PERBANDINGAN ARSITEKTUR *CONVOLUTIONAL NEURAL NETWORK*(CNN) DAN *LONG SHORT TERM MEMORY* (LSTM) DALAM KLASIFIKASI DATASET *EMOTIONAL SPEECH*

SKRIPSI

MUHAMMAD ULIL AMRI

H071181302

PROGRAM STUDI SISTEM INFORMASI

DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

PERBANDINGAN ARSITEKTUR *CONVOLUTIONAL NEURAL NETWORK*(CNN) DAN *LONG SHORT TERM MEMORY* (LSTM) DALAM KLASIFIKASI DATASET *EMOTIONAL SPEECH*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Sistem Informasi Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

MUHAMMAD ULIL AMRI

H071181302

PROGRAM STUDI SISTEM INFORMASI

DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2022

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Ulil Amri
NIM : H071181302
Program Studi : Sistem Informasi
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya dengan judul:

Perbandingan Arsitektur *Convolutional Neural Network*(Cnn) Dan *Long Short Term Memory* (Lstm) Dalam Klasifikasi Dataset *Emotional Speech*

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa Sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 25 Januari 2022
Yang menyatakan,

Muhammad Ulil Amri
H071181302

PERBANDINGAN ARSITEKTUR *CONVOLUTIONAL NEURAL NETWORK*(CNN) DAN *LONG SHORT TERM MEMORY* (LSTM) DALAM KLASIFIKASI DATASET *EMOTIONAL SPEECH*

Disusun dan diajukan oleh:

MUHAMMAD ULIL AMRI

H071181302

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Menyetujui,

Pembimbing Utama	**Pembimbing Pertama**
Dr.Eng. Armin Lawi, S.Si., M.Eng.	**Dr. Hendra, S.Si., M.Kom.**
NIP. 197204281995121001	NIP. 197601022002121001

Ketua Program Studi

Dr. Hendra, S.Si., M.Kom.

NIP. 197601022002121001

Pada tanggal November 2022

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Muhammad Ulil Amri
NIM : H071181302
Program Studi : Sistem Informasi
Judul Skripsi : Perbandingan Arsitektur Convolutional Neural Network (CNN) dan Long Short Term Memory (LSTM) Dalam Klasifikasi Dataset Emotional Speech

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

		Tanda Tangan
Ketua	: Dr.Eng. Armin Lawi, S.Si., M.Eng.	(.....)
Sekretaris	: Dr. Hendra, S.Si., M.Kom.	(.....)
Anggota	: Edy Saputra Rusdi, S.Si., M.Si	(.....)
Anggota	: Rozalina Amran, S.T., M.Eng.	a.n. (.....)

Ditetapkan di : Makassar
Tanggal : November 2022

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS HASANUDDIN DEPARTEMEN MATEMATIKA FAKULTAS MIPA

KATA PENGANTAR

Segala puji bagi Allah Subhanahu Wa ta'ala, Tuhan atas langit dan bumi beserta segala isinya. Karena, berkat nikmat dan karunianya sehingga penulisan skripsi ini dapat terselesaikan. Shalawat serta salam semoga senantiasa tercurahakan kepada Baginda Rasulullah Muhammad Shallallahu Alaihi Wasallam dan kepada para keluarga serta sahabat beliau, yang senantiasa menjadi teladan yang baik.

Alhamdulillah, skripsi dengan judul "Perbandingan Arsitektur Convolutional Neural Network(CNN) Dan Long Short Term Memory (LSTM) Dalam Klasifikasi Dataset Emotional Speech" yang disusun sebagai salah satu syarat akademik untuk meraih gelar Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan. Tentunya, dalam penulisan skripsi ini, penulis mampu melewati berbagai hambatan dan masalah berkat bantuan moril dan materil, serta dorongan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada orang tua penulis, Ayahanda **Muhammad Basir**, dan Ibunda **Amriani**, sebagai tempat kembali setelah pergi, terima kasih atas kasih sayang, doa, dan nasihat yang tulus sebagai bekal kehidupan. Rasa terima kasih juga penulis tujukan kepada saudari tercinta, **Khalizatul Jannah** serta seluruh keluarga yang senantiasa memberikan dukungan dan doa bagi penulis dalam menyelesaikan skripsi ini. Oleh karena itu, penulis menyampaikan ucapan terima kasih dan penghargaan yang tulus kepada:

1. Rektor Universitas Hasanuddin Makassar **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc**.
2. Bapak Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam **Dr. Eng. Amiruddin**, **M.Si** dan para Wakil Dekan serta seluruh staf yang telah memberikan bantuan selama penulis mengikuti Pendidikan di FMIPA Universitas Hasanuddin.

3. Bapak **Prof. Dr. Nurdin, S.Si., M.Si**, selaku Ketua Departemen Matematika FMIPA Unhas. Penulis juga berterima kasih atas dedikasi dosen-dosen pengajar, serta staf Departemen atas ilmu dan bantuan yang bermanfaat.
4. Bapak **Dr. Hendra, S.Si., M.Kom.** sebagai Ketua Program Studi Sistem Informasi Universitas Hasanuddin
5. Bapak **Dr.Eng. Armin Lawi, S.Si., M.Eng.**, dan Bapak **Dr. Hendra, S.Si., M.Kom.** selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan saya dalam penyusunan skripsi ini.
6. Bapak **Edy Saputra Rusdi, S.Si., M.Si.** dan Ibu **Rozalina Amran, S.T., M.Eng.** selaku dosen penguji terima kasih atas ilmu yang diberikan selama proses perkuliahan serta saran dan masukan yang diberikan dalam proses penyusunan skripsi ini.
7. Kepada saudara-saudara seperjuangan grup ***SKUT***, **Islah**, **Maxi**, **Rifky**, **Ramdan**, **Luthfi** terima kasih atas *carry* yang diberikan setiap semester, kebersamaan, kepedulian, dan canda tawa yang telah kita lewati selama ini, semoga kesuksesan selalu kita dapatkan dalam setiap langkah-langkah kita.
8. Kepada teman-teman seperjuangan **Sistem Informasi 2018** tercinta yang selalu menemani, menguatkan dan menyemangati selama masa perkuliahan, serta turut membantu dan memudahkan penulis dalam penyusunan skripsi ini.
9. Keluarga Besar **UKM Karate-Do Universitas Hasanuddin** yang telah mewadahi penulis dalam mengembangkan diri, terutama dibidang organisasi keolahragaan.
10. Teman-teman **KKN Tamalanrea 06,** atas waktu kebersamaan dan berbagi pengalaman saat melaksanakan KKN dan setelahnya.
11. Semua pihak yang telah memberikan bantuan kepada penulis baik berupa materi dan non-materi yang tidak dapat penulis sebutkan satu per satu, terima kasih untuk bantuan dan dukungannya.
12. Kepada *someone*, terima kasih telah menjadi *support system* saya dibalik layar selama pengerjaan skripsi ini. *Someone who willing to spend her time talking with me, even in the ramdom things*. Selalu memberikan saran dan masukan terdahap apapun yang saya lakukan.

13. *Last but not least, i wanna thank me, i wanna thank me for believing in me, i wanna thank me for doing all this hard work, i wanna thank me for having no days off, i wanna thank me for, for never quitting.*

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan bahkan kritik yang membangun dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi yang membacanya, terutama untuk pengembangan ilmu pengetahuan.

Makassar, November 2022

Muhammad Ulil Amri

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan dibawah ini:

Nama : Muhammad Ulil Amri

NIM : H071181302

Program Studi : Sistem Informasi

Departemen : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Perbandingan Arsitektur Convolutional Neural Network(CNN) Dan Long Short Term Memory (LSTM) Dalam Klasifikasi Dataset Emotional Speech

adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa Sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Yang Menyatakan

Muhammad Ulil Amri

ABSTRAK

Manusia berinteraksi melalui tiga saluran utama yaitu penglihatan, peraba dan suara. Ekspresi suara yang pada umumnya dikenali adalah tertawa, bersenandung, dan lain-lain . Walaupun manusia bisa mengenali ekspresi emosional dengan baik, penelitian pengenalan ekspresi emosional yang dilakukan oleh mesin terus dilakukan agar dapat melakukan pengenalan ekspresi emosional dalam interaksi manusia dan komputer. Untuk memanfaatkan sepenuhnya perbedaan saturasi emosional antara kerangka waktu, diusulkan untuk pengenalan ucapan menggunakan ekstraksi fitur yang dimanfaatkan oleh algoritma *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM). Penelitian ini menggunakan arsitektur LSTM dan CNN yaitu VGG16, VGG19, dan InceptionV3 untuk pengklasifikasian emosi suara melalui input berupa citra dan array. Pada penelitian ini dataset yang digunakan yaitu *TESS*, dengan 14 kelas dan 2800 jumlah data. Dengan melakukan *training* sebanyak 100 *epoch* dan akan dilakukan proses evaluasi kinerja model dengan menggunakan *precision*, *recall*, dan *f1-score*.

Kata kunci: Emosi Suara, *Convolutional Neural Network (CNN), Long Short-Term Memory (*LSTM*), Transfer Learning,* InceptionV3, VGG16, VGG19

ABSTRACT

Humans interact through three main channels, namely sight, touch and sound. Voice expressions that are generally recognized are laughing, humming, and others. Even though humans can recognize emotional expressions well, research on emotional expression recognition is carried out by machines so that they can recognize emotional expressions in human-computer interactions. In order to take full advantage of the differences in emotional saturation between timeframes, it is proposed for speech recognition using feature extraction utilized by Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) algorithms. This study uses the LSTM and CNN architectures, namely VGG16, VGG19, and InceptionV3 to classify sound emotions through input in the form of images and arrays. In this study, the dataset used was TESS, with 14 classes and 2800 amounts of data. By training as many as 100 epochs and the process of evaluating the performance of the model will be carried out using precision, recall, and f1-score.

Keywords: Emotional Speech, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Transfer Learning, InceptionV3, VGG16, VGG19.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Manusia berinteraksi melalui tiga saluran utama yaitu visual (penglihatan) haptic (peraba), *audio* (suara). Dalam berinteraksi manusia mengungkapkan emosi dalam bentuk ekspresi. Emosi merupakan suatu kondisi mental seseorang yang dapat mendorongnya untuk melakukan suatu tindakan atau berekspresi yang dapat dipicu dari dalam atau luar dirinya. Manusia memiliki ekspresi emosi yang berbeda dalam menghadapi berbagai situasi dalam kehidupannya. Apabila dikaitkan dengan konteks hubungan antar manusia, ekspresi emosi, termasuk ekspresi balasan yang ditampilkan terhadap stimulus tertentu, dipengaruhi pola komunikasi dan interaksi sosial yang melibatkan emosi. Ekspresi emosi ada verbal dan non verbal. Satu individu dapat menyebabkan orang lain bereaksi terhadap ekspresi tersebut dan pesan yang terkandung di dalamnya. Di samping itu, ekspresi wajah memunculkan respon yang dapat mempengaruhi banyak hal yang tidak berhubungan dengan penilaian, preferensi, dan sikap orang lain.

Pesan melalui ekspresi emosi tidak hanya ditunjukkan secara langsung oleh orang lain, tetapi juga dapat ditangkap melalui media. Sering kali individu tidak menyadari adanya pengaruh media secara tidak langsung terhadap suasana hati. Media dapat digunakan untuk berbagai tujuan dan setiap media memiliki ciri khasnya untuk tujuan tertentu. Pesan ekspresi emosi melalui media *audio* visual dapat menjadi salah satu sarana mengetahui kemampuan individu untuk mengenali ekspresi emosi. Walaupun manusia bisa mengenali ekspresi emosional dengan baik, penelitian pengenalan ekspresi emosional yang dilakukan oleh mesin terus dilakukan agar dapat melakukan pengenalan ekspresi emosional dalam interaksi manusia dan komputer. Oleh karena itu, diperlukan suatu pendekatan untuk mengenali ekspresi emosional dengan lebih baik.

Untuk memanfaatkan sepenuhnya perbedaan saturasi emosional antara kerangka waktu, metode baru diusulkan untuk pengenalan ucapan menggunakan ekstraksi fitur yang dimanfaatkan oleh algoritma *Long Short-Term Memory*

(LSTM) dan *Convolutional Neural Network* (CNN). Kemudian dilakukan perbandingan evaluasi kinerja antara kedua logaritma tersebut. Pengenalan ucapan emosional bertujuan untuk secara otomatis mengklasifikasikan unit ucapan (mis., Ucapan) ke dalam keadaan emosional, seperti marah, bahagia, netral, sedih, muak, dan terkejut.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengimplementasikan model arsitektur LSTM, VGG16, VGG19 dan InceptionV3 ini pada dataset TESS?
2. Bagaimana hasil evaluasi kinerja model arsitektur LSTM, VGG16, VGG19 dan InceptionV3 untuk klasifikasi emosi suara pada dataset TESS?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan adalah *Toronto Emotional Speech (*TESS*).*
2. Dataset yang digunakan hanya menggunakan suara wanita muda dan wanita tua, berumur sekitar 25-65 tahun.
3. Jenis ekspresi wajah yang dijadikan objek penelitian adalah ekspresi senang, sedih, terkejut, jijik, takut, marah, dan emosi netral.
4. Dataset hanya menggunakan suara bahasa Inggris.
5. Arsitektur yang digunakan yaitu LSTM, VGG16, VGG19 dan InceptionV3.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengimplementasikan model arsitektur LSTM, VGG16, VGG19 dan InceptionV3 ini pada dataset TESS.
2. Untuk mengetahui hasil evaluasi kinerja model arsitektur LSTM, VGG16, VGG19 dan InceptionV3 untuk klasifikasi emosi suara pada dataset TESS.

1.5 Manfaat Penelitian

Penelitian ini dapat menghasilkan model Machine Learning yang nantinya dapat digunakan untuk klasifikasi emosi suara, serta memberikan informasi tentang implementasi model arsitektur LSTM, VGG16, VGG19 dan InceptionV3 pada dataset TESS.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian ini merujuk pada beberapa penelitian yang telah dilakukan sebelumnya. Berikut ini merupakan penelitian terkait:

Penelitian yang dilakukan oleh Ververidis, Dimitrios dan Constantine Kotropoulos pada tahun 2015. Pada penelitian ini berfokus pada klasifikasi ucapan emosional berdasarkan informasi gender menggunakan metode Sequential Forward Selection. Ketika pengklasifikasi Bayes dengan PDF Gaussian digunakan, tingkat klasifikasi yang benar sebesar 61,1% diperoleh untuk subjek laki-laki dan tingkat yang sesuai sebesar 57,1% untuk subjek perempuan. Dalam eksperimen yang sama, pengenalan ucapan Emosional acak bertujuan untuk secara otomatis mengklasifikasikan unit ucapan (misalnya ucapan) ke dalam keadaan emosional, seperti kemarahan, kebahagiaan, netral, kesedihan, dan kejutan.

Penelitian yang dilakukan oleh Tamulevičius G., Karbauskaitė R. dan Dzemyda G. pada tahun 2019, melakukan penelitian klasifikasi ucapan emosional menggunakan fitur berbasis dimensi fraktal. Hasil penelitian menunjukkan akurasi rata-ratanya adalah 96,5 %.

Penelitian yang dilakukan oleh Yi-Lim Lin, dan Gang Wei pada tahun 2005, melakukan penelitian pengenalan ucapan emosional menggunakan Hidden Markov Model (HMM) dan Support Vector Machine (SVM). Dengan cara mengklasifikasikan keadaan emosi seseorang berdasarkan gender. Dari hasil penelitian tersebut, tingkat pengenalan oleh pengklasifikasi HMM adalah 98,9% untuk subjek perempuan, 100% untuk subjek laki-laki, dan 99,5% untuk kasus independen gender. Ketika klasifikasi SVM dan vektor fitur yang diusulkan digunakan, tingkat klasifikasi yang benar dari 89,4%, 93,6% dan 88,9% diperoleh masing-masing untuk kasus independen gender laki-laki dan perempuan.

2.2 Ekspresi Emosi

Ekspresi emosi manusia tidaklah bersifat unik tetapi dapat pula ditemukan pada banyak jenis yaitu binatang (Ekman, 2003). Banyak dari peristiwa sosial dialami oleh manusia menghasilkan emosi yang sama juga dialami oleh binatang. Pendapat Darwin ini merupakan hasil dari eksperimen berkesinambungan yang dilakukan merujuk pada teori evolusionernya. Sebagai salah satu ilmuwan yang pertama kali menggunakan foto sebagai ilustrasi dan menggunakan metode judgement untuk mempelajari nilai isyarat dari suatu ekspresi yang sekarang menjadi ekspresi paling sering dibahas dengan menggunakan metode psikologis.

Pada prinsipnya guratan ekspresi emosi adalah tindakan yang bersifat tingkah laku lengkap, dan kombinasi dengan tanggapan jasmani lain yaitu suara, postur, gestur, pergerakan otot, dan tanggapan fisiologis lainnya (Matsumoto dan Ekman, 2007). Misalnya guratan ekspresi emosi yang ditunjukan oleh raut wajah seseorang adalah bagian dari emosi. Guratan ekspresi merupakan bentuk komunikasi seperti kata-kata dan merupakan bentuk komunikasi yang lebih cepat dari kata-kata itu sendiri (Safaria dan Saputra, 2009).

Ekspresi emosi muncul secara spontan bahkan seringkali sulit dikontrol atau disembunyikan (Hude, 2006). Ekspresi emosi dapat terlihat dari perubahan fisiologis yang timbul akibat reaksi terhadap peristiwa atau stimulus tertentu yang mengakibatkan emosi, reaksi ini baik bersifat internal maupun eksternal akan memunculkan ekspresi emosi yang terwujud dalam penampilan fisiologis, meliputi raut wajah, hingga sikap dan tingkah laku. Ekspresi emosi selain diwarisi secara genetis ternyata dipengaruhi juga oleh pengalaman dalam berinteraksi dengan orang lain.

Emotional expression (ekspresi emosi) merupakan perubahan-perubahan dalam otot, kelenjar yang mendalam dan tingkah laku, yang berasosiasi dengan emosi (Chaplin, 2006). Ekspresi emosi adalah kecenderungan seseorang untuk mengungkapkan perasaan yang sedang dirasakan kepada orang lain. Ekspresi emosi adalah suatu upaya yang dilakukan untuk mengkomunikasikan status perasaannya yang berorientasi pada tujuan tertentu (Safaria dan Saputra, 2009).

Ekspresi emosi adalah keadaan kesiapan menanggapi peristiwa-peristiwa mendesak untuk bereaksi atau bertindak dan bagaimana merespon emosi (Ekman, 1997). Ekspresi emosi sebagai suatu perasaan dan pikiran-pikiran khasnya, suatu keadaan biologis dan psikologis, dan serangkaian kecenderungan untuk siap bertindak (Goleman, 2004).

Ekspresi emosi adalah suatu bentuk komunikasi melalui perubahan raut wajah dan gesture yang menyertai emosi, sebagai luapan dari emosi, mengungkapkan, menyampaikan perasaan kepada orang lain, dan menentukan bagaimana perasaan orang lain (Safaria dan Saputra, 2009). Pengekspresian emosional seseorang akan memberikan informasi yang diperlukan oleh individu untuk mengambil suatu keputusan yang dapat dilakukan melalui komunikasi (Zuhana, 2010).

Ekspresi emosi yang tinggi merupakan refleksi sikap negatif berperan sebagai stressor yang dapat meningkatkan kerentanan dan kekambuhan pada seseorang yang mengalami gangguan psikologis (Hertinjung dan Partini, 2010). Sementara itu, Hasanat berpendapat bahwa ekspresi emosi merupakan indeks keseluruhan emosi, sikap dan perilaku yang diekspresikan dalam keluarga, ekspresi emosi juga berkaitan dengan bagaimana cara orangtua atau pasangan berbicara mengenai individu atau seseorang yang mengalami gangguan psikologis. (Hertinjung dan Partini, 2010)

Menurut Barrett dan Fossum emosi adalah manifestasi dari keadaan fisiologis dan kognitif manusia, yang dalam pengungkapannya merupakan cermin dari pengaruh budaya dan sistem sosial. Memperkuat pendapat tersebut, Berry menambahkan bahwa emosi dipelajari individu sebagai nilai-nilai budaya dalam lingkungan sosial yang ditinggali (Kurniawan Hasanat, 2007). Maka kultur dan sistem sosial dimana individu tersebut tinggal dan menetap mengatur serta membatasi kepada siapa, kapan, dan dimana seseorang bisa mengungkapkan dan merahasiakan emosi-emosi yang sedang ia rasakan, serta berhubungan dengan cara pengungkapan emosi tersebut baik verbal maupun nonverbal.

Berdasarkan uraian mengenai definisi ekspresi emosi sebelumnya dapat disimpulkan bahwa ekspresi emosi merupakan usaha yang dilakukan oleh

seseorang untuk mengkomunikasikan status perasaan (emosi) sebagai respon terhadap situasi tertentu baik internal maupun eksternal yang terlihat dari perubahan biologis, fisiologis dan serangkaian kecenderungan tindakan (sikap dan tingkah laku) berorientasi pada tujuan.

2.2.1 Ekspresi Suara

Ekspresi suara yang pada umumnya dikenali adalah tertawa, bersenandung, berteriak-teriak, memaki, atau tiba-tiba terhenyak dengan tatapan kosong. Menandai makna ekspresi suara tidak semudah dengan ekspresi wajah. Orang yang berteriak-teriak tidaklah selalu menandakan bahwa ia sedang marah bahkan ada orang yang marah hanya diam saja, sebaliknya orang yang diam tidak berarti dia sedang dalam keadaan sedih. Para pakar komunikasi menganggap bahwa komunikasi dalam bentuk ekspresi suara lebih mudah dipahami dan lebih berpengaruh daripada komunikasi tertulis.

2.2.2 Jenis-jenis Emosi *Universal*

Ekman menyatakan dalam penelitiannya berdasarkan pada sistematika dan konklusif bahwa secara universal ekspresi emosi (Matsumoto, 2005; Matsumoto dan Ekman, 2007). Meliputi :

Tabel 2.1 Jenis-Jenis Ekspresi Emosi dan Pengertiannya

Jenis Ekspresi Emosi	Pengertian
Anger (Marah)	Perasaan ketidaksenangan terhadap sesuatu yang melukai, menganiaya, menentang dan biasanya muncul dengan spontan serta ingin melawan penyebab perasaan ini. Ekspresi emosi marah sangat bervariasi bentuknya mulai dari perubahan raut muka, dalam bentuk verbal, dalam bentuk tindakan, hingga dalam bentuk sikap dan marah yang tidak diperlihatkan. Pelampiasan marah dapat ditahan atau bahkan dapat pula dieksplorasi. Secara psikologis terlalu sering menahan marah akan menimbulkan kegoncangan mental dan hal tersebut tidak baik untuk kesehatan mental.
Contempt (Muak)	Perasaan atau perilaku ketika seseorang melihat sesuatu atau seseorang yang kualitas tindakan, proses atau kemampuannya menurun atau rendah, rata-rata

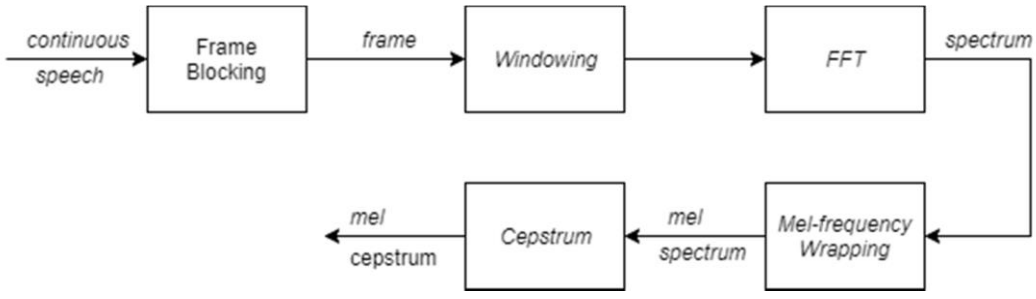
	atau biasa saja, atau tidak layak.
Disgust (Jijik)	Perasaan yang muncul karena suatu objek yang menjijikan, tidak disukai, atau dibenci.
*Fear (*Takut*)*	Perasaan cemas dan menghasut karena adanya kehadiran sesuatu yang berbahaya, kejahatan, atau perasaan yang akan menyakitkan. Rasa takut mendorong manusia untuk mengambil tindakan yang perlu untuk menghindari bahaya yang mengancam kelangsungan hidup. Ekspresi emosi takut dapat berupa tindakan seperti: berteriak histeris (scream), loncat, berlari, merunduk, menutup telinga, atau menghindar. Ekspresi takut juga ditandai dengan perubahan faali seperti: denyut nadi meningkat, jantung berdebar-debar, pandangan mata kabur, keluar keringat dingin, dan persendian terasa lemas.
*Happiness (*Senang*)*	Perasaan terhadap sesuatu yang benar-benar disukai, kepuasan, atau rasa riang gembira. Emosi gembira dan bahagia dalam psikologi ditekankan pada hal yang membawa kebermaknaan pada kehidupan. Seseorang akan mencapai kebahagiaannya ketika kebutuhan-kebutuhannya telah terpenuhi atau tercukupi. Sehingga kesehatan mental akan bermula dan berkembang.
*Sadness (*Sedih*)*	Perasaan dimana semangat yang rendah atau duka cita. Beberapa hal yang biasanya menyebabkan manusia dirundung kesedihan yaitu ketika musibah datang seperti kegagalan, kecelakaan, kematian, dan lain-lain. Emosi sedih dapat terjadi dalam hubungan interpersonal, misalnya pada proses komunikasi pesan yang disampaikan dipahami tidak sesuai dengan harapan sebenarnya (misunderstanding), hal ini dapat menimbulkan kekecewaan. Ekspresi emosi sedih meliputi: menangis dengan air mata bercucuran, mata berkaca-kaca, wajah pucat, dingin, pandangan lesu, tanpa senyum, dan tidak bergairah.
*Surprise (*Terkejut*)*	Perasaan atas sesuatu yang tiba-tiba atau tidak terduga. Emosi heran dan kaget berada pada kontinum yang sama. Biasanya diekspresikan dengan: berteriak spontan, terperanjat, mata terbelalak, merinding, latah, meneteskan air mata, dan tertawa.

Sumber: Matsumoto dan Ekman, 2007; Ekman dan Friesen, 2009; Matsumoto, 2005; Fok, dkk., 2007.

2.3 *Mel-Frequency Cepstral Coefficient* (MFCC)

Mel-Frequency Cepstral Coefficient (MFCC) merupakan ekstraksi ciri yang menghitung koefisien cepstral dengan mempertimbangkan pendengaran manusia (Buono, 2009). MFCC II-6 didasarkan atas variasi *bandwidth* kritis terhadap frekuensi pada telinga manusia yang merupakan filter yang bekerja secara linier pada frekuensi rendah dan bekerja secara logaritmik pada frekuensi tinggi. Filter ini digunakan untuk menangkap karakteristik fonetis penting dari sinyal ucapan. Untuk meniru kondisi telinga, karakteristik ini digambarkan dalam skala mel-frekuensi, yang merupakan frekuensi linier di bawah 1000 Hz dan frekuensi logaritmik di atas 1000 Hz.

Oleh karena itu, menurut waktu singkat analisis spektral merupakan cara yang paling umum dan tepat untuk mengekstraksi ciri suara masukan (Do, 2013). Ekstraksi ciri sinyal suara menggunakan MFCC didasarkan atas variasi *bandwidth* terhadap frekuensi pada telinga manusia yang merupakan filter, yang bekerja secara linier pada frekuensi rendah dan bekerja secara logaritmik pada frekuensi tinggi. Filter ini digunakan untuk menangkap karakteristik fonetis penting dari sinyal suara masukan atau ucapan. Karakteristik ini digambarkan dalam skala mel-frekuensi, yang merupakan frekuensi linier di bawah 1000 Hz dan frekuensi logaritmik di atas 1000 Hz. Diagram proses MFCC dapat dilihat pada Gambar 2.1.



Gambar 2.1 Blok MFCC (Kirti dan Minakshee, 2013)

1. Frame Blocking

Proses ini akan membagi sinyal *audio* kedalam bentuk *frame*. *Frame blocking* merupakan proses pembagian sinyal menjadi beberapa *frame* yang lebih kecil agar sinyal lebih mudah untuk diproses selanjutnya (Syafria, 2014). *Frame blocking* dilakukan dengan cara manual yaitu dengan memotong masing-masing suara menjadi satu huruf. *Frame blocking* dilakukan dengan cara manual menggunakan *software audacity*.

2. Windowing

Windowing adalah proses yang dilakukan untuk meminimalisir diskontinuitas antar frame yang dapat menyebabkan kehilangan informasi yang terdapat pada suatu sinyal (Setiawan, 2011). Konsepnya adalah meruncingkan sinyal ke angka nol pada permulaan dan akhir setiap frame.

3. *Fast Fourier Transform* (FFT)

Fast Fourier Transform (FFT). Tahapan selanjutnya adalah mengubah tiap frame dari domain waktu ke dalam domain frekuensi. FFT adalah algoritma yang mengimplementasikan *Discrete Fouries Transform* (DFT).

4. Mel-Frequency Wrapping

Studi psikofisik telah menunjukkan bahwa persepsi manusia tentang frekuensi suara untuk sinyal ucapan tidak mengikuti skala linier. Jadi, untuk setiap nada dengan frekuensi sesungguhnya f, dalam Hz, sebuah pola diukur dalam sebuah skala yang disebut “mel”. Skala “mel frekuensi” adalah skala frekuensi linier di bawah 1000 Hz dan skala logaritmik di atas 1000 Hz.

5. Cepstrum

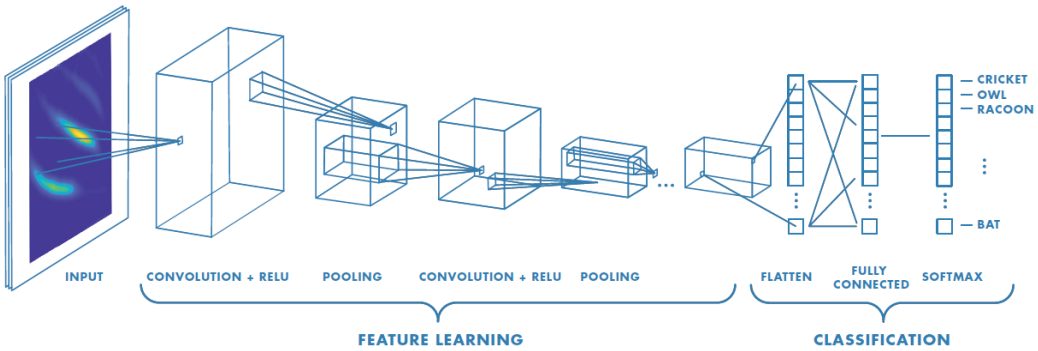
Cepstrum adalah sebutan kebalikan untuk spectrum. Cepstrum biasa digunakan untuk mendapatkan informasi dari suatu sinyal suara yang diucapkan oleh manusia. Pada langkah terakhir ini, spektrum log mel dikonversi menjadi cepstrum menggunakan Discrete Cosine Transform (DCT). Hasil dari proses

ini dinamakan MFCC. MFCC ini adalah hasil alihragam cosinus dari logaritma short term power spectrum yang dinyatakan dalam skala mel frekuensi.

Pada penelitian ini, untuk mengetahui perbedaan suara emosi ketika tertawa, marah, menangis, dan lain-lain. Emosi itu dibedakan melalui suara wanita muda dan tua menggunakan model arsitektur CNN dan LSTM. Sebuah metode yang akan mendeteksi perbedaan suara emosi dengan menghitung nilai mean dari nilai ekstraksi Mel-Frequency Cepstral Coefficient (MFCC).

2.4 *Convolutional Neural Network*

Convolutional Neural Network (CNN), juga disebut *ConvNet*, adalah jenis *Artificial Neural Network* (ANN) atau dalam bahasa indonesia disebut Jaringan Syaraf Tiruan (JST). *ANN* memiliki arsitektur umpan-maju yang dalam dan memiliki kemampuan generalisasi yang luar biasa dibandingkan dengan jaringan lain dengan lapisan *fully connected*, CNN dapat mempelajari fitur-fitur dari objek terutama data spasial dan dapat mengidentifikasi objek dengan lebih efisien. Model *deep* CNN umumnya terdiri dari dua tahap, yang pertama adalah *feature extraction* atau *feature learning* dan kedua adalah *classification* atau klasifikasi dengan *fully connected layer*. Model konseptual dasar CNN ditunjukkan pada Gambar 2.2 (Ghosh dkk., 2019).



Gambar 2.2 Arsitektur *Convolutional Neural Network*

Perbedaan antara CNN dan JST/ANN tradisional adalah CNN digunakan terutama di bidang pengenalan pola dalam citra sehingga dapat mengurangi jumlah parameter saat membuat model (O'Shea dan Nash, 2015). Asumsi terpenting

tentang masalah yang diselesaikan oleh CNN seharusnya tidak memiliki fitur yang bergantung secara spasial. Dengan kata lain, misalnya dalam sebuah aplikasi deteksi wajah, tidak perlu memperhatikan letak wajah di dalam citra (Albawi dkk., 2017). Berikut ini dijelaskan tahapan CNN.

2.4.1 *Convolutional Layer*

Konvolusi merupakan salah satu tahap pada arsitektur CNN. Konvolusi merupakan suatu istilah matematis yang berarti mengaplikasikan sebuah fungsi pada output fungsi lain secara berulang. Dalam pengolahan citra, konvolusi berarti mengaplikasikan sebuah kernel pada citra. Kernel adalah sebuah matriks kecil dengan tinggi dan lebarnya lebih kecil dari matriks citra yang akan di konvolusi. Kernel biasanya juga dikenal dengan istilah filter atau convolution mask (Pokhrel, 2019).

Dalam *machine learning*, *input* citra berbentuk *array* dua dimensi dan *kernel* merupakan parameter berbentuk *array* multidimensi yang disesuaikan dengan model algoritma. Konvolusi dapat digunakan pada lebih dari satu dimensi. Sebagai contoh jika menggunakan gambar dua dimensi I sebagai *input*, maka *kernel K* juga berbentuk dua dimensi:

$$S(i, j) = (I * K)(i, j) = \sum_a \sum_b I(a, b)K(i - a, j - b) \quad (2.1)$$

Keterangan:

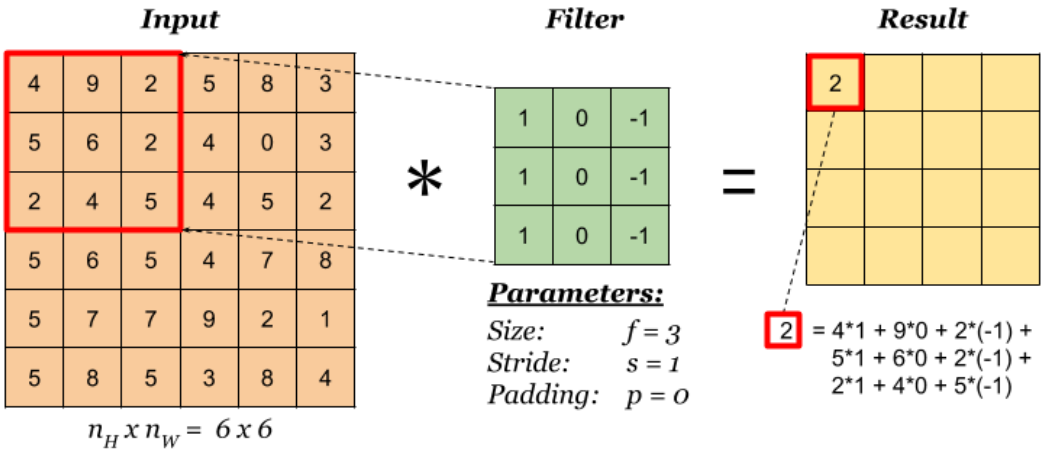
S (i, j) = Fungsi hasil konvolusi

I = *Input*

K = *Kernel* atau *Filter*

(i, j) = *Pixel Input*

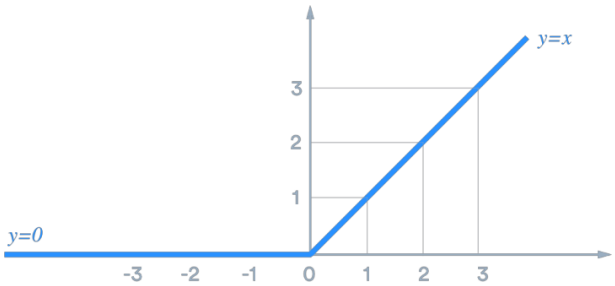
(a, b) = *Pixel Kernel*



Gambar 2.3 Contoh Bentuk Konvolusi

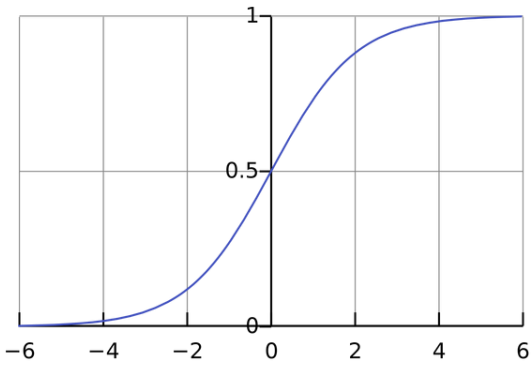
2.4.2 Fungsi Aktivasi

Operasi konvolusi adalah linier dan untuk membuat jaringan syaraf lebih kuat, perlu memperkenalkan beberapa non-linier. Untuk tujuan ini, dapat menerapkan fungsi aktivasi seperti ReLU (Bahuleyan, 2018). *Rectified linear unit* (ReLU) merupakan salah satu fungsi aktivasi yang sering digunakan pada *convolutional neural network* (Chen dkk., 2018). Fungsi aktivasi ReLU mempercepat waktu komputasi karena sangat sederhana. Jika nilai input negatif, maka *outputnya* adalah 0. Jika positif maka *outputnya* adalah nilainya sendiri.



Gambar 2.4 Fungsi aktivasi ReLU

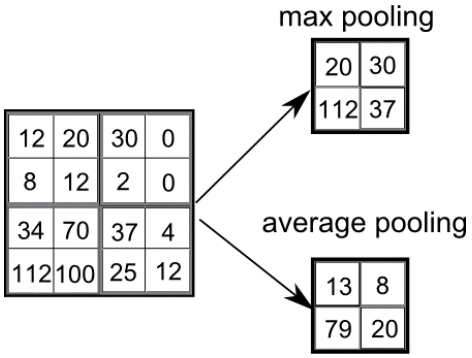
Fungsi aktivasi ReLU digunakan pada proses ekstraksi fitur,tetapi pada proses klasifikasi fungsi aktivasi yang digunakan adalah *softmax*. Fungsi aktivasi *softmax* berada pada *layer output* yang bertujuan untuk klasifikasi. Biasanya fungsi aktivasi *softmax* digunakan pada multikelas.



Gambar 2.5 Fungsi aktivasi *softmax*

2.4.3 *Pooling Layer*

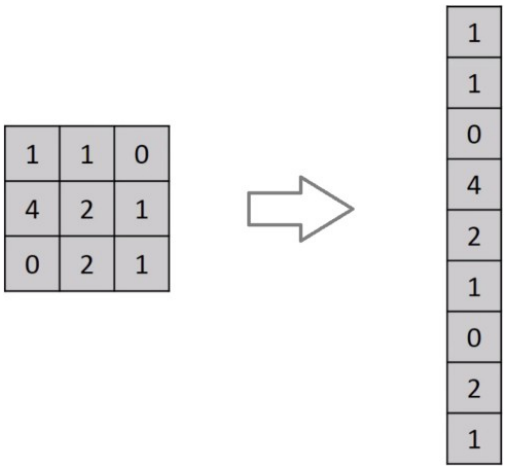
Pooling layer berfungsi menjaga ukuran data setelah proses convolution dengan melakukan *downsampling* (pereduksian sample), yaitu mengambil *feature map* dengan ukuran lebih besar dan menyusutkannya ke *feature map* berukuran lebih rendah. Dengan *pooling* data direpresentasikan ke dalam bentuk yang lebih kecil, sehingga dapat mereduksi waktu komputasi dan mengatasi *overfitting* (Suyanto, 2018). Ada berbagai jenis teknik yang digunakan pada *pooling layers* seperti *max pooling*, *min pooling*, *average pooling*, *gated pooling*, *tree pooling*, dan lain-lain. *Max pooling* adalah teknik penyatuan yang paling populer dan paling banyak digunakan. *Max pooling* bekerja dengan mempartisikan citra ke sub-wilayah persegi, dan hanya mengembalikan nilai maksimum dari di dalam sub-wilayah itu seperti pada Gambar 2.6. Salah satu ukuran filter paling umum yang digunakan dalam max-pooling adalah 2x2 (Gosh et al, 2020).



Gambar 2.6 Ilustrasi *Pooling Layer*

2.4.4 *Flatten*

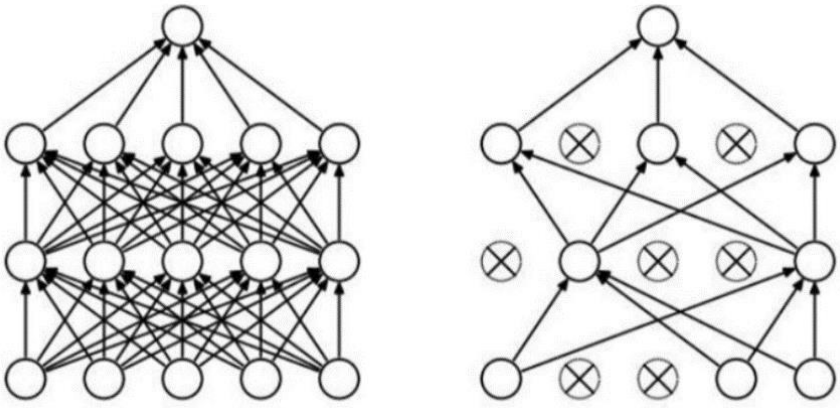
Flatten layer merupakan sebuah metode untuk mengubah hasil dari pooling layer. Flatten mengubah multi-dimensi menjadi satu dimensi, yang *outputnya* akan digunakan oleh *Fully connected layer*.



Gambar 2.7 Ilustrasi *Flatten layer*

2.4.5 *Dropout*

Dropout adalah salah satu pendekatan yang paling banyak digunakan untuk regularisasi. Metode ini secara acak melakukan drop *neuron* pada jaringan selama proses *training*. Dengan melakukan drop pada beberapa *neuron*, kemampuan dari *feature selection* (seleksi fitur) dapat didistribusikan ke semua *neuron* secara merata dan secara langsung memaksa model untuk mempelajari beberapa fitur independen. Dengan melakukan drop pada *neuron* berarti, *neuron* 11 yang di drop tidak akan mengambil bagian dalam propagasi maju atau propagasi mundur selama proses *training*. Tetapi dalam kasus proses pengujian, jaringan skala penuh digunakan untuk melakukan prediksi. Pada Gambar 2.8 ditunjukkan gambaran cara kerja *dropout* pada arsitektur CNN saat proses *training* berlangsung (Gosh dkk, 2020). *Dropout* merupakan salah satu usaha untuk *mencegah* terjadinya *overfitting* dan juga mempercepat proses learning (Santoso dan Ariyanto, 2018).

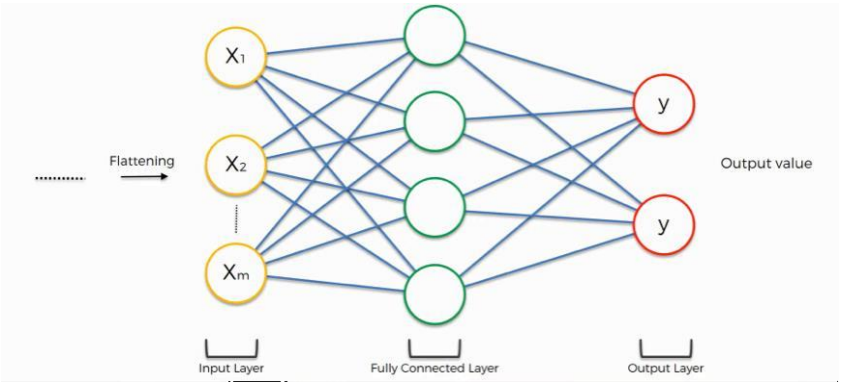


(a) (b)

Gambar 2.8 Sebelum *Dropout* (a) dan Setelah *Dropout* (b)

2.4.6 *Fully Connected Layer*

Fully connected layer mengambil *input* dari hasil *output pooling layer* yang berupa feature map. *Feature map* tersebut masih berbentuk multidimensional *array* maka lapisan ini akan melakukan *reshape feature map* dan menghasilkan vektor sebanyak n-dimensi dimana n adalah jumlah kelas *output* yang harus dipilih program. Misalnya lapisan terdiri dari 500 neuron, maka akan diterapkan sebagai klasifikasi akhir dari jaringan (Dutt dan Dutt, 2017). Gambar 2.9 menampilkan proses yang ada dalam *fully connected layer*.



Gambar 2.9. Proses *fully connected layer*

2.5 Arsitektur *Convolutional Neural Network*

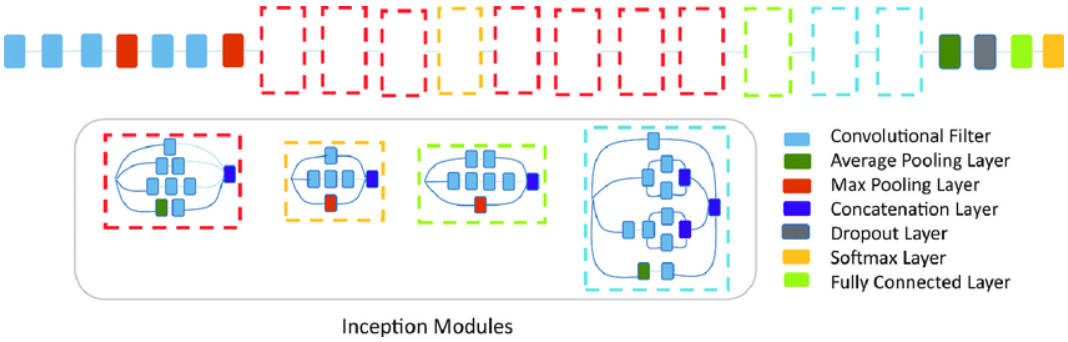
Arsitektur CNN merupakan gabungan dari beberapa layer yang secara umum terdiri dari *Convolutional Layer, Subsampling Layer (Pooling Layer)*, dan

Fully Connected Layer (O'Shea dan Nash, 2015). *Layer* ini tersusun tanpa adanya aturan yang universal dan berbeda-beda tergantung dari dataset yang digunakan. Ada banyak arsitektur CNN yang telah didesain para ahli, yang pertama adalah LeNet (LeCun dkk., 1998) yang digunakan untuk membaca kode pos dan digit (Suyanto, 2018), namun karena keterbatasan perangkat dan waktu komputasi yang tinggi maka arsitektur ini sempat dilupakan karena dianggap tidak efektif saat itu.

Selanjutnya arsitektur CNN terus berkembang dan pada tahun 2012 muncullah AlexNet yang membuat arsitektur CNN kembali dilirik untuk aplikasi *Computer Vision* (Suyanto, 2018). Dari beberapa arsitektur yang didesain beberapa mengikuti kompetisi *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) yang menggunakan dataset *ImageNet* untuk melihat kinerja dari setiap arsitektur. Arsitektur dengan performa terbaik pada ILSVRC dapat digunakan untuk menjadi *base* model atau *pre-trained network* karena akurasi yang tinggi dan *validation loss* yang rendah. Beberapa arsitektur yang digunakan pada penelitian ini antara lain Inception V3,VGG19, dan VGG16.

2.5.1 *InceptionV3*

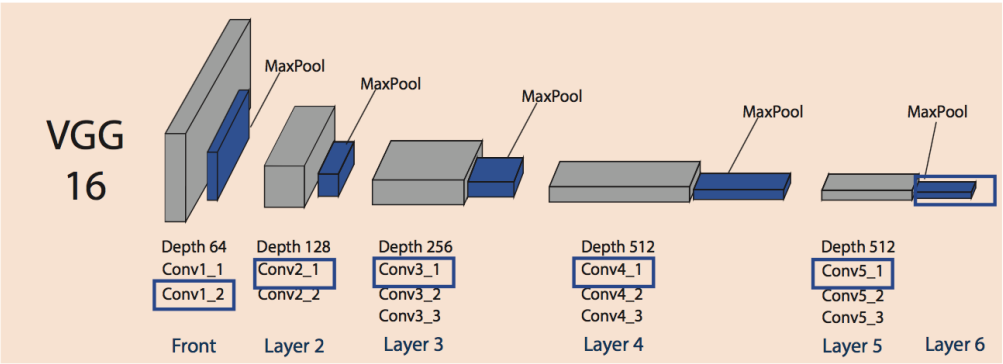
InceptionV3 adalah model pengenalan citra yang banyak digunakan yang memiliki ukuran *input* citra 299 x 299. Model yang dihasilkan arsitektur ini telah terbukti mencapai akurasi lebih dari 78,1% pada dataset *ImageNet*. Arsitektur ini adalah arsitektur jaringan neural konvolusional dari keluarga *Inception* yang membuat beberapa peningkatan dari versi sebelumnya termasuk menggunakan label *Smoothing*, konvolusi Difaktorkan 7 x 7, dan penggunaan *auxiliary classifier* untuk menyebarkan informasi label ke bagian bawah jaringan (bersama dengan penggunaan *batch normalization*) (Szegedy dkk., 2016).



Gambar 2.10 Arsitektur InceptionV3

2.5.2 VGG16

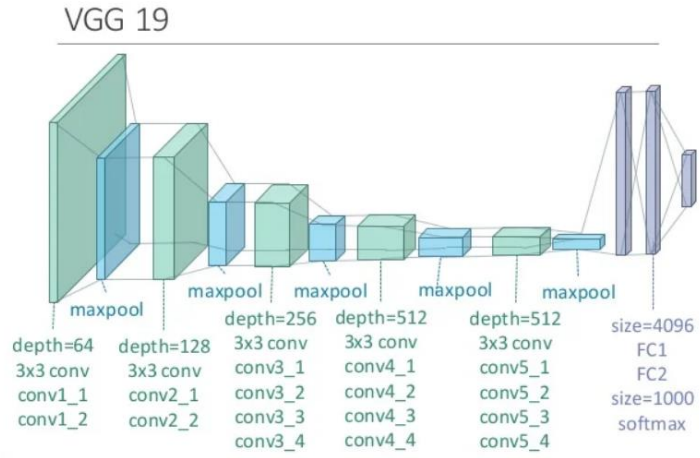
VGGNet merupakan arsitektur yang menunjukkan bahwa kedalaman jaringan merupakan komponen penting untuk menghasilkan performa yang tinggi. Salah satu jaringan VGGNet yang yang digunakan pada penelitian ini adalah *VGG16* yang mana terdapat 13 jaringan konvolusional dan 3 jaringan *fully connected*. VGGNet menggunakan filter konvolusi 3x3 dan *max-pooling* 2x2 diikuti oleh tiga lapisan Tersambung Penuh (*fully connected*), dua yang 13 pertama memiliki masing-masing 4096 saluran, yang ketiga berfungsi Klasifikasi ILSVRC 1000 arah dan karenanya berisi 1000 saluran (satu untuk setiap kelas) (Simonyan dan Zisserman, 2015). VGGNet memiliki 138-140 juta parameter sehingga membutuhkan banyak memori.



Gambar 2.11 Arsitektur Model VGG16

2.5.3 VGG19

VGG19 adalah arsitektur jaringan yang merupakan bentuk variasi dari VGG16 yang diciptakan oleh *Visual Geometry Group* (VGG). Model ini memiliki 16 lapisan konvolusi dengan filter 3x3, kemudian terdapat 5 *pooling layer* yang menggunakan *Max Pooling* dengan pool size sebesar 2 x 2 dan 3 fully connected layer dengan nilai 4096, 4096, 1000 lalu size image untuk input memiliki ukuran 224 x 224 piksel.



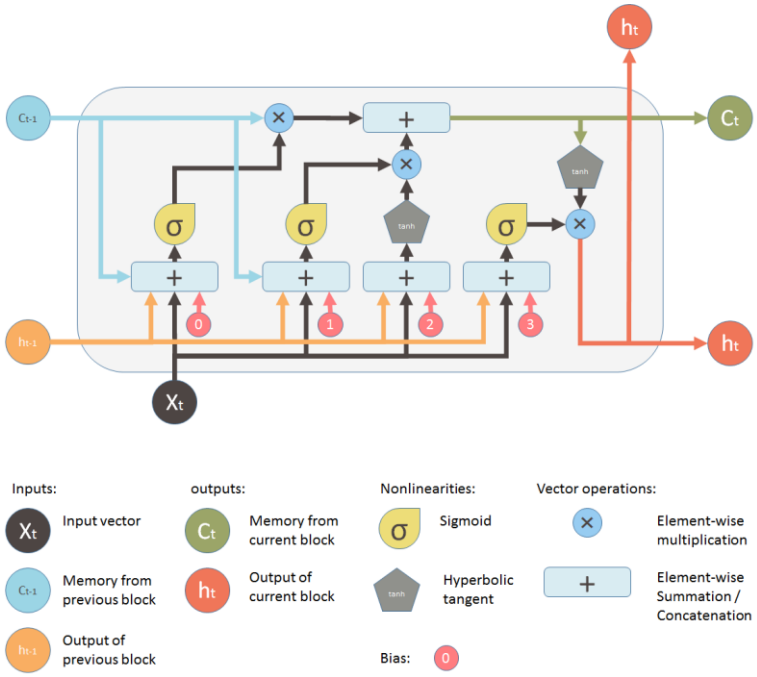
Gambar 2.12 Arsitektur Model VGG19

2.6 *Long Short Term Memory* (LSTM)

Long Short Term Memory (LSTM) merupakan pengembangan dari *Recurrent Neural Network* (RNN) dengan mengatasi salah satu kekurangan RNN yaitu kemampuan pengelolaan informasi dalam periode yang lama. Diusulkan oleh Sepp Hochreiter dan Jurgen Schmidhuber pada tahun 1997, LSTM banyak dipilih untuk prediksi berbasis waktu atau *time-series* karena dikenal lebih unggul dan handal dalam melakukan prediksi dalam waktu lama dibanding algoritma lain.

LSTM mempunyai sel memori dan arsitektur, dalam LSTM itu terdiri dari *input gate*, koneksi berulang, *forget gate*, dan *output gate*. LSTM juga mampu mengingat informasi jangka panjang. Pada *input gate* berfungsi untuk memblokir atau memasukan bagian yang akan diperbaharui. *Output gate* adalah hasil dari lapisan *sigmoid* yang dijalankan untuk menentukan sel mana yang akan menjadi

output nya. *Forget gate* merupakan memori-memori masa lalu untuk melupakan masa lalu (Le dkk., 2019) Berikut merupakan gambar dari arsitektur LSTM yang dapat dilihat pada Gambar 2.12.



Gambar 2.13 Arsitektur Model *LSTM*

Ide dasar dari *LSTM* yaitu adanya jalur yang menghubungkan antara cell state (Ct-1) sebelumnya dengan *cell state* yang sekarang (Ct). Jalur tersebut, merupakan informasi pada *cell state* dengan mudah dapat diteruskan ke *cell state* berikutnya dengan beberapa modifikasi yang diperlukan. Nilai *cell state* adalah vektor yang dirancang untuk menyimpan informasi tentang konteks sekuen data. Contoh dalam pemrosesan kalimat, informasi yang dapat akan disimpan di dalam *cell state* adalah gender dari subyek, apakah subjek tunggal atau jamak dan sebagainya.

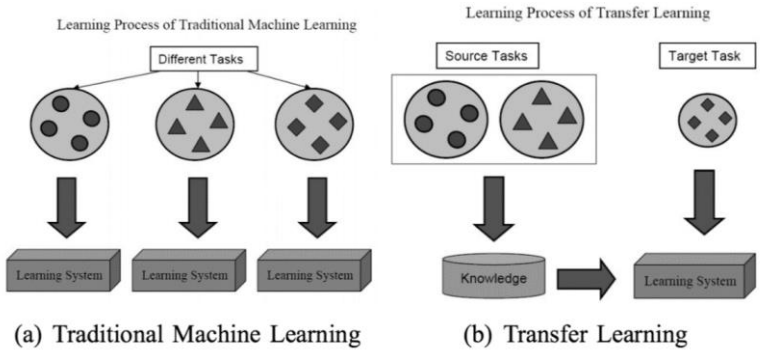
Fitur-fitur ini akan diekstrak oleh *LSTM* selama proses *training*. Ide berikutnya dari *LSTM* adalah penggunaan gerbang *sigmoid* (disimbolkan dengan σ) yang akan mengatur informasi apakah diteruskan atau dihentikan. Seperti diketahui bahwa *output* dari fungsi *sigmoid* adalah antara nol dan satu dengan arti nol adalah informasi dihentikan seluruhnya dan satu adalah informasi diteruskan

seluruhnya. Keluaran dari fungsi *sigmoid* dikalikan dengan suatu nilai lain untuk menentukan seberapa besar informasi tersebut akan digunakan untuk proses berikutnya.

2.7 *Transfer Learning*

Transfer Learning adalah metode menggunakan atau memanfaatkan kembali model atau pengetahuan yang telah dipelajari sebelumnya untuk peningkatan pembelajaran untuk tugas yang lebih baru dan berkaitan (Bali dan Ghosh, 2018). Selain memberikan kemampuan untuk menggunakan kembali model yang sudah dibangun, *transfer learning* dapat membantu mempelajari tugas target dengan cara berikut:

1. Peningkatan kinerja dasar: dengan menambah pengetahuan dari isolated learner (juga dikenal *ignored learner*) dengan pengetahuan dari model sumber atau model yang telah dilatih, kinerja dasar mungkin akan meningkat dengan menerapkan *transfer learning*.
2. Waktu pengembangan model: Memanfaatkan pengetahuan dari model sumber mungkin juga membantu dalam mempelajari tugas target secara penuh, dibandingkan dengan model target yang belajar dari awal. Hal ini menghasilkan peningkatan dalam keseluruhan waktu yang dibutuhkan untuk mengembangkan / mempelajari model.
3. Peningkatan kinerja akhir: Kinerja akhir yang lebih tinggi dapat dicapai dengan memanfaatkan *transfer learning*.



Gambar 2.14 Ilustrasi proses *learning* yang berbeda antara (a) *machine learning* tradisional dan (b) *transfer learning*.

2.8 Evaluasi Kinerja Model

Evaluasi kinerja model merupakan yang terpenting dalam menentukan suatu model bagus atau tidak. Pada kasus klasifikasi, evaluasi kinerja yang digunakan berupa *Confusion matrix*, *precision*, *recall*, *accuracy* dan *F1-score*. Terdapat beberapa pengukuran kinerja model yang digunakan dalam penelitian ini.

1. Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining* atau Sistem Pendukung Keputusan. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 istilah sebagai representasi hasil proses klasifikasi diantaranya *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

1. *True Negative* (TN) merupakan jumlah data negatif yang terdeteksi dengan benar.
2. *False Positive* (FP) merupakan data negatif namun terdeteksi sebagai data positif.
3. *True Positive* (TP) merupakan data positif yang terdeteksi benar.
4. *False Negative* (FN) merupakan kebalikan dari *True Positive*, data positif namun terdeteksi sebagai data negatif

Tabel 2.2. *Confusion matrix*

	Aktual			
Prediksi		Kelas 1	Kelas 2	Kelas 3
	Kelas 1	T11	F12	F13
	Kelas 2	F21	T22	F23
	Kelas 3	F31	F32	T34

Dengan tabel 2.2 tersebut bisa diketahui parameter *Accuracy, Precision, Recall* dan *F1-score*.

2. *Presisi*

Presisi merupakan rasio prediksi benar positif (*True Positive*) dibandingkan dengan keseluruhan hasil yang diprediksi positif. Rumus dari presisi dinyatakan pada Persamaan (2.2)

$$precision = \frac{TP}{(TP + FP)} * 100\% \quad (2.2)$$

3. Recall

Recall atau rasio *true positive* adalah ukuran untuk berapa banyak *true positive* yang diprediksi dari semua positif dalam kumpulan data. Kadang juga disebut kepekaan (sensitivity). Rumus *recall* dapat dituliskan seperti pada Persamaan (2.3).

$$recall = \frac{TP}{(FP + FN)} * 100\% \quad (2.3)$$

4. Akurasi dan Validasi Akurasi

Akurasi digunakan untuk mengukur kinerja algoritma dengan cara yang dapat ditafsirkan. Akurasi suatu model biasanya ditentukan dalam bentuk persentase. Akurasi adalah ukuran seberapa akurat prediksi model dibandingkan dengan data sebenarnya dan Akurasi dihitung berdasarkan *data train*. Rumus untuk akurasi ditunjukkan pada Persamaan (2.4).

$$akurasi = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\% \quad (2.4)$$

Sedangkan validasi akurasi dihitung berdasarkan data validasi. Yang terbaik adalah mengandalkan validasi akurasi dari kinerja model, karena *neural network*

yang baik pada akhirnya akan menyesuaikan data train pada 100%, tetapi akan berkinerja buruk pada data yang baru ditemuinya.

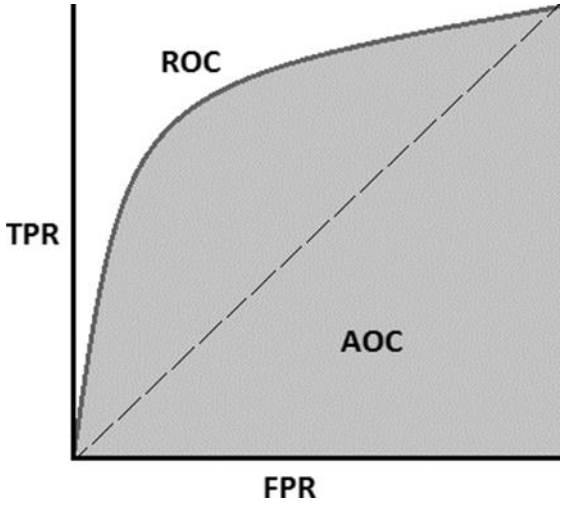
5. F1-Score

F1-Score atau Skor F1 adalah skor *F* yang paling umum digunakan. Ini adalah kombinasi presisi dan recall, yaitu rata-rata harmoniknya. *F1-Score* dapat dihitung melalui Persamaan (2.5).

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{2.5}$$

6. Kurva ROC-AUC

Kurva AUC-ROC adalah pengukuran kinerja untuk masalah klasifikasi pada berbagai pengaturan ambang batas. *Receiver Operator Characteristic* (ROC) adalah kurva probabilitas dan Area *Under the Curve* (AUC) adalah ukuran yang digunakan sebagai ringkasan dari kurva ROC. Ini memperlihatkan seberapa besar model mampu membedakan antar kelas. Semakin tinggi AUC, semakin baik model dalam memprediksi 0 sebagai 0 dan 1 sebagai 1. Dengan analogi, semakin tinggi AUC, semakin baik model dalam membedakan antara pasien dengan penyakit dan tidak ada penyakit.



Gambar 2.15 Kurva *AUC - ROC*

Kurva ROC diplotkan dengan *True Positive Rate*(TPR) terhadap *False Positive Rate*(FPR) dimana TPR berada pada sumbu y dan FPR pada sumbu x seperti yang terlihat pada Gambar 2.19.