

SKRIPSI
PERBANDINGAN CNN DAN VISION TRANSFORMER (ViT) PADA
KLASIFIKASI GENRE MUSIK MENGGUNAKAN LOG-MEL
SPECTROGRAM



OLEH :
SALOMO JULIO ELSADA LAUTT
223020503080

JURUSAN/PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS PALANGKA RAYA

2026

HALAMAN PENGESAHAN

PERBANDINGAN CNN DAN VISION TRANSFORMER (ViT) PADA KLASIFIKASI GENRE MUSIK MENGGUNAKAN LOG-MEL SPECTROGRAM

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan Program Strata-1 pada Jurusan Teknik Informatika Fakultas
Teknik Universitas Palangka Raya

Oleh

Salomo Julio Elsada Lautt

223020503080

Telah dipertahankan didepan tim penguji, pada :

Hari/Tanggal : Selasa, 2 Juni 2026

Waktu : 13:05 - 14:35 WIB

Tempat : Ruangan Ujian TI 1

1. NAHUMI NUGRAHANINGSIH, S.T., M.T., PH.D. 
NIP. 197910092008012016 : (Ketua Penguji)
2. RESSA PRISKILA, S.T., M.T. 
NIP. 199403012019032016 : (Pembimbing Utama/Pertama)
3. DEDDY RONALDO, S.T., M.T. 
NIP. 198012262008121002 : (Pembimbing Pendamping/Kedua)
4. SEPTIAN GEGES, S.KOM., M.KOM. 
NIP. 199109172020121004 : (Anggota)

Mengetahui :

**Jurusan / Program Studi Teknik Informatika
Fakultas Teknik Universitas Palangka Raya
Ketua Jurusan,**



**ARIESTA LESTARI, S.Kom., M.Cs., Ph.D.
NIP. 198003222005012004**

HALAMAN PERNYATAAN

PERNYATAAN

Dengan ini saya menyatakan dengan sebenar-benarnya bahwa dalam Skripsi ini tidak terdapat karya ilmiah yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, serta tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam Skripsi ini dan disebutkan dalam Tinjauan Pustaka.

Palangka Raya, 10 Juni 2026



SALOMO JULIO ELSADA LAUTT

223020503080

RIWAYAT PENYUSUN

Data Diri

Nama : Salomo Julio Elsada Lautt
NIM : 223020503080
Fakultas : Teknik
Jurusan/Program Studi : Teknik Informatika
Jenjang : Strata 1 (S-1)
Jenis Kelamin : Laki-laki
Tempat, Tanggal Lahir : Palangka Raya, 29 Juli 2003
Agama : Kristen Protestan
Status Dalam Keluarga : Kandung
Anak Ke : 2
Alamat : Jl. Ir. H. Juanda No 07A
No Telepon/HP : 082156342551



Data Orang Tua

Nama Ayah : Immanuel Revana Lautt (Alm)
Pekerjaan Ayah : -
Nama Ibu : Marilyn Patmiriantara
Pekerjaan Ibu : Wiraswasta
Alamat Orang Tua : Jl. Ir. H. Juanda No 07A
No.HP Orang Tua : 085167220260

Riwayat Pendidikan

SD : SDN 4 Menteng Palangka Raya (Tahun Lulus 2015)
SMP : SMPN 2 Palangka Raya (Tahun Lulus 2018)
SMA : SMAN 2 Palangka Raya (Tahun Lulus 2021)

Palangka Raya, 10 Juni 2026

Salomo Julio Elsada Lautt
223020503080

HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan penuh rasa syukur, penulis mengucapkan puji dan syukur kepada Tuhan Yang Maha Esa atas kasih, berkat, serta penyertaan-Nya yang senantiasa menyertai setiap langkah penulis hingga skripsi ini dapat diselesaikan. Sebagai ungkapan rasa terima kasih dan penghargaan yang tulus, penulis mempersembahkan karya ini kepada:

1. Orang tua tercinta, khususnya almarhum Ayah yang senantiasa menjadi inspirasi dan motivasi dalam kehidupan penulis, serta Ibu yang selalu memberikan kasih sayang, doa, dukungan, semangat, dan pengorbanan yang tiada henti sehingga penulis dapat menyelesaikan pendidikan dan skripsi ini dengan baik.
2. Saudara penulis, Oska Elsada Lattu, selaku kakak, yang senantiasa memberikan dukungan, motivasi, dan semangat kepada penulis selama menempuh pendidikan hingga penyelesaian skripsi ini.
3. Ibu Ressa Priskila, S.T., M.T. selaku Dosen Pembimbing Utama dan Bapak Deddy Ronaldo, S.T., M.T. selaku Dosen Pembimbing Pendamping, yang dengan penuh kesabaran telah memberikan bimbingan, arahan, ilmu pengetahuan, masukan, serta motivasi kepada penulis selama proses penyusunan skripsi ini.
4. Ibu Nahumi Nugrahaningsih, S.T., M.T., Ph.D. selaku Ketua Penguji dan Bapak Septian Geges, S.Kom., M.Kom. selaku Dosen Penguji, yang telah memberikan kritik, saran, serta masukan yang membangun demi penyempurnaan skripsi ini.
5. Teman-teman dan sahabat penulis di lingkungan kampus yang telah menemani, mendukung, serta berjuang bersama sejak awal perkuliahan hingga akhir masa studi. Terima kasih atas kebersamaan, pengalaman, bantuan, dan kenangan yang telah diberikan selama perjalanan akademik ini.
6. Diri penulis sendiri yang telah berusaha dan menyelesaikan seluruh proses perkuliahan serta penyusunan skripsi ini dengan sebaik-baiknya.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan CNN Dan Vision Transformer (Vit) Pada Klasifikasi Genre Musik Menggunakan Log-Mel Spectrogram” dengan baik.

Skripsi ini membahas perbandingan performa model Convolutional Neural Network (CNN) dan Vision Transformer (ViT) dalam melakukan klasifikasi genre musik menggunakan representasi log-Mel spectrogram. Penelitian ini bertujuan untuk menganalisis kemampuan kedua model dalam mengenali karakteristik genre musik serta mengevaluasi model yang memberikan performa terbaik berdasarkan hasil pengujian yang dilakukan.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan, baik dari segi data, metode penelitian, maupun hasil yang diperoleh. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun sebagai bahan evaluasi dan pengembangan penelitian di masa mendatang.

Akhir kata, penulis berharap skripsi ini dapat memberikan manfaat serta menjadi referensi bagi penelitian selanjutnya, khususnya dalam bidang kecerdasan buatan, *machine learning*, *deep learning*, dan pengolahan audio digital.

Palangka Raya, 10 Juni 2026



SALOMO JULIO ELSADA LAUTT

223020503080

PERBANDINGAN CNN DAN VISION TRANSFORMER (ViT) PADA KLASIFIKASI GENRE MUSIK MENGGUNAKAN LOG-MEL SPECTROGRAM

Salomo Julio Elsada Lauth (223020503080)

Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya, Kampus
Tanjung Nyaho Jl. Yos Sudarso, Palangka Raya 73112

Email : salomojulio0@gmail.com

ABSTRAK

Perkembangan teknologi digital dan kecerdasan buatan telah meningkatkan kompleksitas musik digital sehingga kebutuhan terhadap sistem klasifikasi genre musik yang akurat menjadi semakin penting. Dalam bidang *Music Information Retrieval* (MIR), *log-Mel spectrogram* banyak digunakan karena mampu merepresentasikan informasi waktu dan frekuensi sesuai karakteristik pendengaran manusia. Penelitian ini bertujuan membandingkan performa *Convolutional Neural Network* (CNN) berbasis ResNet-101 dan *Vision Transformer* (ViT-L/16) dalam klasifikasi genre musik menggunakan *log-Mel spectrogram* pada dataset GTZAN, Free Music Archive (FMA), dan MagnaTagATune (MTAT) dengan lima genre musik, yaitu *classical*, *hiphop*, *jazz*, *pop*, dan *rock*.

Tahapan penelitian meliputi *balancing* Dataset, *splitting* Dataset, segmentasi audio 5 detik dan 10 detik, ekstraksi fitur *log-Mel spectrogram*, serta pengecekan *data leakage*. Proses pelatihan menggunakan pendekatan *transfer learning* dan *fine-tuning*, sedangkan evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa kedua model mampu melakukan klasifikasi genre musik dengan baik, di mana ResNet-101 lebih stabil dalam mengekstraksi fitur lokal, sedangkan ViT-L/16 lebih efektif dalam memahami hubungan global dan konteks temporal pada data yang lebih kompleks.

Kata Kunci : Klasifikasi Genre Musik, *Log-Mel spectrogram*, ResNet-101, *Vision Transformer*, Deep Learning.

**PERBANDINGAN CNN DAN VISION TRANSFORMER (ViT) PADA
KLASIFIKASI GENRE MUSIK MENGGUNAKAN LOG-MEL
SPECTROGRAM**

Salomo Julio Elsada Lauth (223020503080)

*Department Of Informatics Engineering, Faculty Of Engineering, University Of
Palangka Raya, Tanjung Nyaho Campus, St. Yos Sudarso, Palangka Raya 73112*

Email : salomojulio0@gmail.com

ABSTRACT

The development of digital technology and artificial intelligence has increased the complexity of digital music, making accurate music genre classification systems increasingly important. In the field of Music Information Retrieval (MIR), log-Mel spectrogram is widely used because it effectively represents time and frequency information according to human auditory perception. This study aims to compare the performance of a Convolutional Neural Network (CNN) based on ResNet-101 and a Vision Transformer (ViT-L/16) for music genre classification using log-Mel spectrogram on the GTZAN, Free Music Archive (FMA), and MagnaTagATune (MTAT) Datasets with five music genres: classical, hip-hop, jazz, pop, and rock.

The research stages include Dataset balancing, Dataset splitting, 5-second and 10-second audio segmentation, log-Mel spectrogram feature extraction, and data leakage checking. The models were trained using transfer learning and fine-tuning approaches, while evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results show that both models achieved good performance in music genre classification, where ResNet-101 was more stable in extracting local features, while ViT-L/16 was more effective in capturing global relationships and temporal context on more complex data.

Keywords : *Music Genre Classification, Log-Mel spectrogram, ResNet-101, Vision Transformer, Deep Learning.*

DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN PERNYATAAN.....	ii
RIWAYAT PENYUSUN.....	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
1.6 Sistematika Penulisan.....	7
1.7 Jadwal Penulisan	8
BAB II LANDASAN TEORI	9
2.1 Tinjauan Pustaka.....	9
2.2 Musik Digital dan Audio Signal.....	15
2.3 Music Genre Classification	15
2.4 Feature Extraction Audio.....	16
2.4.1 Spectrogram.....	16

2.4.2 Mel-Spectrogram	16
2.4.3 Log Mel Spectrogram	17
2.5 <i>Deep Learning</i>	17
2.6 ResNet-101	18
2.7 ViT-L/16	18
2.8 Evaluasi Model Klasifikasi	19
2.8.1 <i>Accuracy</i>	19
2.8.2 <i>Precision</i>	20
2.8.3 <i>Recall</i>	20
2.8.4 <i>F1-score</i>	20
2.8.5 <i>Confusion Matrix</i>	20
2.9 Uji Statistik.....	21
BAB III METODOLOGI PENELITIAN	24
3.1 Kerangka Umum Penelitian	24
3.2 Pengumpulan Dataset	25
3.2.1 Dataset GTZAN.....	26
3.2.2 Dataset <i>Free Music Archive</i> (FMA-Medium).....	26
3.2.3 Dataset MagnaTagATune (MTAT).....	27
3.2.4 Seleksi Genre Musik.....	27
3.3 <i>Preprocessing</i> Dataset	28
3.3.1 <i>Balancing Data Audio</i>	29
3.3.2 <i>Splitting</i> Dataset (<i>Train, Val, Test</i>)	31
3.3.3 Segmentasi 5 Detik dan 10 Detik	33
3.3.4 Ekstraksi Fitur <i>Log-Mel Spectrogram</i>	36
3.3.5 Pengecekan Data <i>Leakage</i>	40

3.4 Pelatihan Model.....	42
3.4.1 Pelatihan Model CNN (ResNet-101).....	43
3.4.2 Pelatihan Model ViT (ViT-L/16).....	52
3.5 Evaluasi Model.....	60
3.5.1 Evaluasi Per Kelas	61
3.5.2 Uji Statistik	61
3.5.3 Analisis Hasil	62
BAB IV HASIL DAN PEMBAHASAN.....	63
4.1 Evaluasi Per Kelas	63
4.2 Uji Statistik.....	67
4.3 Analisis Hasil.....	69
4.4 Visualisasi Hasil	71
BAB V KESIMPULAN DAN SARAN	77
5.1 Kesimpulan.....	77
5.2 Saran.....	77
DAFTAR PUSTAKA.....	79
LAMPIRAN.....	84

DAFTAR GAMBAR

Gambar 3. 1 Kerangka Umum Penelitian	25
Gambar 3. 2 Grafik Class Distribution	28
Gambar 3. 3 Implementasi Splitting Dataset	32
Gambar 3. 4 Implementasi Segmentasi Audio	34
Gambar 3. 5 Implementasi Segmentasi Audio (2)	35
Gambar 3. 6 Implementasi Ekstraksi Log-Mel spectrogram (1)	38
Gambar 3. 7 Implementasi Ekstraksi Log-Mel spectrogram (2)	38
Gambar 3. 8 Implementasi Pengecekan Data Leakage Setelah Splitting	41
Gambar 3. 9 Implementasi Pengecekan Data Leakage Setelah Segmentasi Audio	41
Gambar 3. 10 Implementasi Pengecekan Data Leakage Setelah Ekstraksi Fitur Log Mel Spectrogram	41
Gambar 3. 11 Implementasi Data Generator ResNet-101.....	48
Gambar 3. 12 Implementasi Arsitektur ResNet-101	49
Gambar 3. 13 Implementasi Head Training	50
Gambar 3. 14 Implementasi Fine Tuning.....	50
Gambar 3. 15 Implementasi Callback Pelatihan Head Training	51
Gambar 3. 16 Implementasi Callback Pelatihan Fine Tuning.....	51
Gambar 3. 17 Hasil Pelatihan Model ResNet-101	52
Gambar 3. 18 Implementasi DataLoader ViT-L/16.....	56
Gambar 3. 19 Implementasi Model ViT-L/16 Pretrained (1).....	57
Gambar 3. 20 Implementasi Model ViT-L/16 Pretrained (2).....	57
Gambar 3. 21 Implementasi Freeze dan Unfreeze Encoder ViT-L/16	58
Gambar 3. 22 Implementasi Callback Early Stopping.....	58
Gambar 3. 23 Implementasi Callback Head Training.....	59
Gambar 3. 24 Implementasi Callback Fine Tuning	59
Gambar 3. 25 Penyimpanan Model dan Hasil Evaluasi ViT-L/16.....	60
Gambar 4. 1 Confusion Matrix ResNet-101 5s.....	71
Gambar 4. 2 Confusion Matrix ResNet-101 10s.....	72
Gambar 4. 3 Confusion Matrix ViT-L/16 5s	73

Gambar 4. 4 Confusion Matrix ViT-L/16 10s	75
---	----

DAFTAR TABEL

Tabel 1. 1 Jadwal Penulisan	8
Tabel 3. 1 Dataset Sebelum Balancing.....	30
Tabel 3. 2 Hasil Balancing Dataset Gabungan.....	30
Tabel 3. 3 Hasil Splitting Dataset.....	32
Tabel 3. 4 Hasil Segmentasi Audio 5 Detik	35
Tabel 3. 5 Hasil Segmentasi Audio 10 Detik	36
Tabel 3. 6 Hasil Ekstraksi Fitur Log-Mel spectrogram 5 Detik.....	39
Tabel 3. 7 Hasil Ekstraksi Fitur Log-Mel spectrogram 10 Detik.....	39
Tabel 3. 8 Hasil Pengecekan Data Leakage	42
Tabel 4. 1 Metrik ResNet-101 5s	63
Tabel 4. 2 Metrik ResNet-101 10s	64
Tabel 4. 3 Metrik ViT-L/16 5s.....	65
Tabel 4. 4 Metrik ViT-L/16 10s.....	66
Tabel 4. 5 Paired T-Test Segmentasi 5 Detik	67
Tabel 4. 6 Paired T-Test Segmentasi 10 Detik	68

DAFTAR PUSTAKA

- Adhiyatmaka, I. A., & Kurnia, L. (2024). *Collective Memory , Marginality , and Spatial Politics in.*
- Agustina, A. P., & Subroto, I. M. I. (2026). *Penerapan Convolutional Neural Network Dalam Klasifikasi Genre Musik Pop Islami.* 9(1), 206–214.
- Asanah, N., Pratama, I., Studi, P., Informasi, S., Informasi, F. T., & Buana, U. M. (2025). *Deep Learning Approach for Music Genre Classification using Multi - Feature Audio Representations.* 14, 2045–2054.
- Borovčák, K. (2025). *Instrument Classification in Musical Audio Signals using Deep Learning.* 16(1), 84–99. <https://doi.org/10.2478/crdj-2025-0006>
- Bouckaert, R. R., & Frank, E. (n.d.). *Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms.* 3–4.
- Chen, J., Ma, X., Li, S., Ma, S., Zhang, Z., & Ma, X. (2024). A Hybrid Parallel Computing Architecture Based on CNN and Transformer for Music Genre Classification. *Electronics (Switzerland)*, 13(16), 1–13. <https://doi.org/10.3390/electronics13163313>
- Costa, Y. M. G., Oliveira, L. S., & Silla, C. N. (2017). An evaluation of Convolutional Neural Networks for music classification using spectrograms. *Applied Soft Computing Journal*, 52, 28–38. <https://doi.org/10.1016/j.asoc.2016.12.024>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale. *ICLR 2021 - 9th International Conference on Learning Representations.*
- Gong, Y., Chung, Y., & Glass, J. (2021). *AST : Audio Spectrogram Transformer.*
- Hayes, B., Shier, J., Fazekas, G., Mcpherson, A., & Saitis, C. (2024). *A review of*

- differentiable digital signal processing for music and speech synthesis.*
January, 1–29. <https://doi.org/10.3389/frsip.2023.1284100>
- He, K. (2016). *Deep Residual Learning for Image Recognition.*
- Huang, C. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., & Hoffman, M. D. (2018). *G Enerating Music With Long - Term Structure.* 1–14.
- Iqbal, A., Zim, A. H., Tonmoy, M. A., Zhou, L., Malik, A., & Kuribayashi, M. (2024). EAViT: External Attention *Vision Transformer* for Audio Classification. *APSIPA ASC 2024 - Asia Pacific Signal and Information Processing Association Annual Summit and Conference 2024.*
<https://doi.org/10.1109/APSIPAASC63619.2025.10849238>
- Karpiński, B., & Leszczyński, C. (2025). *Comparison of spectrogram scaling in multi-label Music Genre.*
- Lee, J. (2024). *A Short Survey and Comparison of CNN-Based Music Genre Classification Using Multiple Spectral Features.* December 2023, 245–257.
<https://doi.org/10.1109/ACCESS.2023.3346883>
- Li, J., Han, L., Li, X., Zhu, J., Yuan, B., & Gou, Z. (2021). An evaluation of deep neural network models for music classification using spectrograms.
Multimedia Tools and Applications.
- Lim, K. M., Lee, C. P., Lee, Z. Y., & Alqahtani, A. (2023). *EnViTSA : Ensemble of Vision Transformer with SpecAugment for Acoustic Event Classification.* 1–18.
- Liptáková, E. (2021). *T-test for two dependent samples – some practical notes.* 7(1), 48–55.
- Mahdi, H. (2024). *. : Analysis of Audio Classifier Performance in.* 1–15.
- Mala, C., Studi, P., Perpustakaan, S.-I., Budaya, F. I., Diponegoro, U., Soedarto, J. P., & Tembalang, K. U. (2017). *Digitalisasi Produk Rekaman Studio*

Lokananta Suara Di Studio Lokananta.

- Maurício, J., Domingues, I., & Bernardino, J. (2023). Comparing *Vision Transformers* and Convolutional Neural Networks for Image Classification: A Literature Review. *Applied Sciences (Switzerland)*, 13(9).
<https://doi.org/10.3390/app13095521>
- McLoughlin, I., Pham, L., Song, Y., Miao, X., Phan, H., Cai, P., Gu, Q., Nan, J., Song, H., & Soh, D. (2026). *Spectrogram Features for Audio and Speech Analysis*. 1–29.
- Mehta, A., Chauhan, S., & Choudhury, M. (2025). *Missing Melodies : AI Music Generation and its " Nearly " Complete Omission of the Global South*. 1(1).
- Mehta, J., Gandhi, D., Thakur, G., & Kanani, P. (2021). *Music Genre Classification using Transfer learning on log-based MEL Spectrogram*. *Iccmc*, 1101–1107.
- Mou, A., Milanova, M., & Baillie, M. (2023). *Deep Learning Approaches for Classroom Audio Classification Using Mel Spectrograms*.
- Mycka, J., & Man, J. (2025). *Artificial Intelligence In Music : Recent Trends And Challenges* (Vol. 0123456789). <https://doi.org/10.1007/s00521-024-10555-x>
- Panigrahy, A. (2023). *Music genres classification using vision transformers*
Aparna Panigrahy , Assistant Professor, Department of Information Technology, Nirmala Memorial Foundation College of Commerce and Science Kandivali East, Mumbai. 1, 1–9.
- Parajuli, S., Karna, A., & Dhakal, O. P. (2026). *Music Genre Classification : A Comparative Analysis of Classical Machine Learning and Deep Learning Approaches Using CRNN*.
- Prakash, O., Prasad, L., & Sharma, J. (2021). *Speech Emotion Recognition using Convolutional Recurrent Neural Network*. 8914, 1164–1172.
- Purnama, I. N. (2022). *Music Genre Recommendations Based on Spectrogram*

Analysis Using Convolutional Neural Network Algorithm with RESNET-50 and VGG-16 Architecture. 05(01), 69–74.

Radford, A., Wook, J., Chris, K., Aditya, H., Gabriel, R., Sandhini, G., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2019). *Learning Transferable Visual Models From Natural Language Supervision.*

Rinaldi, R., Tata, J., Seni, K., & Rupa, F. S. (2019). *UPT Perpustakaan ISI Yogyakarta.*

ROSMALA, D., & FADHILAH, M. N. (2025). Audio Conversion for Music Genre Classification Using Short-Time Fourier Transform and Inception V3. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika, 13(1), 84.* <https://doi.org/10.26760/elkomika.v13i1.84>

Soerkarta, R., Aras, S., & Aswad, A. N. (2023). Hyperparameter Optimization of CNN Classifier. *Jurnal Rekayasa Sistem Dan Teknologi Informasi, 5(158), 1205–1210.*

Sridhar, A. (2024). *Attention-guided Spectrogram Sequence Modeling with CNNs for Music Genre Classification.* <http://arxiv.org/abs/2411.14474>

Valerio, V. D., Pereira, R. M., Costa, Y. M. G., Bertolini, D., & Jr, C. N. S. (2018). A Resampling Approach for Imbalanceness on Music Genre Classification Using Spectrograms. *The Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31), 500–505.*

Wairata, C. R., Swedia, E. R., & Cahyanti, M. (2021). Pengklasifikasian Genre Musik Indonesia Menggunakan Convolutional Neural Network. *Sebatik, 25(1), 255–261.* <https://doi.org/10.46984/sebatik.v25i1.1286>

Wu, P., Gao, W., Chen, Y., Xu, F., Ji, Y., Tu, J., & Lin, H. (2025). An improved ViT model for music genre classification based on mel spectrogram. *PLoS ONE, 20(3 March), 1–12.* <https://doi.org/10.1371/journal.pone.0319027>

Zaman, K., Sah, M., Direkoglu, C., & Unoki, M. (2023). A Survey of Audio

Classification Using Deep Learning. *IEEE Access*, 11(September), 106620–106649. <https://doi.org/10.1109/ACCESS.2023.3318015>

Zhang, J. (2023). *Music Genre Classification with ResNet and Bi-GRU Using Visual Spectrograms*. 1–17.