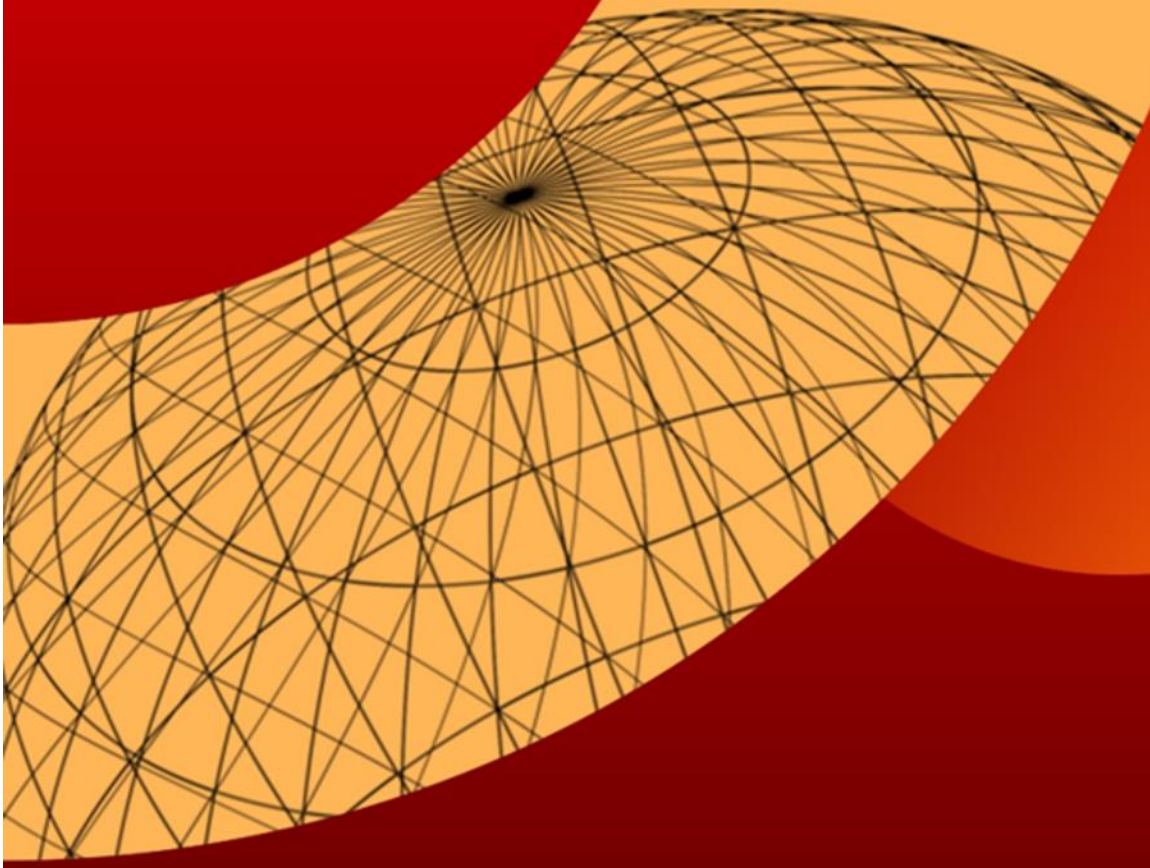
April 2025
Volume 6 Nomor 1
p-ISSN 2723-0325 e-ISSN 2723-0333



TENSOR

Pure and Applied Mathematics Journal

PROGRAM STUDI MATEMATIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PATTIMURA

TENSOR

Pure and Applied Mathematics Journal

is an international academic open-access journal that gains a foothold in mathematics, and its applications are issued twice a year. The focus is to publish original research and review articles on all aspects of pure and applied Mathematics. Editorial board members of the Journal and reviewers will review submitted papers. All submitted articles should report original, previously unpublished research results, experimental or theoretical, and will be peer-reviewed. Articles submitted to the journal should meet these criteria and must not be under consideration for publication elsewhere. Manuscripts should follow the journal template and are subject to both review and editing.

Published by:

Department of Mathematics,
Faculty of Science and Technology,
Pattimura University.
Ambon
2025
Copyright© Program Studi Matematika FST UNPATTI 2025

TENSOR
Pure and Applied Mathematics Journal
Volume 6 Number 1 | April 2025

Person In Charge
Head of Undergraduate Program in Mathematics,
Faculty of Mathematics and Natural Sciences, Pattimura University

Editor in Chief
Dr. H. Batkunde, S.Si, M.Si

Editors
M. I. Tilukay, S.Si, M.Si (Managing and Section Editor)
L. Bakarbessy, S.Si, M.Si (Managing and Section Editor)
Z. A. Leleury, S.Si., M.Si (Copy and Production Editor)
B. P. Tomasouw, S.Si, M.Si (Copy and Production Editor)
Dr. L. K. Beay, S.Pd., M.Si (Proofreader)
N. Dahoklory (Proofreader)

Secretariat and Financial Officer
M. E. Rijoly, S.Si, M.Sc

Graphic Design
V. Y. I. Ilwaru, S.Si, M.Si

Expert Editorial Boards
Prof. Dr. Basuki Widodo, M.Sc (Institut Teknologi Sepuluh November Surabaya, Indonesia)
Prof. Dr. M. Salman A. N, M.Si (Institut Teknologi Bandung, Indonesia)
Dr. H. J. Wattimanela, S.Si., M.Si (Universitas Pattimura, Indonesia)
Dr. Al Azhary Masta, S.Si., M.Si (Universitas Pendidikan Indonesia, Indonesia)
Dr. Muh. Nur, S.Si., M.Si (Universitas Hasanudin, Indonesia)
Dr. Meta Kallista, S.Si., M.Si (Universitas Telkom,Indonesia)
Dr. Teguh Herlambang, S.Si., M.Si (Universitas Nahdlatul Ulama Surabaya, Indonesia)
Asst. Prof. Dr. Anurak Thanyacharoen (Muban Chombueng Rajabhat University, Ratchaburi, Thailand)

Publisher
Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Pattimura University, Ambon, Indonesia

Editorial Address
Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Unversitas Pattimura
Jln. Ir. M. Putuhena, Kampus Unpatti, Poka - Ambon 97233, Provinsi Maluku, Indonesia
Contact : +62 82397854220
Email : tensormathematics@gmail.com

Volume 6 Number 1 | September 2024

TENSOR
Pure and Applied Mathematics Journal

Modeling the Spread of Hepatitis B Disease from the SEIR Model in East Java Using RKF45	Sakinun Na'malia Faisol Tony Yulianto	1-12
Analisis Perbandingan Optimasi *Stochastic Gradient Descent* dan *Adaptive Moment Estimation* dalam Klasifikasi Emosi dari Audio Menggunakan *Convolutional Neural Network*	Aldelia Jocelyn Tutuhatunewa Dorteus Lodewyik Rahakbauw Zeth Arthur Leleury	13-22
Kajian Basis dan Dimensi pada Ruang Hipervektor Atas Lapangan	Loisa Genesis Kambu Henry W. M. Patty Lusye Bakarbessy Novita Dahoklory	23-38
Fungsi Trace dan Fungsi Norm Lapangan Perluasan Atas $\mathbb{Q}$	Novita Dahoklory Henry W. M. Patty	39-48
Optimasi Model LSTM untuk Prediksi Curah Hujan di Kota Ambon: Perbandingan Mean Imputation dan Interpolasi dalam Prediksi Data Time Series	Emanuella M. C. Wattimena Pranaya D. M. Taihuttu Devi V. Waas Citra F. Palembang Victor E. Pattiradjawane[1]	49-56
The Rainbow Vertex Connection Number of Some Amalgamation of Two Cycles	Pranaya D. M. Taihuttu Meilin I. Tilukay Francis Y. Rumlawang E. M. C. Wattimena	57-66

TENSOR
Pure and Applied Mathematics Journal

April | Vol. 6 | No. 1 pp. 13-22
p-ISSN: 2723-0325 e-ISSN: 2723-0333

Analisis Perbandingan Optimasi *Stochastic Gradient Descent* dan *Adaptive Moment Estimation* dalam Klasifikasi Emosi dari Audio Menggunakan *Convolutional Neural Network*

Aldelia Jocelyn Tutuhatunewa[1*], Dorteus Lodewyik Rahakbauw[2], Zeth Arthur Leleury[3]
[1,2,3] Jurusan Matematika Fakultas Sains dan Teknologi Universitas Pattimura, Jalan Ir. M . Putuhena, Ambon, Indonesia.
*Email: aldeliajoe@gmail.com

Manuscript submitted : February 2025;
Accepted for publication : April 2025.
doi: https://doi.org/10.30598/tensorvol6iss1pp13-22

Abstract: Emotion plays a fundamental role in human life, influencing behavior, social interaction, and decision-making. Successful communication and understanding between individuals depend greatly on our ability to recognize and express emotions. In this context, sound or audio plays a key role as a medium that reflects and conveys human emotional expression. In the era of information technology and artificial intelligence, emotion recognition through sound has become a growing focus of research. Machine learning algorithms, particularly neural networks, can be trained to understand and classify emotions conveyed in various forms, including text, images, videos, and audio. Among these algorithms, Convolutional Neural Network (CNN) has shown promising performance in emotion classification tasks. In this study, the comparison between Stochastic Gradient Descent (SGD) and Adaptive Moment Estimation (Adam) optimizers in emotion classification from audio using CNN is investigated. The research aims to determine the optimal optimizer for emotion classification tasks. The results suggest that SGD optimizer outperforms Adam in terms of overall accuracy, with SGD achieving 53% accuracy compared to Adam's 48% accuracy in The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset. Therefore, for emotion classification from audio data, Stochastic Gradient Descent (SGD) optimizer is recommended for better performance.
Keywords: Sound, Emotion recognition, Convolutional Neural Network (CNN), Audio classification, Mel Frequency Cepstral Coefficients (MFCC), Stochastic Gradient Descent (SGD), Adaptive Moment Estimation (Adam).

1. Pendahuluan

Emosi merupakan aspek fundamental dalam kehidupan manusia yang memengaruhi perilaku, interaksi sosial, dan pengambilan keputusan. Keberhasilan komunikasi dan pemahaman antar individu sangat tergantung pada kemampuan kita untuk mengenali dan mengekspresikan emosi. Dalam ranah ini, suara atau audio memainkan peran kunci sebagai medium yang mencerminkan dan membawa ekspresi emosi manusia.

Dalam era teknologi informasi dan kecerdasan buatan, pengenalan emosi melalui suara telah menjadi

fokus penelitian yang semakin berkembang. Mesin dan model pembelajaran mesin (*machine learning*), khususnya neural networks, dapat diajarkan untuk memahami dan mengklasifikasikan emosi yang terkandung terkandung dalam teks, gambar, video maupun audio. Tiga algoritma yang biasa digunakan untuk klasifikasi citra diantaranya adalah *Support Vector Machine* (SVM), K-*Nearest Neighbor* (KNN), dan *Convolutional Neural Network* (CNN). Dimana diantara ketiganya ditemukan bahwa CNN memiliki akurasi paling tinggi pada performa modelnya dengan nilai parameter *Accuracy* 0.942, *Precision* 0.943, *Recall* 0.942, dan F1-*Score* 0.942 [1].

Dalam membangun model CNN, fungsi *Optimizer* seringkali digunakan untuk memaksimalkan performa model. *Optimizer* adalah algoritma atau metode yang digunakan untuk meminimalkan atau memaksimalkan fungsi objektif (*objective function*) selama proses pelatihan (*training*). Banyak optimizer yang digunakan untuk mengelola laju pembelajaran (*learning rate*) yang merupakan parameter yang mengatur seberapa besar perubahan yang akan diterapkan pada bobot selama proses pelatihan. Beberapa *optimizer* yang sering dipakai untuk algoritma CNN antara lain *Adaptive Moment Estimation* (Adam) [2], Stochastic Gradient Descent (SGD) [3], dan *Root Mean Square Propagation* (RMSProp). Klasifikasi emosi dalam audio pernah dilakukan menggunakan metode *Multilayer Perceptron* yang mengklasifikasi 8 jenis emosi mencapai akurasi rata-rata sebesar 96% [4]. Namun pada penelitian tersebut hanya menggunakan *Optimizer* Adam.

Penelitian bertujuan untuk memperoleh hasil tingkat akurasi *optimizer* SGD dalam klasifikasi emosi dari audio menggunakan metode CNN, memperoleh hasil tingkat akurasi *optimizer* Adam dalam klasifikasi emosi dari audio menggunakan metode CNN, dan memperoleh perbedaan performa antara SGD dan Adam dalam mencapai akurasi pada klasifikasi emosi dari audio menggunakan metode CNN.

2. Metode Penelitian

2.1. Tipe Penelitian

Penelitian ini adalah penelitian eksperimental yang melakukan komparasi antara 2 jenis *optimizer*. Penelitian eksperimental digunakan untuk menguji hipotesis kausal tentang hubungan antara dua variabel atau lebih. Peneliti akan melakukan serangkaian eksperimen untuk mengumpulkan data, mengimplementasikan model, dan menganalisis hasil eksperimen untuk mengambil kesimpulan tentang efektivitas dua teknik optimasi dalam konteks yang diberikan. Dalam penelitian ini data yang digunakan adalah data *open-source The Ryerson Audio-Visual Database of Emotional Speech and Song* (RAVDESS). Dataset ini dikumpulkan oleh penulis dari website kaggle.com. Sedangkan alat yang digunakan adalah *software* Python serta *library* TensorFlow dan Keras. Penelitian ini menggunakan algoritma CNN dengan membandingkan 2 *optimizer* (Adam dan SGD) dan mengklasifikasi audio ke lebih banyak kelas, yaitu sebanyak 8 kelas atau jenis emosi. Metode ekstraksi yang digunakan merupakan metode MFCC dan hasil akan dianalisis berdasarkan akurasi model.

2.2. Tahapan Penelitian

Berikut adalah tahapan penelitian yang dilakukan yang secara ringkas ditampilkan pada *flowchart* atau bagan alir dari penelitian ini seperti pada Gambar 1.

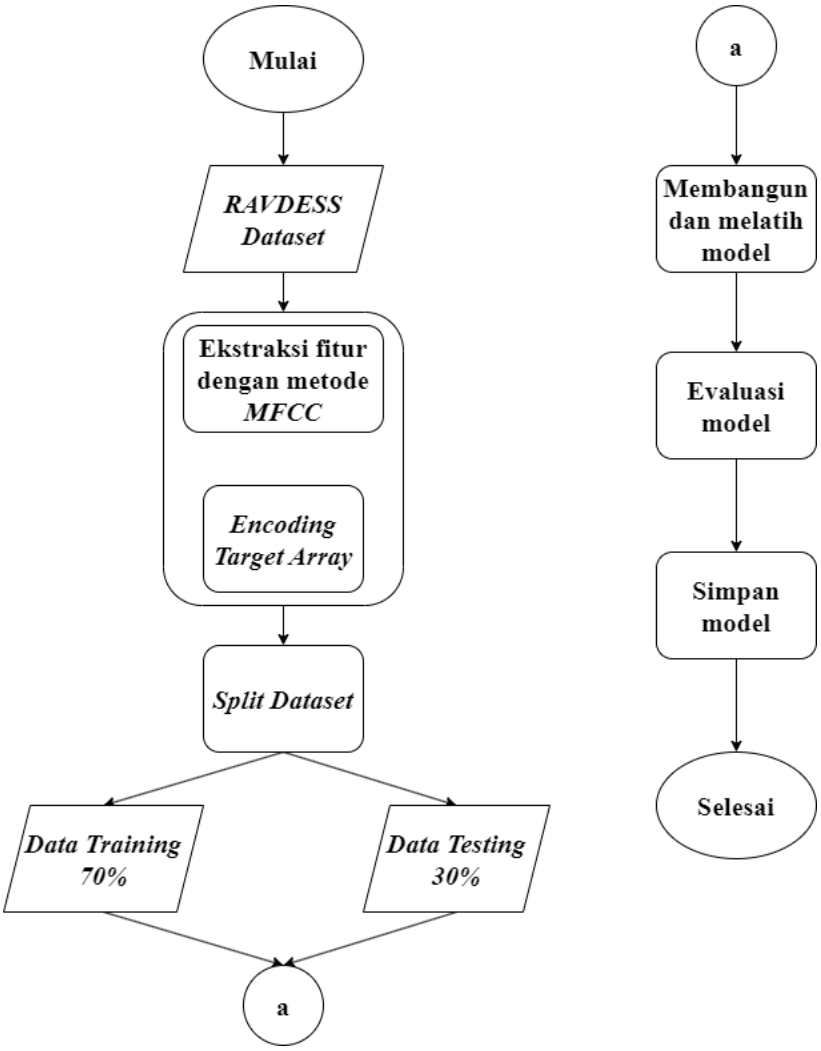
1. Pengumpulan Data
   Melakukan proses mengumpulkan data yang akan dilakukan pada platform *open-source* Kaggle. Data yang akan dipakai adalah RAVDESS *emotional speech* radio dataset yang diambil dari tautan kaggle.com.
2. Pra-pemrosesan Data
   Ekstraksi Fitur Audio: Menggunakan teknik ekstraksi fitur audio seperti *Mel-frequency cepstral coefficients* (MFCCs) untuk mengonversi sinyal audio menjadi representasi fitur yang dapat digunakan oleh model CNN.

Label data: Memastikan setiap data telah diberi label yang benar agar memastikan proses pelatihan mencapai akurasi optimal.

Label encoding: digunakan untuk mengubah label yang berbentuk kategorik menjadi numerik. Dimana angka masing-masing emosi adalah 0 = netral, 1 = tenang, 2= senang, 3 = sedih, 4 = marah, 5 = takut, 6 = jijik, 7 = terkejut.

Pembagian Data: Membagi dataset menjadi subset pelatihan dan pengujian dengan proporsi tertentu untuk melatih dan menguji model. Perbandingan yang dipakai adalah 70% data training dan 30% data testing.

3. Pembangunan Model Convolutional Neural Network (CNN)
   Mendesain arsitektur CNN yang sesuai untuk tugas klasifikasi emosi dari data audio. Melibatkan lapisan konvolusi, pooling, dan lapisan fully connected.
4. Pelatihan Model
   Menggunakan subset pelatihan untuk melatih model CNN dengan menggunakan kedua optimizer, SGD dan Adam. Melibatkan penyesuaian learning rate serta ukuran batch.
5. Evaluasi Model
   Mengukur kinerja model menggunakan subset pengujian. Matrik evaluasi mencakup akurasi training dan akurasi validasi.
6. Analisis Hasil
   Menganalisis hasil eksperimen untuk membandingkan performa model menggunakan optimizer SGD dan Adam.



Gambar 1. Skema Alur Penelitian

2.3. Analisa Hasil

Analisis hasil dilakukan dengan mengevaluasi model untuk mengetahui tingkat performa model dalam mengklasifikasikan berbagai kelas. Salah satu teknik evaluasi model adalah dengan menggunakan *confusion matrix*. *Confusion matrix* adalah sebuah matriks yang menunjukkan dan membandingkan nilai aktual atau nilai sebenarnya dengan nilai hasil prediksi model yang dapat digunakan untuk menghasilkan metrik evaluasi seperti *Accuracy* (akurasi), *Precision*, *Recall*, dan *F1-Score*.

Terdapat 4 nilai yang dihasilkan dalam tabel *confusion matrix*, yaitu *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN).

a. TN merupakan jumlah data negatif yang terdeteksi dengan benar.
b. FP merupakan data negatif namun terdeteksi sebagai data positif.
c. TP merupakan data positif yang terdeteksi benar.
d. FN merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif.

Tabel 1. Tabel *Confusion Matrix*

		True Values	
		True	***False***
Prediction	***True***	*TP* *Correct result*	*FP* *Unexpected result*

False	*FN* *Missing result*	*TN* *Correct absence of result*

a. Precision

Precision atau presisi adalah data yang diambil berdasarkan informasi yang kurang. Dalam klasifikasi biner, presisi dapat dibuat sama dengan nilai prediksi positif. *Precision* adalah ukuran ketepatan model dalam mengidentifikasi sampel positif. Precision menunjukkan seberapa sering prediksi positif model benar. Rumus mencari presisi:

$$Precision = \left(\frac{TP}{(TP + FP)}\right) \times 100\% \quad (1)$$

b. Recall

Recall adalah data penghapusan yang berhasil diambil dari data yang relevan dengan *query*. Dalam klasifikasi biner, *recall* dikenal sebagai sensitivitas. Munculnya data relevan yang diambil adalah menyetujui dengan *query* dapat dilihat dengan recall. *Recall* adalah ukuran sensitivitas model dalam menangkap semua sampel positif. *Recall* menunjukkan seberapa baik model dalam mengidentifikasi semua sampel yang benar-benar positif. Rumus mencari Recall:

$$Recall = \left(\frac{TP}{(TP + FN)}\right) \times 100\% \quad (2)$$

c. F1-Score

Nilai *F1-Score* atau dikenal juga dengan nama F *Measure* didapatkan dari hasil presisi dan *recall* antar kategori hasil prediksi dengan kategori sebenarnya. F1-*Score* adalah ukuran harmonis dari *precision* dan recall. F1-*Score* memberikan gambaran keseluruhan tentang performa model dengan mempertimbangkan keduanya. F1-*Score* memberikan keseimbangan antara *precision* dan *recall*. Rumus mencari F1-*Score*:

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (3)$$

d. Accuracy

Accuracy atau akurasi digunakan untuk mengukur kinerja algoritma dengan cara yang dapat ditafsirkan. Akurasi suatu model biasanya ditentukan setelah parameter model dan dihitung dalam bentuk persentase. Ini adalah ukuran seberapa akurat prediksi model dibandingkan dengan data sebenarnya dan akurasi berada pada *train*. Rumus mencari akurasi:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN))} \times 100\% \quad (4)$$

3. Hasil dan Pembahasan

3. 1 Pemrosesan Dataset

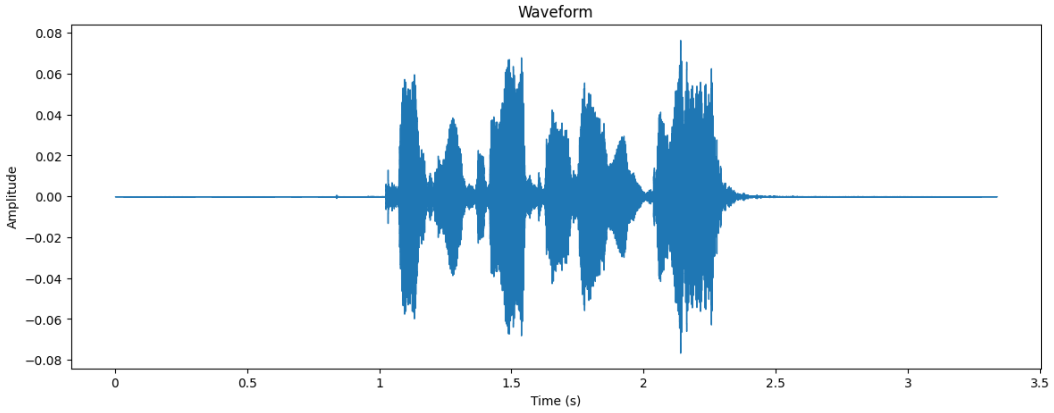
Data yang digunakan dalam penelitian ini adalah rekaman suara manusia dalam bahasa Inggris, yang sudah dilengkapi dengan label sebelumnya. Dataset yang digunakan adalah Ryerson Audio-Visual Dataset of Emotional Speech and Song (RAVDESS), yang dikumpulkan oleh penulis dari situs web Kaggle. Aksen suara dalam dataset ini berasal dari Amerika Utara. Dataset terdiri dari dua jenis file, yaitu file yang berisi percakapan individu dan file yang berisi lagu. Dalam penelitian ini, kami akan menggunakan file rekaman percakapan, yang berjumlah 1440 file audio. Durasi setiap audio dalam dataset ini berkisar antara 3,5 hingga 4 detik. Rekaman percakapan dalam dataset ini melibatkan suara laki-laki dan perempuan, masing-masing terdiri dari 12 individu. Setiap file audio dalam dataset ini memiliki format 16 bit, 48 kHz dengan ekstensi file .wav.

Tabel 2. Jumlah Data Keseluruhan

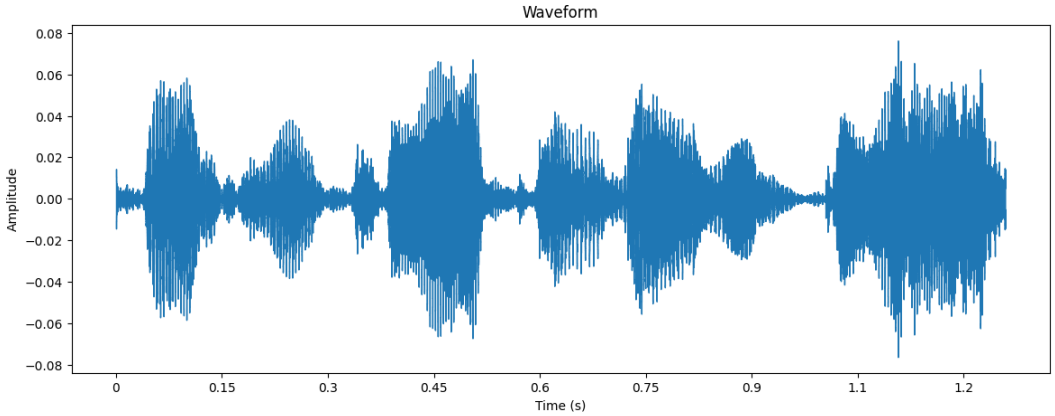
Emosi	Pria	Wanita
Netral	48	48
Tenang	96	96
Bahagia	96	96
Sedih	96	96
Marah	96	96
Takut	96	96
Jijik	96	96
Terkejut	96	96
Total	720	720

Data audio yang digunakan sebelum diekstraksi berbentuk gelombang audio dari representasi amplitudo suara terhadap waktu. Jika divisualisasikan dalam grafik dapat dilihat bagaimana amplitudo suara berubah seiring waktu. Pada rekaman suara manusia, gelombang audio akan menunjukkan fluktuasi amplitudo yang mencerminkan variasi dalam suara yang direkam, seperti pitch (nada), volume, dan durasi. Grafik gelombang audio ini memberikan gambaran visual tentang karakteristik akustik dari rekaman suara sebelum ekstraksi fitur-fitur seperti Mel-Frequency Cepstral Coefficients (MFCC). Pada Gambar 2 merupakan grafik dari salah satu audio dari dataset yang akan digunakan.

Data audio kemudian melewati proses silent removal, yaitu proses pemotongan audio di bagian awal dan akhir untuk menghilangkan bagian diam sebelum dan sesudah aktor berbicara. Proses ini dilakukan dengan bantuan software MATLAB. Sampel hasil data yang telah dipotong atau telah dilakukan silent removal dapat dilihat pada gambar 3 di bawah ini.



Gambar 2. Visualisasi Audio Senang Sebelum *Silent Removal*



Gambar 3. Visualisasi Audio Senang Sesudah *Silent Removal*

Data yang telah melalui proses pemotongan dan dipakai sebagai data dalam penelitian dapat dilihat pada tabel 3 berikut.

Tabel 3. Jumlah Audio Setelah Pemotongan

Kelas	Emosi	Jumlah
1	Netral	96
2	Tenang	190
3	Bahagia	191
4	Sedih	192
5	Marah	192
6	Takut	191
7	Jijik	192
8	Terkejut	191
	Total	1435

Keseluruhan dataset yang berjumlah 1435 audio ini kemudian akan diekstraksi menggunakan metode MFCC dengan pemrograman python yang dijalankan di google colab yang memungkinkan kita memperoleh matriks dari masing-masing audio. Dimana matriks akan berukuran jumlah koefisien MFCC kali jumlah *frame*.

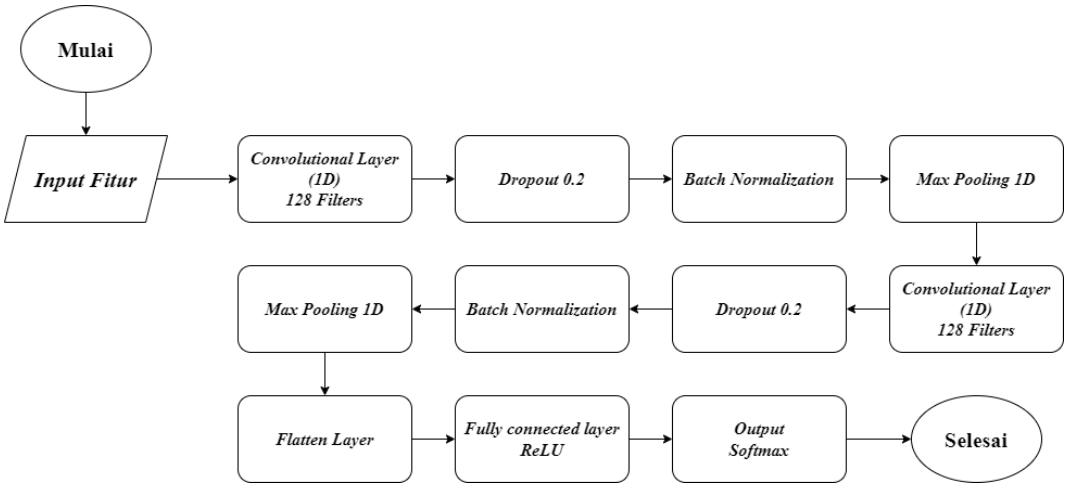
Tabel 4. Sampel Hasil Ekstraksi MFCC dari Audio

	1	2	3	4	...	13
1	-452,251	197,7236	-7,87114	-19,7107	...	0,542556
2	-458,808	204,3042	-23,309	-20,7977	...	-1,46148
3	-485,871	203,8673	-29,3221	-27,0689	...	-9,04032
4	-503,372	209,1617	-11,0603	-30,559	...	-7,3006
5	-533,569	196,6293	-1,95884	-20,9173	...	-5,19639
...	...	...	...	...	...	...
322	-481,685	148,3774	20,60678	2,111137	...	0,33366

Tabel 4 merupakan sampel data hasil ekstraksi MFCC berukuran 13 x 322. Semua data kemudian dikenakan *padding* agar memiliki panjang *frame* yang sama dan sehingga ukuran tiap data sama. Data tersebut kemudian dibagi menjadi data *training* dan data *testing* dengan 70% data *training* dan 30% data *testing*.

3. 2 Pelatihan dan Pengujian Model

Pada tahap ini dilakukan perancangan arsitektur untuk model CNN yang akan digunakan dalam penelitian. Pada penelitian ini akan terdiri dari 2 lapisan konvolusi dengan fungsi aktivasi ReLU dan diikuti dengan *fully connected layer* yang terdiri dari *flatten layer*. Setelah itu terdapat hidden layer dan output layer. Di layer output inilah dimana fungsi aktivasi *Softmax* digunakan untuk mengklasifikasi audio ke dalam 8 kelas emosi yang ada.



Gambar 4. Diagram Arsitektur CNN

Tahapan selanjutnya adalah pelatihan model convolutional neural network dengan data training. Sebelum dilakukan proses training, perlu dilakukan inisialisasi parameter yang akan digunakan seperti *loss function*, *optimizer*, jumlah *epoch*, jumlah *batch*, dan *callback*.

Tabel 5. Inisialisasi Parameter

Parameter	Masukkan
Output class	8
Output layer activation	*Softmax*
Hidden layer activation	ReLU
Epoch	50
Batch size	10 / 32
Optimizer function	SGD / Adam
Loss function	*Categorical cross entropy*

Kemudian setelah inisialisasi parameter, selanjutnya data training akan dilatih sesuai arsitektur CNN yang telah dirancang. Pada proses ini juga akan dilakukan proses validasi dengan untuk melihat apakah model

dengan data training sudah belajar dengan baik. Hasil pelatihan dilakukan dengan 2 *batch size* dan 9 *learning rate*. Berikut data akurasi training dan validasi model dapat dilihat pada tabel 6 dan tabel 7.

Tabel 6. Data Akurasi Hasil Pelatihan dan Validasi Model dengan *Optimizer* SGD

Learning Rate	*Batch_size = 10*		*Batch_size = 32*	
	Accuracy	*Val_ Accuracy*	*Accuracy*	*Val_ Accuracy*
0.001	0.93	0.47	0.73	0.42
0.002	0.96	0.53	0.84	0.46
0.003	0.97	0.44	0.91	0.42
0.004	0.98	0.47	0.92	0.44
0.005	0.97	0.45	0.90	0.44
0.006	0.97	0.47	0.91	0.45
0.007	0.98	0.48	0.96	0.47
0.008	0.98	0.45	0.95	0.48
0.009	0.99	0.34	0.90	0.47

Tabel 7. Data Akurasi Hasil Pelatihan dan Validasi Model dengan *Optimizer* Adam

Learning Rate	*Batch_size = 10*		*Batch_size = 32*	
	Accuracy	*Val_ Accuracy*	*Accuracy*	*Val_ Accuracy*
0.0001	0.91	0.43	0.93	0.44
0.0002	0.89	0.44	0.89	0.48
0.0003	0.96	0.45	0.83	0.40
0.0004	0.85	0.45	0.59	0.32
0.0005	0.49	0.35	0.61	0.32
0.0006	0.45	0.37	0.88	0.43
0.0007	0.92	0.42	0.77	0.38
0.0008	0.81	0.41	0.71	0.39
0.0009	0.73	0.36	0.73	0.40

3. 3 Analisis Hasil

Tabel 8. Hasil Perhitungan *Precision*, *Recall*, dan F1-*Score* dengan *Optimizer* SGD

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Precision	43,47	56,41	46,55	55	62,06	63,82	47,69	43,18
Recall	52,63	77,19	46,55	42,3	62,06	51,72	53,44	32,75
F1-Score	0,38	0,65	0,47	0,55	0,62	0,57	0,5	0,37

Untuk model yang menggunakan optimizer SGD, Label 1 memiliki performa terbaik dengan F1-*Score* 0,65, *precision* 56,41%, dan *recall* 77,19%. Ini menunjukkan model mampu mengidentifikasi sebagian besar sampel Label 1 dengan akurat dan konsisten. Sedangkan Label 7 memiliki performa terendah dengan F1-*Score* 0,37, *precision* 43,18%, dan *recall* 32,75%. Ini menandakan model kesulitan dalam mengidentifikasi dan menangkap sampel untuk kategori ini.

Tabel 9. Hasil Perhitungan *Precision*, *Recall*, dan F1-*Score* dengan *Optimizer* Adam

	0	1	2	3	4	5	6	7
Precision	36,33	46,15	52,08	44,61	61,01	55,88	47,5	42,85
Recall	27,58	73,68	43,1	50	62,06	32,75	65,51	25,86
F1-Score	0,31	0,57	0,47	0,47	0,62	0,41	0,55	0,32

Model dengan menggunakan optimizer Adam menunjukkan performa terbaik dalam mendeteksi emosi Marah (Label 4), yang ditandai dengan *precision*, *recall*, dan F1-*Score* yang tinggi. Ini berarti model dapat dengan baik mengidentifikasi dan memprediksi emosi ini dengan akurasi yang baik. Sebaliknya, model mengalami kesulitan terbesar dalam mendeteksi emosi Netral (Label 0), yang ditandai dengan nilai *precision* dan F1-*Score* terendah, serta emosi Terkejut (Label 7), yang memiliki nilai *recall* terendah. Hal ini menunjukkan bahwa model perlu perbaikan lebih lanjut untuk meningkatkan akurasi dan konsistensi dalam mendeteksi emosi-emosi tersebut.

4. Kesimpulan dan Saran

4. 1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dengan menggunakan model *Convolutional Neural Networ*k (CNN) yang memakai 9 jenis *learning rate* dan 2 jenis *batch size*, maka dapat disimpulkan bahwa:

1. Hasil pelatihan *Convolutional Neural Network* (CNN) dengan menggunakan *optimizer* SGD, memiliki nilai akurasi validasi sebesar 53% dengan menggunakan *learning rate* 0,002 dan *batch size* 10.
2. Sedangkan untuk *optimizer* Adam hasil pelatihan terbaik berada pada *learning rat*e 0,0002 dan *batch size* 32 dimana akurasi validasi yang diperoleh sebesar 48%.
3. Hasil analisis *confusion matrix* model CNN dengan optimizer SGD memiliki akurasi umum lebih tinggi, yaitu sebesar 53% dibandingkan *optimizer* Adam yang sebesar 48%.

Selain itu, hasil dari pelatihan juga menunjukkan model lebih mudah mengenali emosi berlabel 1 yaitu audio dengan emosi tenang dan paling sulit mengenali audio berlabel 0 yaitu audio emosi netral. Maka dalam kasus klasifikasi emosi dari dataset audio RAVDESS ini, optimizer yang lebih baik adalah *optimizer Stochastic Gradient Descent* (SGD).

4. 2 Saran

Untuk meningkatkan literasi terkait klasifikasi emosi dengan menggunakan *Convolutional Neural Network* (CNN) diperlukan penyesuaian parameter yang lebih teliti dan beragam agar dapat menghindari adanya overfitting yang ditandai dengan tingginya akurasi pelatihan namun rendahnya akurasi validasi.

References

[1] Naufal, M. F. (2021). Analisis Perbandingan Algoritma Svm, Knn, Dan Cnn untuk Klasifikasi Citra Cuaca. Jurnal Teknologi Informasi dan Ilmu Komputer, 8(2), 311-317. doi: 10.25126/jtiik.2021824553.

[2] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[3] Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[4] Ardiyansyah, D., & Jayanta, J. (2021). Model Klasifikasi Emosi Berdasarkan Suara Manusia dengan Metode Multilayer Perceptron. In Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya 2,(1), pp. 689-702).