

Sistem Klasifikasi Tangisan Bayi Menggunakan *Machine Learning*

1st Aisha Medina Arafat

Fakultas Teknik Elektro

Universitas Telkom

Bandung, Indonesia

aishaarafat@student.telkomuniversity.ac.id

2nd Wahmisari Priharti

Fakultas Teknik Elektro

Universitas Telkom

Bandung, Indonesia

wpriharti@telkomuniversity.ac.id

3rd Dien Rahmawati

Fakultas Teknik Elektro

Universitas Telkom

Bandung, Indonesia

dienrahmawati@telkomuniversity.ac.id

Abstrak— Salah satu permasalahan yang dihadapi oleh orang tua yang baru memiliki anak adalah sulitnya memahami maksud dari tangisan bayi mereka. Pengklasifikasian penyebab tangisan bayi sudah mulai diteliti dari tahun 1960 dengan menggunakan bantuan manusia (suster anak berpengalaman) untuk mengenali ‘sinyal’ penyebab tangisan bayi. Pada sebuah studi tahun 2006 oleh Dunstan, ditemukan pola refleks suara pada bayi yang dapat diartikan kedalam 5 penyebab tangisan, yaitu butuh bersendawa, tidak nyaman, lapar, sakit perut, dan lelah. Teori ini bernama Dunstan Baby Language (DBL). Dengan berkembangnya teknologi kecerdasan buatan berupa audio classification, hal ini dapat menjadi solusi untuk permasalahan orang tua baru yang belum memahami arti tangisan bayinya. Untuk itu, pada Tugas Akhir ini penulis merancang sebuah simulasi sistem klasifikasi penyebab tangisan bayi menggunakan Machine Learning dengan ekstraksi fitur MFCC (Mel-Frequency Cepstral Coefficients) dan pengklasifikasian ANN (Artificial Neural Network). Pembuatan dan pengujian sistem ini menggunakan sampel suara dari dataset penelitian sebelumnya yang berisi suara-suara tangisan bayi. Dengan model yang telah dibuat menggunakan *n_mfcc* 40, 1 hidden layer dengan dense 300, dihasilkan akurasi validasi saat training sebesar 65% dan F1 Score saat testing sebesar 49%.

Kata kunci— klasifikasi tangisan bayi, DBL, MFCC, ANN

I. PENDAHULUAN

Merawat bayi bukanlah sesuatu yang mudah dilakukan, maka dari itu ada banyak sekali ilmu khususnya. Walaupun demikian, ada beberapa hal yang lebih sulit untuk dipelajari oleh orangtua, yaitu komunikasi. Satu-satunya cara seorang bayi berkomunikasi dengan dunia luar ialah melalui tangisan, sementara tidak semua orangtua dapat memahami arti tangisan tersebut, terlebih orangtua yang baru saja memiliki bayi. Studi mengenai klasifikasi penyebab tangisan bayi dimulai pada tahun 1960 oleh kelompok riset Wasz-Hockert dengan bantuan suster yang berpengalaman dan menghasilkan 4 tipe tangisan bayi; sakit, lapar, saat baru lahir, dan kesenangan [1].

Sebuah penelitian pada tahun 2006 oleh Dunstan merincikan ada 5 kelas penyebab tangisan bayi berdasarkan suara refleksnya, yaitu bersendawa (Eh), tidak nyaman (Heh), lapar (Neh), sakit perut (Eairh), dan lelah (Owh) [2]. Teori ini dinamakan Dunstan Baby Language, dan dapat dipelajari oleh semua kalangan. Pada tahun-tahun belakangan ini, dimana teknologi sudah semakin maju dan adanya AI (Artificial Intelligence) atau kecerdasan buatan, sebuah riset dari India oleh Mukhopadhyay pada tahun 2013 menyatakan

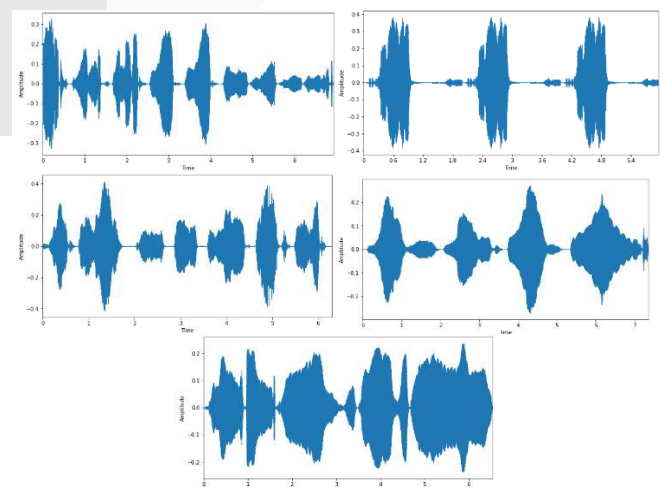
bahwa proses identifikasi atau klasifikasi tangisan bayi dapat dilakukan dengan algoritma machine learning [3].

Pada penelitian kali ini penulis merancang sebuah sistem klasifikasi tangisan bayi yang dapat mengetahui penyebab dari sebuah tangisan. Sistem ini menggunakan metode ekstraksi fitur suara agar sistem dapat menangkap ciri dari setiap suara lebih baik dengan data yang lebih sedikit, metode yang digunakan adalah Mel Frequency Cepstral Coefficients karena memiliki akurasi ekstraksi tinggi sebesar 99.78% dibandingkan dengan metode ekstraksi lainnya [4], dan Artificial Neural Network sebagai metode klasifikasi penyebab tangisan karena metode ini lebih sering digunakan pada speech recognition, salah satunya dengan akurasi 98.125% [5], dan masih jarang digunakan untuk klasifikasi tangisan bayi. Dengan adanya sistem ini, maka orang tua dapat mengetahui penyebab bayi menangis.

II. KAJIAN TEORI

A. Klasifikasi Tangisan Bayi

Studi mengenai klasifikasi penyebab tangisan bayi dimulai pada tahun 1960 oleh kelompok riset Wasz-Hockert dengan bantuan suster yang berpengalaman dan menghasilkan 4 tipe tangisan bayi; sakit, lapar, saat baru lahir, dan kesenangan [1]. Sebuah penelitian pada tahun 2006 oleh Dunstan merincikan ada 5 kelas penyebab tangisan bayi berdasarkan suara refleksnya, yaitu bersendawa (Eh), tidak nyaman (Heh), lapar (Neh), sakit perut (Eairh), dan lelah (Owh) [2]. Teori ini dinamakan Dunstan Baby Language, dan dapat dipelajari oleh semua kalangan.



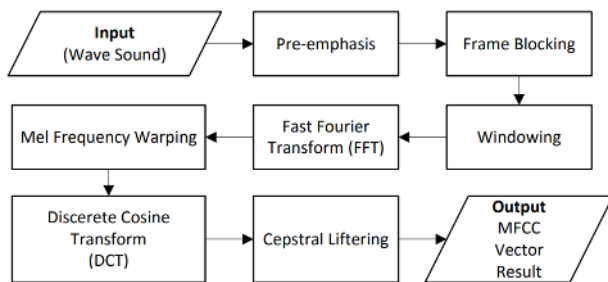
GAMBAR 1

Sinyal Suara Tangisan Bayi (Hu, Bu, Bp, Dc, Ti)

B. Mel-Frequency Cepstral Coefficients

Mel-Frequency Cepstral Coefficients (MFCC) adalah sebuah metode ekstraksi fitur yang mengubah sinyal suara menjadi vektor parameter [6]. Berikut adalah beberapa kelebihan pada MFCC:

1. Dapat menangkap karakter suara untuk mengetahui pola pada suara tertentu.
2. Memberikan output berupa vektor dengan ukuran data yang kecil namun tidak menghilangkan karakteristik dari suara yang diekstrak.
3. Cara kerja MFCC mirip dengan cara kerja pendengaran manusia dalam memberikan persepsi terhadap suara yang didengarkan dan memiliki akurasi sebesar 99.78% dibandingkan dengan metode ekstraksi lain sebesar 95.82% untuk LPCC dan 95.56% untuk BFCC [4].

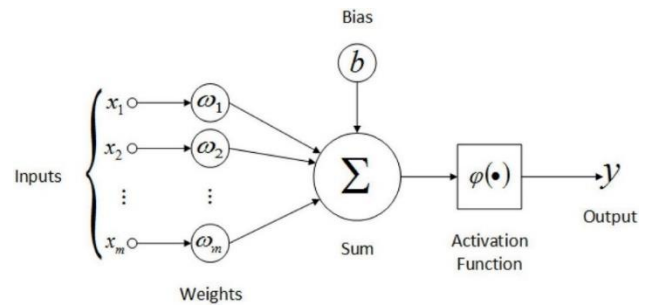


GAMBAR 2
Proses MFCC

1. Pre-emphasis merupakan salah satu jenis filter yang digunakan untuk mempertahankan frekuensi-frekuensi tinggi pada sebuah spektrum,
2. Frame blocking berfungsi untuk menghindari sinyal suara mengalami perubahan akibat pergeseran artikulasi dari produksi suara, sehingga sinyal harus diproses secara short segments (short frame).
3. Windowing menghindari terjadinya kebocoran spektral atau aliasing, yaitu munculnya sinyal baru yang frekuensinya berbeda dengan frekuensi sinyal asli.
4. Fast Fourier Transform digunakan untuk mengubah setiap frame frekuensi pada sample N dari domain waktu menjadi domain frekuensi.
5. Mel-Frequency Warping biasanya dilakukan filterbank, yaitu salah satu bentuk dari filter yang bertujuan untuk mengetahui ukuran energi frequency band tertentu dalam sinyal suara.
6. Discrete Cosine Transform merupakan tahap terakhir dalam rangkaian MFCC yang fungsinya mengubah mel spectrum menjadi domain yang mirip waktu (quefrency).
7. Cepstral Lifting digunakan untuk memperhalus hasil dari seluruh proses MFCC.

C. Artificial Neural Network

Artificial Neural Network salah satu metode klasifikasi dari machine learning yang metode pembelajarannya adalah supervised yaitu mempelajari data yang sudah memiliki label terlebih dahulu [7]. Pada arsitekturnya, ANN termasuk ke dalam Multi-Layer Perceptron (MLP) yang merupakan feed forward neural network dimana informasi masuk dari input menuju hidden layer dan langsung ke output, tidak mengalami proses berulang atau looping sama sekali seperti Recurrent Network. MLP melibatkan banyak layer yang saling terhubung dengan semua neuron di layer selanjutnya (fully connected layer).



GAMBAR 3
Multi-Layer Perceptron

Gambar 2.3 menunjukkan bentuk neural network atau jaringan saraf dengan model Multi-Layer Perceptron dimana x_n adalah masukan yang terhubung dengan neuron dengan bobot w_n pada setiap koneksinya. Neuron akan menjumlahkan semua sinyal yang masuk, dengan setiap sinyal masukan dikali bobot, dan bias adalah faktor lain berupa nilai yang ditambahkan agar hasil perhitungan lebih akurat, dan ϕ merupakan fungsi aktivasi. Persamaannya dinyatakan sebagai berikut:

$$y = \sum (x_n \times w_n) + b = \phi$$

D. Confusion Matrix

Confusion matrix adalah sebuah fungsi analitik prediktif yang sering digunakan untuk mengukur kinerja dari model klasifikasi pada machine learning [21] [22]. Ada 4 nilai dalam sebuah confusion matrix, yaitu:

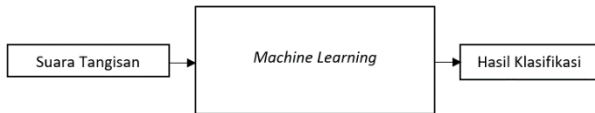
- True Positive (TP), jumlah data yang bernilai positif dan diprediksi benar sebagai positif.
- True Negative (TN), jumlah data yang bernilai negatif, dan diprediksi benar sebagai negative.
- False Positive (FP), jumlah data yang bernilai negatif tetapi diprediksi salah sebagai positif.
- False Negative (FN), jumlah data yang bernilai positif tetapi diprediksi benar sebagai negatif.

TABEL 1
Confusion Matrix

Actual Label	Label	Predicted Label				
		Bu	Dc	Hu	Bp	Ti
Actual Label	Bu	TN		FP		
	Dc		TN	FP		
	Hu	FN	FN	TP	FN	FN
	Bp			FP	TN	
	Ti			FP		TN

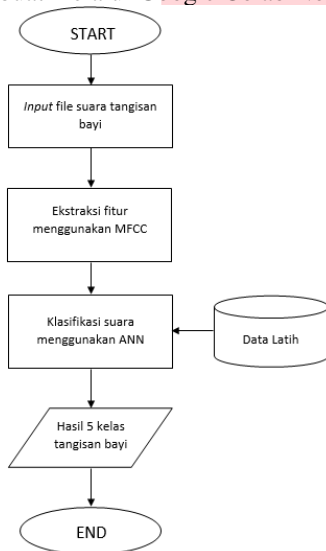
Tabel diatas menunjukkan bentuk multi-label confusion matrix dengan kondisi mencari nilai TP, TN, FP dan FN untuk label 'Hu'. Confusion matrix menampilkan dan membandingkan nilai aktual dengan nilai prediksi dari model dan akan menghasilkan analisa hasil accuracy, precision, recall, dan F1 Score.

III. METODE



GAMBAR 4
Diagram Blok

Sistem ini bekerja saat menerima masukan file suara tangisan bayi, lalu diteruskan dan diproses pada Google Colab Notebook menggunakan metode supervised machine learning, dimana sebelumnya mesin sudah mempelajari data training yang dibuat sehingga dapat mengklasifikasikan penyebab dari suara tangisan bayi dan diproses dengan fitur ekstraksi MFCC, lalu diklasifikasikan menggunakan metode ANN. Untuk merealisasikan sistem ini, maka penulis akan merancang sebuah sistem menggunakan hardware laptop atau komputer, dimana file dataset disimpan di Google Drive dan program dibuat melalui Google Colab Notebook.



GAMBAR 5
Flowchart

Pertama, pilih file suara tangisan bayi yang akan dicoba pada program, setelahnya sistem akan memulai ekstraksi fitur, dan pengklasifikasian pada suara tangisan. Sebelumnya, pastikan laptop sudah tersambung dengan internet terlebih dahulu.

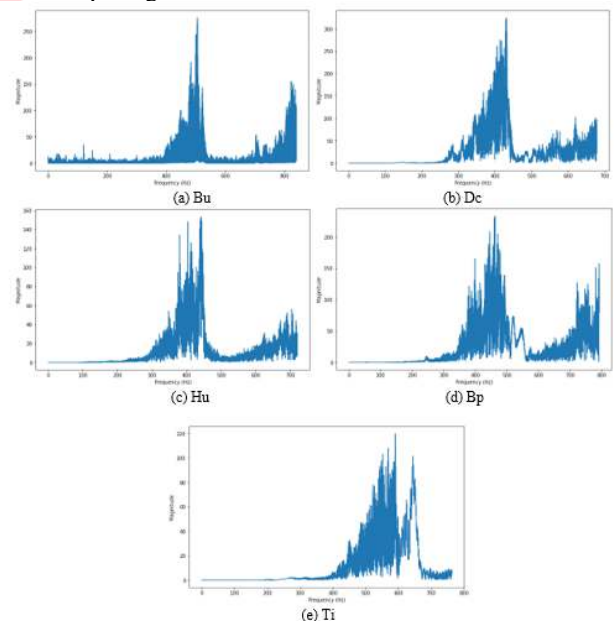
IV. HASIL DAN PEMBAHASAN

Pada tugas akhir ini menggunakan dataset suara dari penelitian tangisan bayi [8] yang berjumlah 200 suara (total 240 dengan dataset pengujian) dan memiliki 5 klasifikasi, yaitu perlu sendawa (burping) pada folder 1 dengan kode label 'bu', tidak nyaman (discomfort) pada folder 2 dengan kode label 'dc', lapar (hungry) pada folder 3 dengan kode label 'hu', sakit perut (belly pain) pada folder 4 dengan kode label 'bp', dan lelah (tired) pada folder 5 dengan kode label 'ti'.

TABEL 2
Pengujian Koefisien MFCC

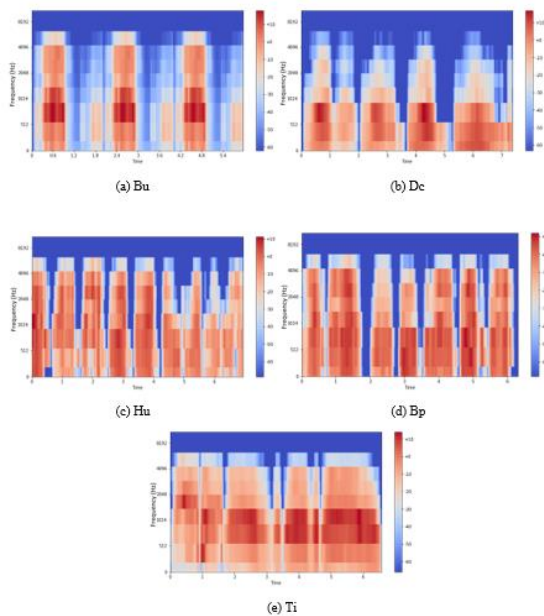
Fungsi Aktivasi	Tanh		
Dense	n_mfcc		
1st Layer	50	40	30
150	55.0%	62.5%	57.5%
200	62.5%	60.0%	62.5%
250	62.5%	60.0%	52.5%
275	62.5%	57.5%	60.0%
300	55.0%	65.0%	52.5%
350	65.0%	65.0%	60.0%
375	52.5%	60.0%	60.0%

Pada pengujian terhadap koefisien MFCC, dilakukan dengan nilai 30, 40, dan 50. Serta pada ANN menggunakan 1 *hidden layer* dengan nilai *dense* 300, dan fungsi aktivasi tanh. Dan didapatkan hasil kombinasi terbaik saat koefisien bernilai 40, menghasilkan akurasi 65%. Pada salah satu proses MFCC, yaitu FFT, sinyal diubah menjadi domain frekuensi seperti gambar 5.



GAMBAR 6
Sinyal Suara Setelah FFT

Selanjutnya, sinyal yang sudah dalam bentuk frekuensi diubah menjadi Mel-Spectrum.



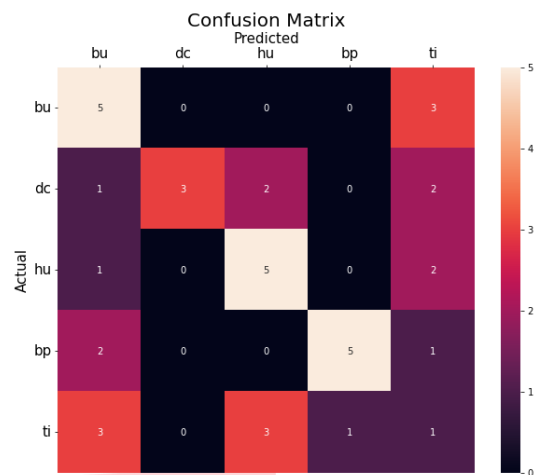
GAMBAR 7
Mel Spectrum

Setelah dalam bentuk *Mel spectrum*, dilakukan konversi menjadi domain miri waktu (quefreny). Hasil konversi ini yang disebut Mel-Frequency Cepstral Coefficients yang berbentuk deretan vektor akustik seperti pada gambar 6.

	feature	class
0	[-444.64023, 62.37537, -67.536064, -1.2063481, ...]	bu
1	[-467.31485, 78.43643, -87.0122, -5.442688, 14, ...]	bu
2	[-459.2786, 59.81047, -40.301186, -13.204036, ...]	bu
3	[-486.5376, 102.80255, -25.359304, -4.6115785, ...]	bu
4	[-459.20743, 58.279217, -39.7656, -11.944051, ...]	bu
...	...	...
195	[-436.24567, 127.134285, -39.832638, -5.714138, ...]	ti
196	[-380.9529, 180.07541, -85.47753, 2.5431292, 3, ...]	ti
197	[-419.96085, 105.63415, -10.739681, -10.275775, ...]	ti
198	[-389.04074, 178.5511, -43.712986, -22.131716, ...]	ti
199	[-368.00372, 122.27825, -64.497505, -1.0286558, ...]	ti

GAMBAR 8
Hasil Ekstraksi Ciri

Sehingga dihasilkan validasi akurasi sistem 65%. Pada pengujian terhadap data suara dilakukan sebanyak 40 kali dengan 8 suara setiap label kelasnya. Dihasilkan confusion matrix seperti gambar 9.



GAMBAR 9
Confusion Matrix

Lalu dihasilkan Analisa terhadap pengujian seperti pada tabel

TABEL 3
Analisa Confusion Matrix

	Accuracy	Precision	Recall	F1 Score
Burp	66%	42%	63%	50%
Discomfort	79%	100%	38%	55%
Hungry	70%	50%	63%	56%
Belly Pain	83%	83%	63%	71%
Tired	56%	13%	14%	13%

Sehingga nilai F1 Score untuk pengujian ini adalah 49%.

V. KESIMPULAN

Pada Tugas Akhir ini dapat disimpulkan bahwa perancangan sistem klasifikasi suara pada tangisan bayi dapat dilakukan menggunakan metode MFCC dan ANN. Sistem ini berfungsi untuk mengetahui dan mengklasifikasi penyebab dari sebuah suara tangisan bayi. Sistem ini menggunakan laptop dengan bantuan aplikasi berbasis web yaitu Google Colaboratory. Pengklasifikasian dimulai dengan mengekstraksi fitur atau karakteristik suara menggunakan metode MFCC, dan selanjutnya diklasifikasikan dengan metode ANN. Pada pengujian sistem, didapat hasil terbaik untuk koefisien MFCC (n_{mfcc}) sebesar 40, epoch 250, batch size 16, serta pada ANN menggunakan 1 hidden layer, dengan nilai dense sebesar 300, dan fungsi aktivasi tanh yang menghasilkan nilai akurasi validasi saat training sebesar 65%, dan nilai F1 Score pada confusion matrix untuk testing model sebesar 49%.

REFERENSI

- [1] E. V. O. Wasz-Höckert, T. J. Partanen, V. Vuorenkoski, K. Michelsson, "The identification of some specific meanings in infant vocalization. Experientia.," 1964.
- [2] M. D. Renanti, "Identifikasi Arti Tangisan Bayi Versi Dunstan Baby Language Menggunakan Jarak Terpendek Dari Jarak Mahalanobis (Infant Cries Identification of Dunstan Baby Language Version using the Shortest Distance of Mahalanobis)," Jurnal Sains Terapan, vol. 3, no. 1, pp. 27–34, 1970, doi: 10.29244/jstsv.3.1.27-34.
- [3] B. K. J. Mukhopadhyay, B. Saha, B. Majumdar, A. K. Majumdar, S. Gorain and A. S. Arya, S. D. Bhattacharya,

“An evaluation of human perception for neonatal cry using a database of cry and underlying cause,” 2013 Indian Conference on Medical Informatics and Telemedicine, ICMIT 2013, 2013.

[4] T. Gulzar, A. Singh, and S. Sharma, “Comparative Analysis of LPCC, MFCC and BFCC for the Recognition of Hindi Words using Artificial Neural Networks,” *Int J Comput Appl*, vol. 101, no. 12, pp. 22–27, 2014, doi: 10.5120/17740-8271.

[5] G. Dede and M. H. Sazli, “Speech recognition with artificial neural networks,” *Digital Signal Processing: A Review Journal*, vol. 20, no. 3, pp. 763–768, 2010, doi: 10.1016/j.dsp.2009.10.004.

[6] A. H. Mansour, G. Zen Alabdeen Salh, and K. A. Mohammed, “Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms,” *Int J Comput Appl*, vol. 116, no. 2, pp. 34–41, 2015, doi: 10.5120/20312-2362.

[7] aditya yanuar R, “Artificial Neural Network (ANN) – Universitas Gadjah Mada Menara Ilmu Machine Learning,” *Machinelearning.Mipa.Ugm.Ac.Id*. p. 5, 2018. [Online]. Available: <https://machinelearning.mipa.ugm.ac.id/2018/05/24/artificial-neural-network-ann/>

[8] R. H. Prasetyo, I. I. Tritoasmoro, L. Novamizanti, and F. T. Elektro, “Sistem Identifikasi Arti Tangisan Bayi Menggunakan Metode Mfcc , Dwt Dan Knn Pada Raspberry Pi Identification System of Baby Cry Using Mfcc , Dwt and,” vol. 7, no. 2, pp. 4203–4210, 2020.