *Sci. Chin. Inf. Sci.*, 64:1–3.
- Ilahiyah, S. and Nilogiri, A. (2018). Implementasi deep learning pada identifikasi jenis tumbuhan berdasarkan citra daun menggunakan convolutional neural network. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 3(2):49–56.
- Kelleher, J. D. (2019). *Deep learning*. MIT press.
- Khan, A., Sohail, A., Zahoor, U., and Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, 53:5455–5516.
- Kim, W., Son, B., and Kim, I. (2021). Vilt: Vision-and-language transformer without convolution or region supervision. In *International Conference on Machine Learning*, pages 5583–5594. PMLR.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kristiyanti, D. A. and Saputra, I. (2023). Machine learning untuk pemula.
- Kusuma, A. W. and Ellyana, R. L. (2018). Penerapan citra terkompresi pada segmentasi citra menggunakan algoritma k-means. *Jurnal Terapan Teknologi Informasi*, 2(1):65–74.
- Kusuma, I. W. A. W. and Kusumadewi, A. (2020). Penerapan metode contrast stretching, histogram equalization dan adaptive histogram equalization untuk meningkatkan kualitas citra medis mri. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 11(1):1–10.
- Lau, J. J., Gayen, S., Ben Abacha, A., and Demner-Fushman, D. (2018). A dataset of clinically generated visual questions and answers about radiology images. *Scientific data*, 5(1):1–10.
- Li, K. and Malik, J. (2017). Learning to optimize neural nets. *arXiv preprint arXiv:1703.00441*.
- Li, M., Cai, W., Liu, R., Weng, Y., Zhao, X., Wang, C., Chen, X., Liu, Z., Pan, C., Li, M., et al. (2021). Ffa-ir: Towards an explainable and reliable medical report generation benchmark. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.
- Lin, Z., Zhang, D., Tao, Q., Shi, D., Haffari, G., Wu, Q., He, M., and Ge, Z. (2023). Medical visual question answering: A survey. *Artificial Intelligence in Medicine*, page 102611.

- Mauli, D. (2018). Tanggung jawab hukum dokter terhadap kesalahan diagnosis penyakit kepada pasien. *Cepalo*, 2(1):33–42.
- Monedero, Í., Barbancho, J., Márquez, R., and Beltrán, J. F. (2021). Cyber-physical system for environmental monitoring based on deep learning. *Sensors*, 21(11):3655.
- Nguyen, B. D., Do, T.-T., Nguyen, B. X., Do, T., Tjiputra, E., and Tran, Q. D. (2019). Overcoming data limitations in medical visual question answering. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*.
- Nguyen, T.-H., Nguyen, T.-N., and Ngo, B.-V. (2022). A vgg-19 model with transfer learning and image segmentation for classification of tomato leaf disease. *AgriEngineering*, 4(4):871–887.
- Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*.
- Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.
- Puttagunta, M. and Ravi, S. (2021). Medical image analysis based on deep learning approach. *Multimedia tools and applications*, 80:24365–24398.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Srivastava, Y., Murali, V., Dubey, S. R., and Mukherjee, S. (2021). Visual question answering using deep learning: A survey and performance analysis. In *Computer Vision and Image Processing: 5th International Conference, CVIP 2020, Prayagraj, India, December 4-6, 2020, Revised Selected Papers, Part II 5*, pages 75–86. Springer.
- Tschandl, P., Rinner, C., Apalla, Z., Argenziano, G., Codella, N., Halpern, A., Janda, M., Lallas, A., Longo, C., Malvehy, J., et al. (2020). Human–computer collaboration for skin cancer recognition. *Nature Medicine*, 26(8):1229–1234.
- Wang, S. and Summers, R. M. (2012). Machine learning and radiology. *Medical image analysis*, 16(5):933–951.
- Xuehai, H., Yichen, Z., Luntian, M., Eric, X., and Pengtao, X. (2020). Pathvqa: 30000+ questions for medical visual question answering. *arXiv preprint arXiv:2003.10286*.
- Zhang, Z. (2016). Neural networks: further insights into error function, generalized weights and others. *Annals of translational medicine*, 4(16).

LAMPIRAN