

Klasifikasi Tangisan Bayi Berdasarkan Amplitudo Frekuensi Suara Menggunakan Algoritma MFCC dan CNN

Dwi Puspa Martiyaningsih¹⁾, Chandra Ayunda Apta Soemedhy²⁾, Vina Afrilia Kurniawan³⁾

1,2,3)Program Studi Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto
Jln. DI. Panjaitan No. 128, Purwokerto

Abstract - *The sound of a baby's cry can identify a baby's problems, such as sad, happy, sleepy, and hungry. But not all adults understand and understand the meaning of a baby's cry. Especially for new parents who do not have the experience to raise a child. In today's digital era, a system is needed to identify baby's voice. Research has been done on baby's cry using the Method algorithm Moments of Distribution and K-Nearest Neighbors with an accuracy rate still below 50%. So it is necessary to develop the identification using the algorithms Mel Frequency Cepstrum Coefficient (MFCC) and Convolutional Neural Network (CNN). The first step is to perform feature extraction with amplitude using the MFCC algorithm, then the results of feature extraction can be determined the mean value of the amplitude. For sad and happy categorization, the median value is 0.5, while for hunger and sleepiness, the median value is 0.6 in data I and III and 0.875 in data II. Furthermore, it is classified by the CNN algorithm and the accuracy obtained for the sad and happy categories is 95% with a percentage of the probability of the truth of the test on the sample being above 75%, while for the hungry and sleepy category is 100% with the percentage of the probability of the truth of the test on the sample ranging from 63% to 100%.*

Keywords - CNN, MFCC, Machine Learning, Deep Learning, Baby Cry, Klasifikasi, Hungry, sleepy.

Abstrak - *Suara tangisan bayi dapat mengidentifikasi masalah bayi, seperti sedih, bahagia, mengantuk, dan lapar. Tetapi tidak semua orang dewasa mengerti dan memahami arti tangisan bayi. Terutama bagi orang tua baru yang belum memiliki pengalaman untuk mengasuh seorang anak. Pada era digital saat ini, dibutuhkan suatu sistem untuk mengidentifikasi suara bayi. Pernah dilakukan penelitian mengenai tangis bayi*

*) penulis korespondensi: Dwi Puspa Martiyaningsih
Email: 1910244@ittelkom-pwt.ac.id

menggunakan algoritma Metode Moments of Distribution dan KNearest Neighbours dengan tingkat akurasi masih dibawah 50%. Sehingga perlu dilakukan pengembangan Identifikasi tersebut dengan menggunakan metode algoritma Mel Frequency Cepstrum Coefficient (MFCC) dan Convolutional Neural Network (CNN). Langkah pertama yaitu dengan melakukan ekstraksi ciri dengan amplitudo menggunakan algoritma MFCC, selanjutnya hasil dari ekstraksi ciri dapat ditentukan nilai tengah dari amplitudonya. Untuk pengkategorian sedih dan Bahagia didapatkan nilai tengah 0.5, sedangkan untuk lapar dan mengantuk didapat nilai tengah 0.6 pada data I dan III dan 0.875 pada data II. Selanjutnya diklasifikasikan dengan algoritma CNN dan akurasi yang didapatkan untuk kategori sedih dan bahagia adalah 95% dengan presentase probabilitas kebenaran pengujian terhadap sampel diatas 75%, sedangkan untuk kategori lapar dan mengantuk adalah 100% dengan presentase probabilitas kebenaran pengujian terhadap sampel berkisar 63% sampai 100%

Kata Kunci – CNN, MFCC, Machine Learning, Deep Learning, Tangis Bayi, Klasifikasi, Lapar, Mengantuk.

I. PENDAHULUAN

Komunikasi adalah pengiriman atau penerimaan suatu informasi atau berita kepada suatu individu atau kelompok yang nantinya akan diproses sehingga informasi tersebut dapat tersampaikan dengan baik [1]. Alasan umum manusia melakukan komunikasi adalah untuk mempengaruhi orang lain, membangun hubungan, memperoleh pengetahuan, membantu orang, dan komunikasi untuk bermain [2]. Komunikasi dapat dilakukan dengan dua cara, tertulis maupun lisan. Komunikasi secara tertulis dilakukan dengan cara proses penukaran pesan melalui surat ataupun melalui media sosial yaitu aplikasi tukar pesan secara online seperti *whatsapp* dan yang lainnya.

Salah satu komunikasi yang dapat dilakukan oleh bayi adalah menangis. Menangis adalah sebuah

ungkapan perasaan sedih, kecewa, dan sebagainya dengan mengeluarkan air mata dan suara. Menangis adalah salah satu cara bayi untuk berkomunikasi [1], baik dengan orang tuanya maupun orang asing. Menangis merupakan respons fisiologis normal terhadap berbagai stimulus nonverbal pada bayi [3]. Bayi mengeluarkan suatu tangisan sesuai dengan kondisi tertentu yang dialaminya. Suara tangis bayi tersebut dapat digunakan untuk mengatasi masalah pada bayi, seperti sedang lapar ataupun mengantuk. Dari suara tangis tersebut, seseorang yang telah terbiasa mengasuh bayi dapat cepat mengenali masalah bayi dari suara tangisnya, tetapi tidak semua orang dapat cepat tanggap dalam memahami masalah yang dialami bayi dari tangisnya, terutama bagi para ibu muda yang berada jauh dari keluarga sehingga tidak mendapat bimbingan dari orang yang sudah berpengalaman. Sebelumnya, telah dilakukan penelitian mengenai klasifikasi tangis bayi dengan menggunakan algoritma Metode *Moments of Distribution* dan *K-Nearest Neighbours* dengan tingkat akurasi yang masih rendah, sehingga perlu dilakukan penelitian lagi dengan algoritma berbeda, salah satunya *Mel Frequency Cepstrum Coefficient* (MFCC) dan *Convolutional Neural Network* (CNN). MFCC digunakan untuk proses ekstraksi dari suara dengan mengubah sinyal suara menjadi sebuah parameter [4]. Penggunaan MFCC pada penelitian ini karena metode tersebut memiliki kemampuan yang dapat mewakili spektrum bicara secara ringkas [5]. Kemudian, CNN digunakan untuk proses pengklasifikasian karena sangat efektif untuk mewakili pola spesial yang memungkinkan untuk mengekstrak beragam ciri dari proses ekstraksi citra suara [6]. Penelitian ini menggunakan variabel berupa tangis bayi yang meliputi bahagia, sedih, mengantuk, dan lapar sebagai kelas untuk pengklasifikasian.

II. PENELITIAN YANG TERKAIT

Di Indonesia, ada penelitian yang meneliti mengenai suara tangis bayi seperti penelitian yang dilakukan oleh Aditya Singgi Prayogi, dkk yang berjudul *Klasifikasi Suara Tangisan Bayi Berdasarkan Prosodic Features Menggunakan Metode Moments of Distribution dan K-Nearest Neighbours*. Pada penelitian ini diuji Langkah awal adalah melakukan *preprocessing* dari video Dunstan Baby Language sebagai acuan untuk label klasifikasi. Data dari video dipotong menjadi 150 file dengan format wav dan diubah menjadi 5 versi suara tangisan yang berbeda. Setelah melakukan *preprocessing*, metode yang dilakukan selanjutnya adalah ekstraksi fitur dengan *Prosodic Features*. Dalam *Prosodic Feature* ada tiga

karakteristik utama berupa *fundamental frequency*(f_0), *frame range*, dan *loudness*. Fitur *pitch* menangkap *fundamental frequency*(f_0), sedangkan energi mencerminkan naik turun dari level *loudness* dan *frame time* mempresentasikan setiap detik yang didapat dari *pitch* dan *loudness*.

Kemudian untuk pengidentifikasian pola dari setiap sampel suara tangis bayi dilakukan berdasar nilai rata-rata, varian, standar deviasi, kemiringan dan kecondongan, dengan diterapkan metode *Moments of Distribution*. Lalu dilanjutkan pengenalan suara dengan klasifikasi menggunakan metode *KNearest Neighbour* serta menghitung tingkat akurasi prediksi yakni dengan cara mencocokkan antara suara tangis bayi secara aktual dengan sample bayi yang telah disediakan [7].

Penelitian lain mengenai klasifikasi suata tangis bayi juga pernah dilakukan oleh Retno Hadi Prasetyo, dkk menggunakan metode yang sama dengan penelitian sebelumnya, yaitu *K-Nearest Neighbours*. Yang menjadi pembeda dengan penelitian sebelumnya yaitu menggunakan metode gabungan dari *Discrete Wavelet Transform* (DWT) dan *Mel Frequency Cepstrum Coefficient* (MFCC), dimana metode tersebut digunakan untuk mengekstraksi ciri dengan cara menganalisis sinyal suara berdasarkan tinggi rendahnya frekuensi yang dapat digunakan sebagai identitas individu, sedangkan penelitian sebelumnya menggunakan *Prosodic Features* [8].

Carl Ray Wairata, dkk juga pernah melakukan penelitian mengenai pengklasifikasian suara. Penelitian tersebut dilakukan untuk membedakan genre music Indonesia. Perbedaan dengan penelitian sebelumnya yaitu penggunaan metode *Convolutional Neural Network* (CNN) [9]. CNN merupakan salah satu model *Deep Learning* yang banyak diterapkan dibidang penelitian yang mana salah satunya adalah pengenalan atau pengklasifikasian suara [10].

Penulis juga melakukan penelitian mengenai klasifikasi suara tangis bayi dengan menggunakan penggabungan metode *Mel Frequency Cepstrum Coefficient* (MFCC) dan *Convolutional Neural Network* (CNN) yang dirasa akan memiliki tingkat akurasi yang cukup tinggi.

III. METODE PENELITIAN

A. Subjek Penelitian

Penelitian yang dilakukan adalah pengklasifikasian suara tangisan bayi hingga menghasilkan sebuah kesimpulan yang final untuk mengetahui apakah bayi tersebut lapar atau mengantuk. Sampel yang digunakan adalah rekaman suara tangisan bayi berjumlah 25 dengan format wav dengan uraian 15 suara bayi laki-laki dan 10 suara bayi perempuan. Setiap sampel kemudian diproses dengan MFCC untuk kemudian diamati amplitudonya.

B. Metode Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan dua metode, yaitu :

- 1) *Observasi*, pengumpulan data dilakukan dengan cara mengamati suara tangis bayi sehingga dapat mengetahui perbedaannya.
- 2) *Wawancara*, bertanya kepada orang tua bayi mengenai sebab dari tangisan bayinya mengenai tangis bayi tersebut termasuk dalam kelas bahagia, sedih, lapar, dan mengantuk. Sehingga penulis dapat tahu pasti perasaan apa yang dirasakan oleh bayi tersebut

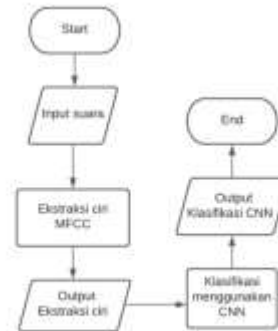
C. Alat dan Bahan

Alat dan bahan untuk penyusunan sistem ini diantaranya berupa hardware dan software:

- 1) Smartphone
- 2) Laptop
- 3) Google Colab
- 4) Dataset suara bayi
- 5) VSCod (*Visual Studio Code*)

D. Perancangan Sistem

Sistem ini dilakukan dengan cara menginputkan file suara tangisan bayi dengan format wav, kemudian dilakukan ekstraksi ciri dengan menggunakan metode MFCC yang kemudian akan diklasifikasikan dengan metode CNN yang akan diproses sehingga muncul kesimpulan apakah bayi tersebut merasa bahagia, sedih, lapar, ataupun mengantuk. Pemrosesan sistem seperti pada Gbr. 1



Gbr. 1 Proses Ekstraksi Ciri dan Klasifikasi

IV. MEL FREQUENCY CEPSTRAL COEFFICIENT (MFCC)

Feature extraction (ekstraksi ciri) adalah suatu proses yang digunakan untuk menentukan nilai atau vektor yang akan menjadi suatu ciri dari obyek atau individu [11]. Proses tersebut dilakukan dengan mengekstrak informasi dari data sampel yang didapat agar data bisa diklasifikasikan sesuai kategorinya masing-masing [12]. Ekstraksi tersebut menghasilkan sinyal suara dari setiap frame dan kemudian diubah ke dalam domain cepstral [13]. Proses MFCC dapat dilihat pada Gbr. 2.



Gbr. 2. Proses MFCC

- *Pre-emphasis*, merupakan proses dimana sinyal akan dilewatkan kepada sebuah filter yang dapat menerima frekuensi tinggi [14]. Persamaan *Preemphasis* dapat dilihat pada persamaan (1) [15].

$$y(n) = s(n) - \alpha s(n-1) \quad (1)$$

Dimana :
 S_n : nilai sampel ke-n
 α : konstanta pre-emphasis, $0,9 \leq \alpha \leq 1,0$

- *Frame Blocking*, berfungsi untuk membagi sinyal menjadi beberapa frame yang memiliki Panjang tertentu. Sinyal suara yang terdiri dari S sampel (X(S)) akan dibagi kedalam beberapa frame dengan N sampel yang masing-masing akan dipisahkan oleh M(M<N) [16].
- *Windowing*, berfungsi untuk menghilangkan noise yang muncul diujung frame dengan mempertimbangkan blok [14].
- *Fast Fourier Transform(FFT)*, N dan frame diubah dari domain waktu ke domain frekuensi dengan tujuan agar sinyal dapat diproses dalam spectral substraksi. Proses FFT dapat dituliskan kedalam persamaan(2) [17].

$$F(u) = \frac{1}{n} \sum_{x=0}^{x=N-1} f(x) \exp\left[-\frac{2j\pi ux}{N}\right] \quad (2)$$

- *Mel-Frequency Wrapping*, pada tahap ini sinyal suara dari domain frekuensi diubah menjadi domain frekuensi mel pada frekuensi Hz sehingga menghasilkan persamaan (3) dan (4).

$$Mel f = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (3)$$

$$f = 700\left(10^{\frac{m}{2595}} - 1\right) \quad (4)$$

dimana:

Mel f = nilai frekuensi f

Tahap ini menghasilkan mel filter bank yang akan mempresentasikan besar energi pada rentang frekuensi masing-masing [18].

- *Discreate Cosine Transform (DCT)* merupakan Teknik merubah sinyal suara menjadi frekuensi dasar dengan memperhitungkan nilai riil dari hasil transformasi. Persamaan DCT dapat dilihat pada persamaan (5) dan (6) [19].

$$X(k) = C(k) \sqrt{\frac{2}{n}} \sum_{x=0}^{n-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad (5)$$

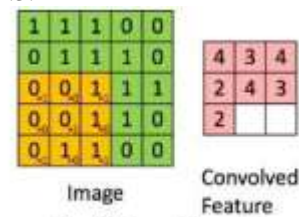
dimana:

$$C(k) = \begin{cases} 2^{-\frac{1}{2}}, & \text{untuk } k = 0 \\ 1, & \text{untuk } k \text{ lainnya} \end{cases} \quad (6)$$

V. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional Neural Network (CNN) merupakan salah satu kelas *feed-forward artificial neural networks* yang terinspirasi dari pola konektivitas antar neuron menyerupai organisasi visual cortex [20]. *Convolutional Neural Network (CNN)* adalah *Deep Neural Network* dengan jenis neural network yang didesain untuk mengolah data dua dimensi [21]. CNN Memiliki beberapa Layer, yaitu *Convolutional Layer*, *Pooling Layer*, dan *Fully connected Layer*.

- *Convolutional Layer*, merupakan tahap dimana seluruh data menyentuh lapisan *Convolutional* yang mengalami proses *konvolusi* dan akan difilter kemudian menghasilkan sebuah activation map. Lapisan pada layer ini memiliki tiga parameter, yaitu *depth*, *stride*, dan *zero padding* [21]. Operasi *konvolusi* dapat dilihat pada Gbr.3.



Gbr. 3. Operasi Konvolusi

- *Pooling Layer*, merupakan lapisan yang menggunakan *Feature Map* sebagai input dan mengolahnya berdasar nilai piksel terdekat. Lapisan ini disisipkan ke dalam lapisan konvolusi secara teratur. Bentuk *Pooling Layer* yang paling umum adalah 2x2 [22]. Operasi pada *Pooling Layer* dapat dilihat pada Gbr.4.

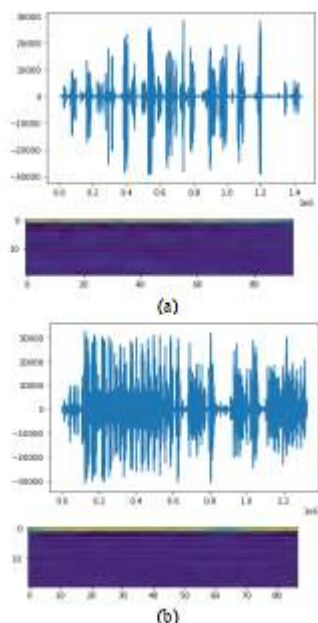


Gbr. 4. Operasi pada *Pooling Layer*

- *Fully connected Layer*, merupakan lapisan yang digunakan untuk tujuan melakukan transformasi pda dimensi data agar dapat diklasifikasikan secara linear [22].

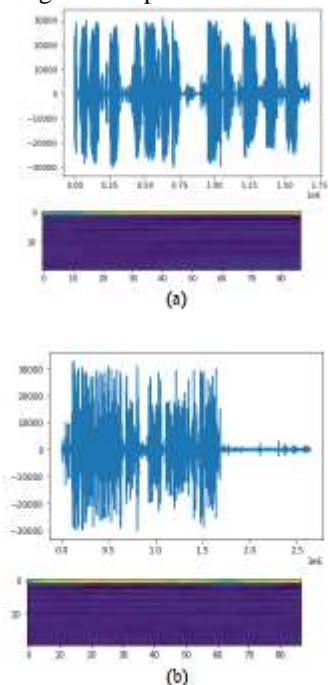
VI. HASIL DAN PEMBAHASAN

Pengujian awal dilakukan dengan ekstraksi ciri menggunakan metode MFCC. Setiap Sampel diinputkan ke dalam sistem, kemudian dapat memvisualisasikan sebuah audio menjadi grafik frekuensi dan spektrum. Grafik audio mengenai perasaan bayi sedang senang atau sedih dapat dilihat dari Gbr. 2.



Gbr. 5 Grafik Tangis Bayi. (a) bahagia, (b) sedih.

Grafik audio mengenai perasaan bayi sedang lapar atau mengantuk dapat dilihat dari Gbr. 3.



Gbr. 6 Grafik Tangis Bayi. (a) mengantuk, (b) lapar.

Setelah melakukan ekstraksi ciri menggunakan MFCC, selanjutnya data sampel hasil observasi dan wawancara ditraining dengan membagi data ke dalam dua kategori data, yaitu data latih dan data uji. Data latih pada penelitian ini bernilai 0,9 atau setara dengan 22 data tangis bayi dan data uji bernilai 0,1 atau setara dengan 3 data tangis bayi. Pertama, penulis melakukan

klasifikasi tangis bayi untuk membedakan perasaan sedih dan bahagia. Langkah awal dengan melakukan perhitungan untuk mencari *Sampling Rate* terkecil dan terbesar. Didapatkan hasil *Sampling rate* terkecil dari sampel Gbr.1 (b) yang bernilai 30.0 milik audio bayi menangis (sedih). Dari hasil tersebut dapat dibentuk dataset untuk melakukan training dengan mencari amplitude terbesar dan terkecil, setelah itu dataset diubah menjadi bentuk array untuk melakukan testing dari data yang sudah ditraining.

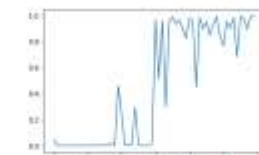
Langkah selanjutnya dilakukan klasifikasi dengan training menggunakan *epoch* sebanyak 20 kali dan didapat akurasi hingga 95%. Gbr. 7 adalah proses training dan akurasi yang didapat.

```

Epoch 0/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 1/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 2/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 3/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 4/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 5/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 6/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 7/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 8/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 9/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 10/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 11/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 12/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 13/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 14/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 15/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 16/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 17/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 18/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 19/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
Epoch 20/20: Loss: 0.0000 - Acc: 0.0000 - Accuracy: 0.0000
    
```

Gbr. 7 Proses training data sedih dan bahagia beserta hasil akurasinya

Gbr. 8 adalah grafik hasil training yang dilakukan pada dua data sampel, sedih dan Bahagia



Gbr. 8 Grafik klasifikasi sedih dan bahagia

Pada Gbr. 5 dapat dilihat bahwa dari detik 0 sampai 30 yang merupakan sampel audio bayi menangis (sedih) menunjukkan nilai 0.0 sampai 0.5, sedangkan detik 31 sampai 60 milik sampel audio bayi tertawa (bahagia) menunjukkan kisaran nilai dari 0.5 sampai 1.0. Dari grafik tersebut, penulis mengambil median 0.5 untuk mencari probabilitas bayi sedih ataupun bahagia. Apabila nilai berkisar antara 0.0 sampai 0.5, maka dikategorikan bayi tersebut sedang bahagia, sedangkan nilai 0.6 sampai 1.0 dikategorikan bayi sedang sedih.

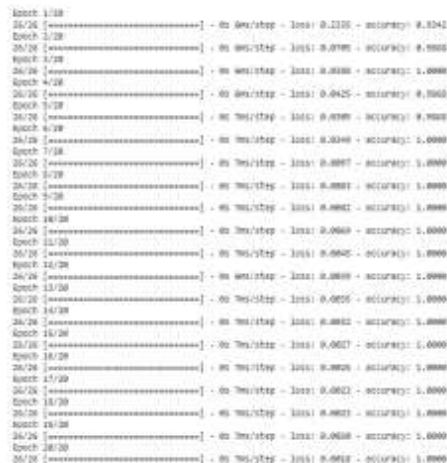
TABEL I
 TABEL PENGUJIAN TERHADAP SAMPEL SEDIH DAN BAHAGIA

Sampel	Probabilitas	
	Bahagia	Sedih
A	3%	97%
B	81,81%	18,18%
C	0%	100%

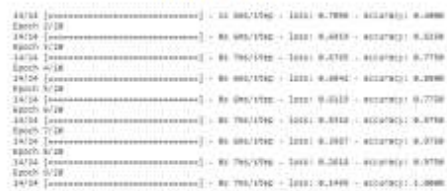
Pada TABEL I dapat dilihat bahwa presentase benar dari pengujian tersebut cukup tinggi, yaitu diatas 75%.

Setelah dilakukannya pengkategorian antara sedih dan bahagia, penulis melakukan klasifikasi pada audio menangis untuk mengkategorikan apakah bayi sedang lapar atau mengantuk. Dengan cara yang sama, penulis melakukan proses training terhadap 3 pasang sampel untuk mendapatkan akurasi yang lebih tepat. Sampel I mendapat *Sampling Rate* terkecil yaitu 38.0, sampel II 20.0, dan sampel III adalah 29.0. Dengan langkah yang sama seperti saat klasifikasi sedih dan bahagia, data sampel diubah ke bentuk array, kemudian ditraining dengan *epoch* sebanyak 20 kali.

Gbr. 9, 10, 11 menunjukkan proses training dan akurasi yang didapat.



Gbr. 9 Proses training data sedih dan bahagia beserta hasil akurasiya terhadap data I



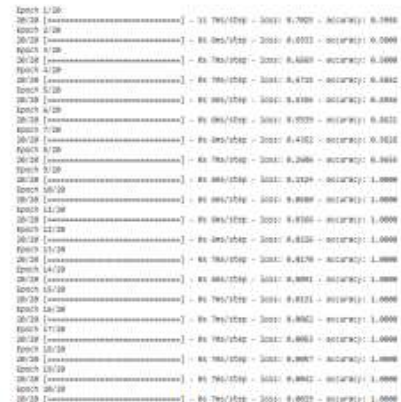
Gbr. 10 Proses training data sedih dan bahagia beserta hasil akurasiya terhadap sampel II



Gbr. 11 Proses training data sedih dan bahagia beserta hasil akurasiya terhadap data III

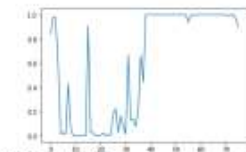


Gbr. 12 Grafik klasifikasi lapar dan mengantuk pada data I



Gbr. 13 Grafik klasifikasi lapar dan mengantuk data II

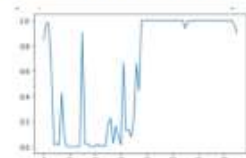
Gbr. 12, 13, 14 adalah grafik hasil training yang dilakukan pada dua data sampel, sedih dan Bahagia



Gbr. 13 Grafik klasifikasi lapar dan mengantuk data III



Gbr. 14 Grafik klasifikasi lapar dan mengantuk data III



Pada Gbr. 12 yang merupakan training pada sampel I dapat dilihat bahwa dari detik 0 sampai 20 yang merupakan sampel audio bayi lapar menunjukkan nilai 0.0 sampai 0.6, sedangkan detik 21 sampai 40 milik sampel audio bayi mengantuk menunjukkan kisaran nilai dari 0.7 sampai 1.0. Pada grafik tersebut, diambil nilai median 0.6 untuk mencari probabilitas bayi lapar ataupun mengantuk.

Apabila nilai berkisar 0.0 sampai 0.6, maka dikategorikan bayi tersebut sedang lapar, sedangkan nilai 0.7 sampai 1.0 dikategorikan bayi sedang mengantuk.

Pada Gbr. 13 yang merupakan training pada sampel II dapat dilihat bahwa dari detik 0 sampai 20 yang merupakan sampel audio bayi lapar menunjukkan nilai 0.0 sampai 0.875, sedangkan detik 21 sampai 40 milik sampel audio bayi mengantuk menunjukkan kisaran nilai dari 0.876 sampai 1.0. Pada grafik tersebut, diambil nilai median 0.875 untuk mencari probabilitas bayi lapar ataupun mengantuk. Apabila nilai berkisar 0.0 sampai 0.875, maka dikategorikan bayi tersebut sedang lapar, sedangkan nilai 0.876 sampai 1.0 dikategorikan bayi sedang mengantuk.

Pada gambar 14 yang merupakan training pada sampel III hasil yang didapat sama dengan training pada sampel I.

TABEL II
 TABEL TRAINING TERHADAP SAMPEL LAPAR DAN MANGANTUK

Data	Waktu Komputasi (s)	Akurasi	Presentase Benar Tiap Sampel (%)			
			A	B	C	D
I	7,06	100%	100	97	86	73
II	5,08	100%	93	93	63	90
III	6,75	100%	93	62	93	63

Pada TABEL II dapat dilihat bahwa data II memiliki waktu komputasi paling cepat yaitu 5,08 detik. Dilihat dari presentase kebenaran data yang ditraining, data I memiliki rata-rata presentase kebenaran paling tinggi, yaitu 89. Kemudian diikuti data II dengan presentase kebenaran 84,75, dan yang terakhir data III dengan presentase kebenaran paling rendah yaitu 77,75.

Pada Penelitian sebelumnya yaitu menggunakan Metode *Moments of Distribution* dan *K-Nearest Neighbours* didapatkan tingkat akurasi terbaik 90% dengan waktu komputasi selama 8,8 detik dan akurasi terendah 48% dengan waktu komputasi 4,1 detik menggunakan data sampel umum yang dilakukan oleh seorang musisi asal Australia. Sedangkan penelitian yang dilakukan oleh penulis telah mengalami perkembangan dimana akurasi yang didapat sudah mencapai 100% dengan waktu komputasi 5,08 detik dengan presentase kebenaran 89% menggunakan data sampel yang diambil secara langsung dilakukan saat observasi di lingkungan penulis

VII. KESIMPULAN

Pada penelitian ini telah dirancang suatu sistem untuk mengklasifikasikan suara tangisan bayi. Sistem ini menggunakan metode MFCC sebagai ekstraksi ciri

dan CNN untuk pengkategorian. Sistem ini dapat mengkategorikan apakah bayi sedang sedih atau bahagia dengan menggunakan dataset berupa audio suara bayi yang didapat secara langsung melalui observasi dan wawancara. Sistem dapat membedakan suara bayi yang sedang tertawa dan menangis hingga mencapai akurasi sebesar 95% dan juga tingkat kebenaran terhadap pengujian yang cukup tinggi, yaitu diatas 75%. Selain itu, sistem ini juga dapat membedakan secara suara tangis bayi, apakah bayi sedang lapar atau mengantuk hingga mencapai akurasi sebesar 100% dengan tingkat kebenaran berkisar 59% sampai 93%. Berikut ini adalah beberapa saran yang diberikan untuk pengembangan penelitian ini:

1. Mencoba menggunakan metode klasifikasi yang lainnya seperti KNN, LDA, atau PCA.
2. Mengintegrasikannya kedalam aplikasi berbasis android sehingga mudah digunakan.
3. Mencari data lebih banyak untuk dijadikan dataset sebagai sampel pengujian.

UCAPAN TERIMA KASIH

Ucapan terima kasih untuk pihak - pihak yang terkait, terutama kepada Tuhan YME yang telah memberikan limpahan kesehatan bagi penulis, sehingga penulis dapat menyelesaikan sistem klasifikasi suara bayi dengan baik. Selanjutnya, ucapan terimakasih kepada kedua orang tua karena telah memberikan asupan gizi yang cukup sehingga penulis dapat menjaga kesehatan dengan situasi dan kondisi saat ini. Tak lupa penulis ucapkan terimakasih kepada dosen pembimbing, Bapak Agi Prasetiadi, S.T., M. Eng. yang telah membimbing penulis dalam menyelesaikan sistem klasifikasi tangisan bayi ini. Kemudian penulis ucapkan terimakasih kepada orang tua bayi karena telah bersedia melakukan record untuk tangisan suara bayi yang dijadikan sebagai bahan penelitian sistem klasifikasi suara bayi. Serta terimakasih kepada teman-teman, terutama untuk Abdul Hafidh Priyanto yang turut serta memberikan dukungan dan bantuan dalam proses menyelesaikan sistem klasifikasi tangisan bayi dan jurnal ini.

DAFTAR PUSTAKA

- [1] KBBI, "Kamus Besar Bahasa Indonesia," *KBBI*, 2021. <https://kbbi.web.id/> (accessed Jul. 14, 2021).
- [2] A. Liliweri, *Komunikasi Antar Personal*.

- Jakarta Timur: Prenada Media, 2017.
- [3] D. Anurogo, "Manajemen Menangis pada Bayi," *Cermin Dunia Kedokt.*, vol. 46, pp. 8–13, 2019, [Online]. Available: <http://103.13.36.125/index.php/CDK/article/view/530>.
- [4] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Networks*, vol. 130, pp. 22–32, 2020, doi: 10.1016/j.neunet.2020.06.015.
- [5] D. Astuti, "Aplikasi Identifikasi Suara Hewan Menggunakan Metode Mel-Frequency Cepstral Coefficients (MFCC)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 26–34, 2019, doi: 10.20895/inista.v1i2.50.
- [6] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, no. January, pp. 24–49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.
- [7] A. S. Prayogi, M. Rizqi, and T. M. Fahrudin, "Klasifikasi Suara Tangisan Bayi Berdasarkan Prosodic Features Menggunakan Metode Moments of Distribution dan K-Nearest Neighbours," *Teknika*, vol. 8, no. 2, pp. 119–125, 2019, doi: 10.34148/teknika.v8i2.206.
- [8] R. H. Prasetyo, "Sistem Identifikasi Arti Tangisan Bayi Menggunakan Metode Mfcc, Dwt Dan Knn Pada Raspberry Pi," vol. 7, no. 2, pp. 6–28, 2020.
- [9] C. R. Wairata, E. R. Swedia, and M. Cahyanti, "Pengklasifikasian Genre Musik Indonesia Menggunakan Convolutional Neural Network," *Sebatik*, vol. 25, no. 1, pp. 255–261, 2021, doi: 10.46984/sebatik.v25i1.1286.
- [10] Y. Yohannes and R. Wijaya, "Klasifikasi Makna Tangisan Bayi Menggunakan CNN Berdasarkan Kombinasi Fitur MFCC dan DWT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 599–610, 2021, doi: 10.35957/jatisi.v8i2.470.
- [11] E. Susanti, S. Sasongko, and I. G. Suta W, "Klasifikasi Suara Berdasarkan Usia Menggunakan Mel Frequency Cepstral Coefficient (Mfcc)," vol. 4, no. 2, pp. 120–126, 2017.
- [12] M. M. Effendi, "Pengenalan Pengaruh Suara Konsonan Terhadap Vokal Menggunakan MFCC dan SVM," *IT Soc.*, vol. 3, no. 02, pp. 34–39, 2018, doi: 10.33021/itfs.v3i02.583.
- [13] F. D. Adhinata, D. P. Rakhmadani, and A. J. T. Segara, "Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, pp. 28–33, 2021, doi: 10.20895/dinda.v1i1.198.
- [14] A. Anggoro, S. Herdjunto, and R. Hidayat, "MFCC dan KNN untuk Pengenalan Suara Artikulasi P," *Avitec*, vol. 2, no. 1, pp. 13–19, 2020, doi: 10.28989/avitec.v2i1.605.
- [15] H. Heriyanto, "Deteksi Suara Ucapan Salam Bahasa Arab Menggunakan Mel Frequency Cepstral Coefficient (Mfcc) Dan Pemilihan Fitur Min Max," *Semin. Nas. Inform.*, vol. 2020, no. Semnasif, pp. 405–414, 2020, [Online]. Available: <http://103.23.20.161/index.php/semnasif/article/view/4133>.
- [16] M. T. N. Azizah, A. Hidayatno, and Y. Christyono, "Aplikasi Pengenal Pengucap Berbasis Identifikasi Suara Dengan Ekstraksi Ciri Mel-Frequency Cepstrum Coefficients (Mfcc) Dan Kuantisasi Vektor," *Transient J. Ilm. Tek. Elektro*, vol. 6, no. 4, pp. 638–643, 2017.
- [17] Muhathir, R. Muliono, and Susilawati, "Fast Fourier Transform Analysis For Voice Recognition Registers Women In Techniques Singing," *J. Inf. Technol. Educ. Res.*, vol. 2, no. 2, pp. 92–98, 2019, doi: 10.31289/jite.v2i2.2166.
- [18] D. K. Putra, I. I. Triasmoro, and R. D. Atmaja, "Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (MFCC) Dan Gaussian Mixture Model (GMM)," *eProceedings Eng.*, vol. 4, no. 2, pp. 1766–1772, 2017, [Online]. Available: <http://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/487/460>.
- [19] I. Prayoga Prawiro, R. Magdalena, and I. N. Apraz Ramatryana, "SIMULASI DAN ANALISIS PERBANDINGAN ANTARA METODE DISCRETE COSINE TRANSFORM (DCT) DAN MODIFIED DISCRETE COSINE TRANSFORM

- (MDCT) PADA PEMISAHAN REFF LAGU,” *ISSN 2355-9365 e-Proceeding Eng.*, vol. 5, no. 3, pp. 5513–5520, 2018.
- [20] M. H. Ashshiddieqy, Jondri, and A. Rizal, “Klasifikasi Suara Paru Dengan Convolutional Neural Network (CNN),” *eProceedings Eng.*, vol. 07, no. 02, pp. 8506–8512, 2020.
- [21] P. A. Nugroho, I. Fenriana, and R. Arijanto, “Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia,” *Algor*, vol. 2, no. 1, pp. 12–21, 2020.
- [22] S. Ilahiyah and A. Nilogiri, “Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network,” *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.*, vol. 3, no. 2, pp. 49–56, 2018.