

Pendekatan *Transfer Learning* Untuk Klasifikasi Tangisan Bayi Dengan Imbalanced Dataset

Theresia Herlina Rochadiani¹

¹Universitas Pradita

Informasi Artikel

Diterima : 19 Mar 2024

Direview : 23 Mar 2024

Disetujui : 1 Apr 2024

Kata Kunci

imbalanced dataset,
klasifikasi, tangisan bayi,
transfer learning,
YAMNet

Abstrak

Klasifikasi tangisan bayi dapat dimanfaatkan untuk mengidentifikasi masalah kesehatan bayi dan memenuhi kebutuhan bayi dengan cepat. Dalam studi ini, teknik *transfer learning*, dengan model terlatih YAMNet, diterapkan untuk klasifikasi bayi dengan dataset terbatas dan tidak seimbang. YAMNet, sebuah model Convolutional Neural Network khusus untuk analisis audio, mengatasi keterbatasan metode tradisional yang bergantung pada interpretasi manusia. Dengan mempelajari fitur-fitur audio secara otomatis, memungkinkan kinerja klasifikasi yang lebih akurat. Dalam studi ini, dilakukan eksplorasi dan analisis manfaat penggunaan YAMNet, melalui perbandingan dengan model baseline tanpa teknik *transfer learning*. Hasilnya menunjukkan bahwa model YAMNet tidak hanya nilai akurasinya yang tinggi 0.8106, namun juga nilai skor-F1nya tinggi yaitu mencapai 0.9831. Terbukti bahwa penggunaan *transfer learning* dapat meningkatkan kinerja dalam klasifikasi tangisan bayi, terutama dalam mengatasi ketidakseimbangan data dan meningkatkan prediksi untuk kelas minoritas.

Keywords

baby crying, classification, imbalanced dataset, transfer learning, YAMNet

Abstract

The classification of baby cries can be used to identify baby health problems and meet the baby's needs quickly. In this study, the transfer learning technique, with a YAMNet pre-trained model, is applied for infant classification with a limited and imbalanced dataset. YAMNet, a Convolutional Neural Network model specifically for audio analysis, overcomes the limitations of traditional methods that rely on human interpretation. By automatically learning audio features, it enables more accurate classification performance. In this study, exploration and analysis of the benefits of using YAMNet were carried out, through comparison with a baseline model without transfer learning techniques. The results show that the YAMNet model not only has a high accuracy value of 0.8106, but also a high F1-score value of 0.9831. It is proven that the use of transfer learning can improve performance in classification of baby cries, especially in overcoming imbalanced data and improving predictions for minority classes.

A. Pendahuluan

Klasifikasi tangisan bayi telah lama menjadi topik yang menarik di berbagai sektor, termasuk di bidang kesehatan, seperti untuk memantau bayi di inkubator [1] dan juga topik dalam pembelajaran mesin [2], [3]. Kemampuan membedakan berbagai jenis jeritan atau tangisan mempunyai potensi besar untuk mendeteksi masalah kesehatan secara dini dan memenuhi kebutuhan bayi dengan cepat. Bahkan dengan kemajuan kecerdasan buatan, teknologi rekognisi suara tangisan bayi dimanfaatkan dalam sistem keamanan rumah untuk mengawasi bayi [4] dan menjadi alat bantu orang tua menjaga bayi mereka [5]. Dalam beberapa tahun terakhir, kemajuan dalam algoritma pembelajaran mendalam telah membuka jalur baru untuk klasifikasi tangisan bayi otomatis, memberikan solusi yang efisien dan dapat diandalkan. Salah satu metode yang mendapatkan popularitas adalah penggunaan teknik *transfer learning* dengan menggunakan model terlatih seperti YAMNet. Teknik *transfer learning* dapat menjadi solusi apabila memiliki dataset dalam jumlah terbatas.

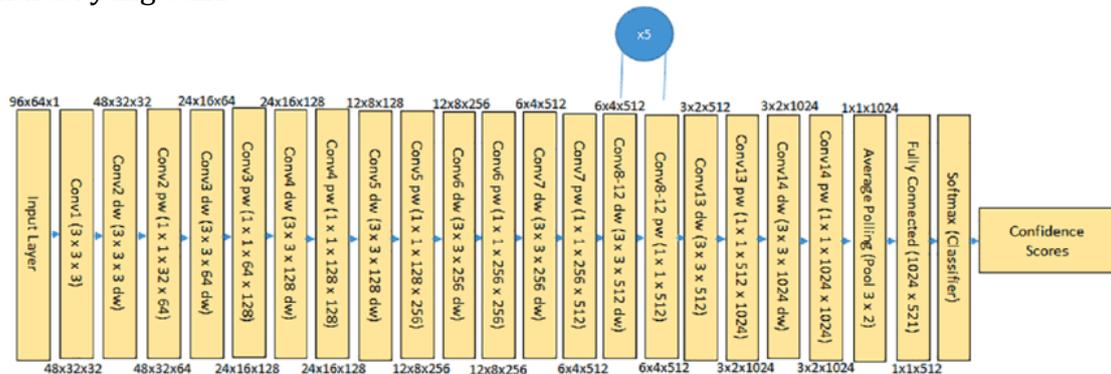
Tangisan bayi berfungsi sebagai cara komunikasi utama [6], menyampaikan berbagai informasi tentang kesehatan dan kebutuhan mendesak mereka. Tangisan yang berbeda dapat menandakan rasa lapar, nyeri, ketidaknyamanan [7], atau bahkan gangguan medis seperti kolik [8] atau gangguan pernapasan [9]. Metode analisis tangisan tradisional terutama bergantung pada interpretasi subjektif oleh perawat atau pakar kesehatan, yang dapat menyebabkan ketidakakuratan dan inkonsistensi. Menggabungkan metode pembelajaran mesin, khususnya model pembelajaran mendalam seperti YAMNet, akan memberikan metode yang lebih obyektif dan sistematis untuk mengklasifikasikan dan menafsirkan tangisan bayi. YAMNet menggunakan *deep neural network* untuk menganalisis data audio dan memiliki kemampuan untuk mengidentifikasi suara.

YAMNet, atau *Yet Another Multi-scale Convolutional Neural Network*, adalah model pembelajaran mendalam yang dibuat oleh Google, khusus untuk menganalisis sinyal audio. YAMNet, yang dilatih pada kumpulan data besar, yaitu dataset Audio [10], yang mencakup 2.084.320 klip suara berdurasi 10 detik dari video YouTube, dan dibagi ke dalam 632 kelas [11], sehingga memiliki kinerja yang sangat baik dalam tugas klasifikasi suara. Dengan menggunakan *transfer learning*, studi klasifikasi tangisan bayi dapat menggunakan model YAMNet yang telah dilatih sebelumnya dengan hasil yang lebih akurat, khususnya untuk mengatasi keterbatasan data berlabel dan sumber daya komputasi.

Salah satu keuntungan utama menggunakan YAMNet untuk klasifikasi tangisan bayi adalah kemampuannya mendeteksi sinyal akustik kecil yang menunjukkan jenis tangisan yang berbeda. Tidak seperti metode tradisional yang terutama bergantung pada analisis spektrogram atau penentuan fitur yang dilakukan oleh manusia, arsitektur jaringan saraf dalam YAMNet secara otomatis mempelajari fitur-fitur yang relevan dari data audio mentah, memungkinkan klasifikasi yang lebih baik. Fitur bawaan ini memungkinkan YAMNet menangani variasi pola tangisan pada bayi secara tepat.

YAMNet menggunakan desain *deep neural network* yang sebagian besar terdiri dari lapisan konvolusional, diikuti oleh *global average pooling* dan *classification layer*, seperti yang ditunjukkan Gambar 1. Arsitektur tersebut bertujuan untuk merekam representasi hierarki properti audio pada berbagai

skala, memungkinkan klasifikasi akurat dari berbagai macam suara. YAMNet dilatih menggunakan kumpulan data yang besar dan beragam yang mencakup jutaan sampel audio dari ribuan kategori suara. Data pelatihan yang substansial ini memungkinkan model memperoleh karakteristik yang kuat dan diskriminatif yang dapat diterapkan pada input audio yang sebelumnya tidak diketahui dengan akurasi yang baik.



Gambar 1. Arsitektur YAMNet [12]

YAMNet dapat membuat komputasi menjadi lebih efisien. YAMNet juga memungkinkan inferensi real-time atau hampir real-time pada berbagai perangkat, termasuk ponsel pintar dan sistem tertanam. Efisiensi ini menjadikannya ideal untuk digunakan dalam konteks sumber daya yang terbatas atau aplikasi yang memerlukan pemrosesan latensi rendah. YAMNet dibangun berdasarkan kerangka pembelajaran mendalam TensorFlow, sehingga tersedia bagi komunitas besar pengembang dan akademisi. Ketersediaan model terlatih YAMNet dan API TensorFlow memudahkan integrasi YAMNet ke dalam aplikasi.

Beberapa studi telah menggunakan pembelajaran mesin untuk mengklasifikasikan tangisan bayi. Beberapa studi menggunakan algoritma KNN [2], [3], SVM [2], *Linear Discriminant Analysis (LDA)*, *Principal Component Analysis (PCA)* [13]. Studi lain bereksperimen menggunakan teknik ensemble dalam mengklasifikasikan tangisan bayi, dan hasilnya sekitar 48% lebih baik dibandingkan pengklasifikasi individu dengan kinerja paling buruk, dan sekitar 20% lebih baik dibandingkan pengklasifikasi individu dengan kinerja terbaik [14].

Adapun metode pembelajaran mendalam juga telah digunakan dalam beberapa studi untuk mengkategorikan tangisan bayi. Algoritma ANN, CNN, dan LSTM digunakan dalam studi [15] dan diperoleh hasil CNN dan LSTM berkinerja baik dalam membedakan antara bayi baru lahir yang sehat dan tidak sehat, dengan tingkat akurasi, presisi, dan recall sekitar 95%. CNN mengungguli LSTM dan ANN dalam mengenali kebutuhan bayi baru lahir, dan mencapai akurasi hingga 60%. Studi lain menggunakan CNN-RNN [16]. Model dievaluasi dengan dataset Dunstan Baby Language, model tersebut mengungguli metode sebelumnya dengan akurasi klasifikasi rata-rata sebesar 94,97%.

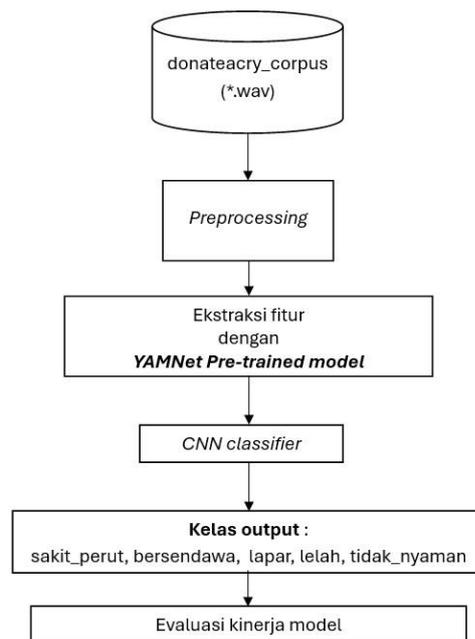
Studi lain mendiskusikan mengenai teknik *transfer learning*. Studi mengenai perbandingan kinerja metode *transfer learning* pada 2 model network, yaitu VGG16 dan ResNet 50. Hasilnya menunjukkan bahwa kinerja yang lebih baik diperoleh melalui pelatihan model dengan *trainable layer* yang paling sedikit [17]. Sedangkan studi [18] memanfaatkan *transfer learning* untuk klasifikasi gambar

burung, dengan menggunakan 3 model yaitu Big Transfer (BiT), DenseNet121, dan VGG16 Performa terbaik dari model yang dibangun dicapai oleh model BiT, yang memiliki akurasi 96,53%. Model DenseNet121 dan VGG16, yang dibangun, memiliki akurasi lebih rendah, masing-masing 92,36% dan 81,94%.

Dalam studi ini klasifikasi tangisan bayi menggunakan dataset *donateacry_corpus* [19] yang terdiri dari 5 kelas dengan jumlah data tiap kelas terbatas dan berbeda-beda. Karena keterbatasan dataset inilah maka pada studi ini digunakan *transfer learning* dengan menggunakan *pre-trained model* YAMNet. Melalui studi ini dapat dilihat bagaimana dengan menggunakan model terlatih YAMNet dapat meningkatkan kinerja klasifikasi tangisan bayi pada dataset yang terbatas dan imbalance.

B. Metode Penelitian

Tahapan dalam studi ini mengikuti metodologi yang ditunjukkan oleh Gambar 2 berikut. Dimulai dari dataset audio berupa file berekstensi *.wav* yang akan dilakukan proses *preprocessing* terlebih dahulu sebelum masuk ke dalam model YAMNet. Ketika masuk ke dalam model YAMNet, fitur penting dari audio tangisan bayi akan diekstrak dan diteruskan ke model CNN. Dari model tersebut akan diklasifikasikan ke dalam 5 kelas, yaitu sakit perut, bersendawa, lapar, lelah, dan tidak nyaman.



Gambar 2. Metodologi dalam studi ini

1. Dataset

Dataset *donateacry_corpus* terdiri dari 457 audio file berekstensi *wav* yang memiliki *sampling rate* 16 kHz dan durasi sekitar 7 detik tiap klip audio. Seperti yang ditunjukkan pada Tabel 1, setiap file audio dikategorikan ke dalam 5 kelas. Dari tabel terlihat distribusi file tiap kelas tidak sama, memperlihatkan *imbalance*

dataset, dengan jumlah tertinggi filenya adalah 382 untuk kelas lapar dan jumlah file terendahnya adalah 8 file untuk kelas bersendawa.

Tabel 1. Distribusi File Tiap Kelas Dataset

Kategori	Jumlah File
Lapar	382
Tidak nyaman	27
Lelah	24
Sakit perut	16
Bersendawa	8

Jumlah data yang terbatas dan imbalance dataset tersebut dapat menjadi kendala dalam menghasilkan model yang mumpuni untuk klasifikasi tangisan bayi. Sehingga dalam studi ini dilakukan pemodelan dengan teknik *transfer learning* dalam mengatasi hal tersebut.

2. Preprocessing

Data audio yang berekstensi wav akan diubah ke dalam bentuk tensor float. Hal ini dilakukan karena data ini akan menjadi input pada model YAMNet yang dibangun dengan framework Tensorflow. Selain proses konversi ke tensor float, audio file dipastikan memiliki *sampling rate* 16 kHz dan merupakan *mono channel*.

Proses pemetaan kelas menjadi data numerik juga dilakukan pada tahapan ini. Kemudian file audio yang telah dikonversi ke dalam tensor float beserta kelasnya akan disimpan dalam dataset tensorflow. Dataset ini yang nantinya akan dibagi menjadi 2, yaitu data latih dan data uji dengan proposional 80:20.

3. CNN classifier

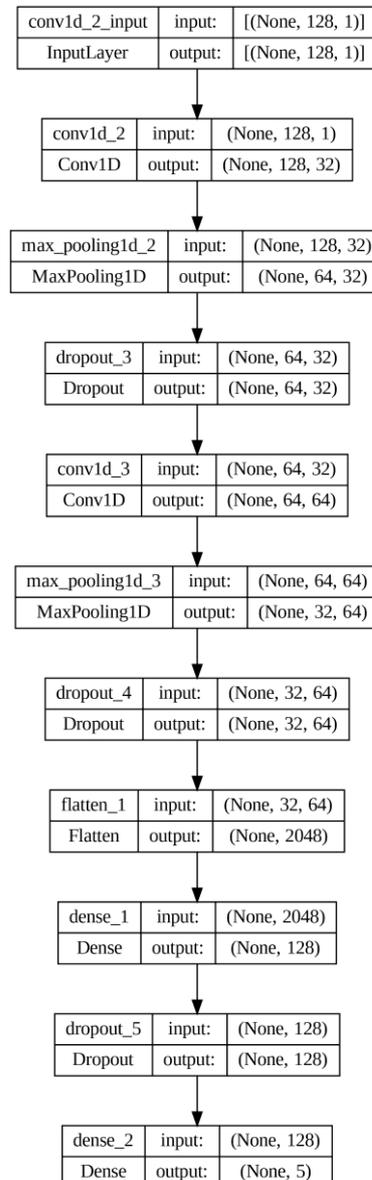
Dalam studi ini akan digunakan model CNN sebagai model baseline dan model *classifier* untuk *downstream task*. Kinerja dari model baseline dan model yang menggunakan teknik *transfer learning* dengan model terlatih YAMNet akan dibandingkan sebagai analisis pengaruh teknik *transfer learning* terhadap klasifikasi tangisan bayi yang memiliki *imbalance* dan keterbatasan jumlah data.

Model CNN yang akan digunakan dalam studi ini menggunakan 1D convolutional layer yang memiliki arsitektur seperti yang ditunjukkan pada Gambar 3. Model ini terdiri dari 2 layer utama yaitu *feature extraction layer* dan *fully connected layer (FC layer)*.

Sesuai namanya, *feature extraction layer* berfungsi untuk mengekstrak fitur audio yang merepresentasikan audio tersebut. Layer ini terdiri dari *convolutional layer* dan *pooling layer*. Pada gambar terlihat ada 2 *convolutional layer* yang tiap layernya diikuti *max pooling layer*.

Fully connected layer berfungsi sebagai *classifier*. Pada model ini FC layer terdiri dari 1 hidden layer dan 1 output layer. *Dropout layer* dapat digunakan untuk meningkatkan kinerja model dengan mencegah adanya *overfitting*. Dengan menambahkan *dropout layer*, neuron dipilih secara acak dan mengubah nilai neuron tersebut menjadi nol. Neuron yang diubah nilainya menjadi nol ini tidak akan dipakai atau dibuang selama pelatihan. Pada model ini, dropout pada *feature extraction layer* diberi nilai 0.25 dan pada FC layer diberi nilai 0.5, yang artinya 25% dan 50% neuron, secara berurutan, akan dibuang.

Sebagai algoritma optimisasi digunakan algoritma Adam, sedangkan *sparse categorical cross entropy* dipilih sebagai fungsi loss, dimana setiap sampel hanya memiliki 1 label atau kelas. Untuk fungsi aktivasi dalam model digunakan relu dan untuk lapisan Dense terakhir menggunakan softmax sebagai fungsi aktivasi yang akan menghasilkan probabilitas untuk setiap kelas output.



Gambar 3. Model CNN untuk klasifikasi tangisan bayi

4. Evaluasi Kinerja

Untuk membandingkan kinerja model baseline dan model dengan menggunakan YAMNet sebagai *feature extractor*, akan dilakukan evaluasi menggunakan 4 metrik yaitu skor F1, presisi, recall, dan akurasi. Selain itu, akan dibandingkan pula kurva loss dan akurasi untuk data training dan data validasi dalam membandingkan kinerja 2 model dalam studi ini.

C. Hasil dan Pembahasan

Untuk mengetahui efektivitas *transfer learning* dalam meningkatkan kinerja model klasifikasi tangisan bayi pada studi ini, maka akan dibandingkan 2 model. Model pertama, yang disebut sebagai model baseline, dimana model ini tidak menggunakan teknik *transfer learning*. Model baseline hanya berupa model CNN yang arsitekturnya ditunjukkan pada Gambar 3. Sedangkan model kedua adalah model yang menggunakan *transfer learning*, dimana YAMNet digunakan sebagai *feature extractor*, yang hasilnya diteruskan sebagai input dari model CNN yang sama dengan model baseline.

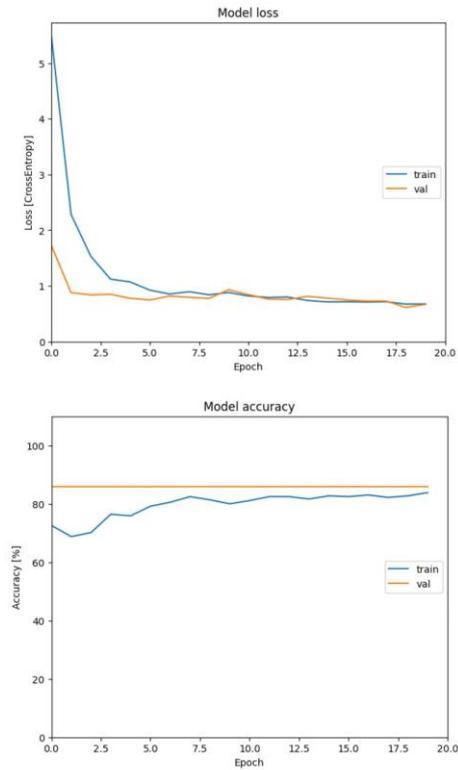
Proses pelatihan yang dilakukan dalam 20 epoch dari 2 model menghasilkan kurva loss dan akurasi untuk setiap model yang ditunjukkan Gambar 4 dan Gambar 5. Untuk model baseline, loss data training dan data validasi mulai konvergen setelah epoch ke-5 dengan nilai loss kurang lebih 0.70. Untuk akurasi, dari awal, akurasi data validation stabil dengan nilai sekitar 0.86, sedangkan untuk data training mengalami fluktuatif di awal epoch namun cenderung meningkat akurasinya, dan fluktuatif semakin mengecil dan mencapai nilai akurasi sekitar 0.83 di akhir epoch, seperti yang ditunjukkan pada Gambar 4. Adanya gap antara kurva akurasi data training dan data validasi, dimana akurasi data training lebih rendah daripada data validasi, menunjukkan bahwa model mengalami *underfitting*.

Gambar 5 menunjukkan kurva loss dan akurasi dari model YAMNet. Terlihat dari kurva, bahwa loss dari data validasi cenderung lebih fluktuatif dibandingkan dengan data training. Loss data validasi rata-rata bernilai kurang dari 1, meskipun di epoch ke-19 mencapai nilai loss tertinggi, sekitar 1.16 Untuk data training, loss cenderung stabil di sekitar 0.50. Meskipun gap antara kurva loss data training dan data validasi terlihat cukup besar, namun perbedaan nilai lossnya kurang dari 0.50. Untuk akurasi model kedua, antara data training dan data validasi gapnya tidak signifikan. Hal ini menunjukkan tidak terjadi *overfitting* ataupun *underfitting*. Nilai akurasinya pun cukup baik, yaitu kurang lebih 0.83

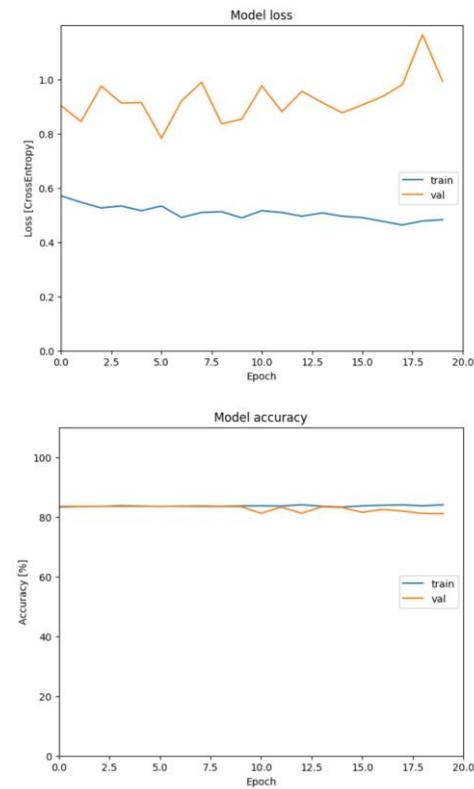
Dari perbandingan kurva loss dan akurasi model baseline dan model YAMNet, terlihat kinerja model YAMNet lebih baik meskipun nilai akurasinya lebih rendah sedikit daripada akurasi model baseline.

Selain dengan melihat kurva loss dan akurasi, kinerja model baseline dan model YAMNet dapat dilihat berdasarkan metrik yang ada pada Tabel 2. Nilai akurasi model baseline lebih tinggi, yaitu 0.8587, daripada nilai akurasi model YAMNet, yaitu 0.8106. Namun nilai skor-F1, presisi, dan recall model baseline jauh lebih rendah dibandingkan model YAMNet. Nilai skor-F1 baseline hanya mencapai 0.1848, sedangkan nilai skor-F1 YAMNet mencapai 0.9831. Hal ini disebabkan oleh ketidakseimbangan jumlah data. Seperti yang ditunjukkan pada Tabel 1, bahwa kelas lapar memiliki jumlah data yang jauh lebih banyak dibandingkan dengan kelas yang lain.

Dalam ketidakseimbangan kelas di mana satu kelas memiliki jumlah sampel yang jauh lebih besar daripada yang lain, tingkat akurasi yang tinggi tetapi tingkat presisi dan recall yang rendah dapat menunjukkan bahwa model cenderung melakukan prediksi yang benar terhadap kelas mayoritas, tetapi seringkali salah dalam mengidentifikasi kelas minoritas.



Gambar 4. Kurva loss dan akurasi model baseline



Gambar 5. Kurva loss dan akurasi model YAMNet

Tabel 2. Perbandingan Kinerja Model

Model	Metrik			
	Skor-F1	Presisi	Recall	Akurasi
Baseline	0.1848	0.1717	0.2000	0.8587
YAMNet	0.9831	0.9667	1.000	0.8106

Mengingat kembali definisi akurasi, presisi, dan recall Akurasi adalah ukuran seberapa banyak prediksi yang benar yang dibuat oleh model dibandingkan dengan jumlah total prediksi. Jika akurasi tinggi, tetapi presisi dan recall rendah, itu berarti model cenderung melakukan prediksi yang benar untuk kelas mayoritas (yang memiliki banyak sampel), tetapi dapat mengabaikan kelas minoritas. Seberapa banyak prediksi positif yang benar dibuat dibandingkan dengan total prediksi positif yang dilakukan oleh model disebut presisi, dan jika presisi rendah, itu berarti banyak prediksi positif yang sebenarnya salah. Recall adalah ukuran dari seberapa banyak prediksi positif yang benar dibuat dibandingkan dengan total jumlah kemungkinan sampel positif yang ada. Jika recall rendah, itu berarti model gagal menemukan sebagian besar kasus positif yang benar.

Dalam situasi seperti deteksi anomali, kelas minoritas (anomali) memiliki frekuensi yang jauh lebih rendah daripada kelas mayoritas (data normal). Karena kelas minoritas tidak dapat diprediksi dengan benar, model dapat memiliki akurasi yang tinggi tetapi kurang presisi dan recall karena kelas minoritas tidak dapat diprediksi dengan benar.

Ini menunjukkan bahwa, dalam kasus ketidakseimbangan kelas, penting untuk mempertimbangkan metrik evaluasi selain akurasi, seperti skor F1, presisi, atau recall. Meterik ini memberikan informasi lebih rinci tentang kinerja model, terutama untuk kelas yang kurang representatif.

D. Simpulan

Transfer learning dapat meningkatkan kinerja model klasifikasi tangisan bayi. Studi ini membandingkan dua model: model baseline dan model dengan *transfer learning*, YAMNet. Hasilnya, meskipun nilai akurasi dari model baseline lebih tinggi daripada model YAMNet, yaitu 0.8587 untuk model baseline, dan 0.8106 untuk model YAMNet, namun nilai skor-F1, presisi, dan recall model YAMNet jauh lebih tinggi daripada model baseline. Nilai skor-F1 YAMNet mencapai 0.9831, sedangkan nilai skor-F1 model baseline hanya 0.1848.

Dalam kasus ketidakseimbangan kelas di mana kelas minoritas memiliki frekuensi yang lebih rendah, skor-F1, presisi, dan recall model dengan *transfer learning*, yaitu YAMNet, lebih baik daripada model baseline. Hal ini menunjukkan bahwa meskipun nilai akurasi dapat menjadi indikator kinerja yang bermanfaat, penting untuk mempertimbangkan metrik evaluasi lainnya juga. Kesimpulannya, penggunaan *transfer learning* dapat meningkatkan kemampuan model untuk mengklasifikasikan tangisan bayi, terutama untuk mengatasi ketidakseimbangan data dan meningkatkan keakuratan prediksi untuk kelas minoritas.

E. Referensi

- [1] E. Sutanto, F. Fahmi, W. Shalannanda, and A. Aridarma, "Cry Recognition for Infant Incubator Monitoring System Based on Internet of Things using Machine Learning," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 1, pp. 444–454, 2021, doi: 10.22266/IJIES2021.0228.41.
- [2] D. Widhyanti and D. Juniati, "Classification of Baby Cry Sound Using Higuchi's Fractal Dimension with K-Nearest Neighbor and Support Vector Machine," *J. Phys. Conf. Ser.*, vol. 1747, no. 1, 2021, doi: 10.1088/1742-6596/1747/1/012014.
- [3] A. M. Mahmoud, S. Mohamed Swilem, A. S. Alqarni, and F. Haron, "Infant Cry Classification Using Semi-supervised K-Nearest Neighbor Approach," in *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, Dec. 2020, vol. 2020-Decem, pp. 305–310, doi: 10.1109/DeSE51703.2020.9450239.
- [4] J. Rathi, J. Rojanarowan, and N. Aphiratsakoun, "AI in Home Security : The Potential of Sound and Facial Recognition AI," *IEET-International Electr. Eng. Trans.*, vol. 9, no. 2, 2023.
- [5] K. Ravindran, M. Ibrahim, H. Lee, U. Qureshi, and K. A. Hafeez, "Design & Development of an AI-Powered Baby Monitoring System," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2021, pp. 1–6, doi: 10.1109/CCECE53047.2021.9569040.
- [6] C. Ji, T. B. Mudiyansele, Y. Gao, and Y. Pan, "A review of infant cry analysis and classification," *EURASIP J. Audio, Speech, Music Process.*, vol. 2021, no. 1, p. 8, Dec. 2021, doi: 10.1186/s13636-021-00197-5.
- [7] M. Viragova and S. O'Curry, "Understanding persistent crying in infancy," *Paediatr. Child Health (Oxford)*, vol. 31, no. 3, pp. 116–121, Mar. 2021, doi: 10.1016/j.paed.2020.12.004.
- [8] D. Ghio *et al.*, "Parents' concerns and understandings around excessive infant crying: Qualitative study of discussions in online forums," *SSM - Qual. Res. Heal.*, vol. 2, no. July, p. 100146, Dec. 2022, doi: 10.1016/j.ssmqr.2022.100146.
- [9] Z. Khalilzad, A. Hasasneh, and C. Tadj, "Newborn Cry-Based Diagnostic System to Distinguish between Sepsis and Respiratory Distress Syndrome Using Combined Acoustic Features," *Diagnostics*, vol. 12, no. 11, p. 2802, Nov. 2022, doi: 10.3390/diagnostics12112802.
- [10] S. Choudhary, C. R. Karthik, P. S. Lakshmi, and S. Kumar, "LEAN: Light and Efficient Audio Classification Network," in *2022 IEEE 19th India Council International Conference (INDICON)*, Nov. 2022, pp. 1–6, doi: 10.1109/INDICON56171.2022.10039921.
- [11] I. Kuzminykh, D. Shevchuk, S. Shiaeles, and B. Ghita, "Audio Interval Retrieval Using Convolutional Neural Networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12525 LNCS, 2020, pp. 229–240.
- [12] A. Tena, F. Claria, and F. Solsona, "Automated detection of COVID-19 cough," *Biomedical Signal Processing and Control*. Elsevier, 2022, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1746809421007722>.
- [13] L. Novamizanti, A. L. Prasasti, and B. S. Utama, "Study of Linear Discriminant

- Analysis to Identify Baby Cry Based on DWT and MFCC," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 982, no. 1, p. 012009, Dec. 2020, doi: 10.1088/1757-899X/982/1/012009.
- [14] H. A. Fahmy, S. F. Fahmy, A. A. Del Barrio García, and G. Botella Juan, "An ensemble multi-stream classifier for infant needs detection," *Heliyon*, vol. 9, no. 4, p. e15098, Apr. 2023, doi: 10.1016/j.heliyon.2023.e15098.
- [15] Y. C. Liang, I. Wijaya, M. T. Yang, J. R. Cuevas Juarez, and H. T. Chang, "Deep Learning for Infant Cry Recognition," *Int. J. Environ. Res. Public Health*, vol. 19, no. 10, 2022, doi: 10.3390/ijerph19106311.
- [16] T. Nadia Maghfira, T. Basaruddin, and A. Krisnadhi, "Infant cry classification using CNN - RNN," *J. Phys. Conf. Ser.*, vol. 1528, no. 1, 2020, doi: 10.1088/1742-6596/1528/1/012019.
- [17] F. Muhammad, A. M. Arimurthy, and D. Chahyati, "Transfer Learning dengan Metode Fine Tuning pada Model Network VGG16 dan ResNet50," *Indones. J. Comput. Sci. Attrib.*, vol. 12, no. 1, pp. 361–374, 2023.
- [18] M. F. A. Yusuf, C. D. Suhendra, and K. M. Auditama, "Klasifikasi Gambar Burung Konservasi di Wilayah Papua Barat Menggunakan Transfer Learning," *Indones. J. Comput. Sci.*, vol. 13, no. 1, pp. 1179–1193, 2024, [Online]. Available: <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>.
- [19] G. Veres, "donateacry-corporus: Infant cry audio corpus repository," 2019. <https://github.com/gveres/donateacry-corporus>.