

# Speech Recognition Penderita Bibir Sumbing di Kabupaten Sumbawa

Anggreni <sup>1\*</sup>, Siti Agrippina Alodia Yusuf <sup>2</sup>, Muhammad Imam Dinata <sup>3</sup>, Nani Sulistianingsih <sup>4</sup>

<sup>1</sup>Program Studi Sistem dan Teknologi Informasi, Universitas Muhammadiyah Mataram

<sup>1</sup>[anggrenimuchali@gmail.com](mailto:anggrenimuchali@gmail.com), <sup>2</sup>[siti.agrippina@ummat.ac.id](mailto:siti.agrippina@ummat.ac.id), <sup>3</sup>[imam.dinata@ummat.ac.id](mailto:imam.dinata@ummat.ac.id), <sup>4</sup>[nani.sulistianingsih@ummat.ac.id](mailto:nani.sulistianingsih@ummat.ac.id)

## Abstrak

Kegagalan bersatunya dua sisi bagian rongga mulut dengan sempurna pada masa kehamilan mengakibatkan adanya celah pada langit-langit mulut (*palatoschisis*), dan menjadi kelainan bawaan pada bibir bagian atas disebut sebagai sumbing bibir (*labioschisis*). Bibir sumbing dapat dibedakan dalam dua kondisi yaitu sumbing bibir komplit mengakibatkan kelainan pada gusi, bantalan gigi (*prosesus alveolis*), langit-langit mulut (*palatum*) dan lubang hidung (*nostril*). Indonesia sebagai salah satu negara di Asia Tenggara menduduki peringkat ke tujuh, dengan 20,4 % penduduk menderita bibir sumbing dan lelangit. Dalam kaitan dengan kelahiran per tahun, prevalensi bayi lahir dengan bibir sumbing dan lelangit ditemukan semakin tinggi yaitu lebih dari 8.900 bayi lahir dengan bibir sumbing dan lelangit per tahun (Kemenkes RI, 2018). Kabupaten Sumbawa yang terdapat di Provinsi Nusa Tenggara Barat merupakan salah satu kabupaten dengan jumlah 21 orang penderita terdaftar sebagai calon pasien di Rumah Sakit Umum Sumbawa Besar. Pengabdian Masyarakat ini bertujuan untuk mengaplikasikan penggunaan Sistem Pengenalan Suara (*Speech Recognition*) pada penderita bibir sumbing berfokus pada pengucapan Huruf P. Data suara pengucapan kata kapak, paku dan atap diperoleh dari 7 orang penderita bibir sumbing di Kabupaten Sumbawa. Setiap kata diulang sebanyak 15 kali pengucapan. Metode K-nearest neighbor (KNN) digunakan untuk mengevaluasi performa dari . Hasil eksperimen menunjukkan akurasi masing-masing kata sebesar 90% untuk atap, 71% untuk kapak, dan 90% untuk paku. Akurasi keseluruhan sistem pengenalan adalah 84%. Dengan adanya sistem ini diharapkan dapat menjadi Solusi bagi orang terdekat penderita dalam mengoreksi ucapan penderita terkait huruf P.

**Kata kunci:** Sumbing bibir, pengenalan suara, *K-nearest Neighbour*

## Abstract

The failure of the two sides of the oral cavity to fuse properly during fetal development results in a congenital abnormality known as a cleft palate (*palatoschisis*) or a cleft lip (*labioschisis*). Cleft lip can manifest in two distinct forms: a complete cleft lip, which involves abnormalities in the gums, alveolar processes, palate (*palatum*), and nostrils, and an incomplete cleft lip. Indonesia, as a Southeast Asian country, ranks seventh globally in the prevalence of cleft lip and palate, with 20.4% of the population affected. According to the Ministry of Health of the Republic of Indonesia (2018), the annual prevalence of infants born with cleft lip and palate exceeds 8,900 cases. In Sumbawa Regency, West Nusa Tenggara Province, 21 individuals were registered as prospective patients at Sumbawa Besar General Hospital. This community service initiative aimed to implement a Speech Recognition System for individuals with cleft lip, with a specific focus on the articulation of the phoneme /p/. Speech data consisting of the words atap (roof), paku (nail), and kapak (axe) were collected from seven individuals with cleft lip in Sumbawa Regency, with each word repeated 15 times. The K-nearest neighbor (KNN) algorithm was employed to evaluate the system's performance. Experimental results demonstrated recognition accuracies of 90% for atap, 90% for paku, and 71% for kapak, yielding an overall system accuracy of 84%. This system holds the potential to serve as a corrective tool for individuals in close proximity to cleft lip patients, aiding in improving speech articulation related to the phoneme /p/.

**Keyword:** Cleft palate, speech recognition, *K-nearest Neighbour*

## I. PENDAHULUAN

Kabupaten Sumbawa dengan 24 kecamatan merupakan salah satu kota kabupaten yang terletak di Provinsi Nusa Tenggara Barat, dengan jumlah penduduk 509,75 Juta jiwa (BPS Sumbawa, 2021) tercatat 21 orang menderita bibir sumbing yang tersebar di beberapa kecamatan. Penyebab

kelainan sumbing bibir dan langit-langit sampai saat ini belum dapat dipastikan. Akan tetapi beberapa hasil studi menunjukkan penyebab terpenting terjadinya kelainan sumbing bibir dan langit-langit bersifat kompleks dan multifaktorial yang melibatkan faktor genetik, lingkungan dan interaksi antara genetik dengan lingkungan. Risiko terjadinya sumbing bibir dan langit-langit adalah jika terdapat riwayat

keluarga positif menderita kelainan ini maka orang tua yang terkena memiliki peluang 3-5% untuk melahirkan anak yang terkena, dan jika terdapat anak yang mengalami kecacatan maka orang tua memiliki peluang 40% untuk melahirkan anak yang lain mengalami kecacatan (Bender,2000). Kelainan pada bibir ini menjadi sebab kesulitan berkomunikasi, karena suara menjadi bagian penting dalam teknologi seperti pengenalan suara pembicara dan pada investigasi forensic (Nelson,2000). Selain itu masalah yang ditimbulkan akibat cacat ini ialah masalah fungsi, estetik dan psikis, ketiganya saling berhubungan yang paling penting ialah mengenai asupan gizi penderita sumbing bibir, apabila terdapat celah pada bibir dan gusi maka penderita tidak bisa menyusui dengan baik, sehingga asupan nutrisi dapat terganggu. Selain itu juga masalah pada penampilan wajah dari pasien sumbing yang dapat mempengaruhi psikis penderita sumbing bibir tersebut dan pengaruhnya terhadap kehidupan social (Supandi, 2013). Penanganan penderita bibir sumbing dapat berjalan dengan baik jika dilakukan oleh ahli dari berbagai disiplin keilmuan (Bisono,2018). Usaha Rehabilitasi pada penderita bibir sumbing secara menyeluruh membutuhkan pengobatan dan perhatian khusus dalam waktu yang cukup lama hingga bertahun-tahun dari tim yang terdiri dari dokter bedah, dokter umum, dokter anak, ahli bedah plastic untuk tindakan operasi, ahli THT untuk penanganan jika terjadi kelainan pada telinga dan lainnya. Selain itu juga dibutuhkan dokter gigi anak, prostodontis, dan ortodontis yang akan mengawasi perkembangan rahang dan gigi, ahli terapi wicara yang mengawasi dan membimbing kemampuan bicara, ahli psikologi dan psikiater anak untuk menangani masalah psikologis yang timbul (Sjamsuhidayat,2005).

Speech recognition adalah kemampuan komputer untuk memahami dan mengidentifikasi suara seseorang, terlepas dari siapa yang berbicara. Implementasi speech recognition contohnya adalah mengeksekusi program yang ada di komputer memanfaatkan bantuan perintah suara (Prayoga,2019). Fungsi Automatic Speech Recognition (ASR) adalah sebagai alat konversinversi bahasa ucapan ke dalam beberapa data (Levis,2012). "PythonVAs" (kependekan dari Python Voice Assistant) adalah aplikasi yang dapat membantu pengguna untuk mempermudah pekerjaan sehari-hari di komputer desktop dimana *Library Speech Recognition* dari bahasa pemrograman Python dapat mengenali ucapan Bahasa Inggris dengan menggunakan bantuan Google dan *Microsoft Speech API* (Sjamsuhidayat,2005).

Teknologi pengenalan suara diimplementasikan ke aplikasi asisten virtual pribadi berbasis desktop dengan tujuan membantu menyederhanakan tugas sehari-hari melalui komputer. Hasil penelitian ini menghasilkan sebuah aplikasi

pengenalan suara (speech recognition) untuk menjalankan tugas umum tertentu di desktop dan menyalakan perangkat elektronik. Dari pengujian yang sudah dilakukan, aplikasi dapat mengenali perintah suara untuk menjalankan perintah yang diberikan dan dapat menyalakan ataupun mematikan perangkat listrik seperti yang diharapkan.

## II. METODE

Metode pelaksanaan pengenalan suara pada penderita bibir sumbing di Kabupaten Sumbawa ini dilakukan dalam beberapa tahapan yaitu :

1. Pengumpulan data
2. Pre proses ekstraksi data
3. Ekstraksi data input
4. Pengenalan suara



**Gambar 1.** Metode Pelaksanaan

Keterangan Gambar 1

1. Pengumpulan data: Kumpulan data yang digunakan dalam penelitian ini dikumpulkan dengan bantuan 8 relawan yang memiliki kondisi sumbing, seperti CL, CP, atau CL/P, berusia antara 4 sampai dengan 45 tahun. Setiap orang mengucapkan tiga kata yang mengandung fonem /p/: 'Paku', 'Kapak', dan 'Atap'. Setiap kata diartikulasikan 10 kali, menghasilkan 240 sampel ucapan. Sampel ini, direkam dalam bahasa Indonesia, dan untuk meminimalkan kebisingan lingkungan, sinyal ucapan ditangkap di dalam ruangan menggunakan smartphone.



**Gambar 2.** Proses Pengambilan Data

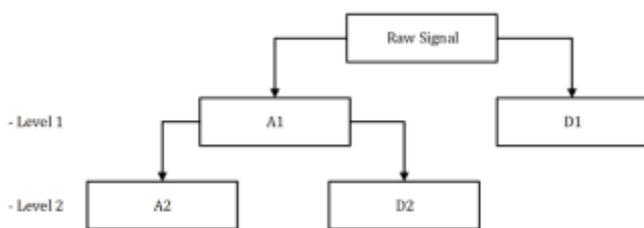
2. Pra Proses

Selama tahap ini, sinyal ucapan mengalami peningkatan untuk meminimalkan kebisingan melalui pra-penekanan. Teknik ini meningkatkan kualitas sinyal ucapan pada frekuensi yang lebih tinggi. Persamaan 1 mengilustrasikan proses pra-penekanan. Sinyal yang ditingkatkan diwakili oleh  $y[n]$ , dan  $x[n]$  mewakili sinyal input, sementara  $x[n-1]$  mewakili sinyal sebelumnya— nilai  $\alpha=0,97$  yang digunakan dalam proses ini.

$$y[n] = x[n] - \alpha * x[n - 1] \tag{1}$$

3. Ekstraksi Data

Ekstraksi fitur melibatkan turunan fitur dari sinyal input. Fitur-fitur ini berfungsi untuk mengkarakterisasi sinyal. Dalam transformasi wavelet diskrit, penggambaran sebuah skala waktu sinyal digital didapatkan dengan menggunakan teknik filterisasi digital. Secara garis besar proses dalam teknik ini adalah dengan melewati sinyal yang akan dianalisis pada filter dengan frekuensi dan skala yang berbeda (Al Fatwa,2009). Sinyal diuraikan pada tahap ini menggunakan Discrete Wavelet Transformation (DWT). Teknik ini dapat melokalisasi sinyal dalam domain frekuensi waktu dan memiliki sifat multiresolusi. Transformasi Wavelet membagi sinyal menjadi dua sub-band, sub-band Approximation dan Detail, pada tingkat dekomposisi pertama. Pada tingkat kedua, perkiraan dibagi lagi menjadi dua sub-band. Prosedur ini berlanjut sampai jumlah tingkat dekomposisi yang diinginkan tercapai. Gambar 3 menunjukkan bagaimana dekomposisi 2 level.



Gambar 3. Dekomposisi 2 Level

4. Pengenalan Suara (Speech Recognition)

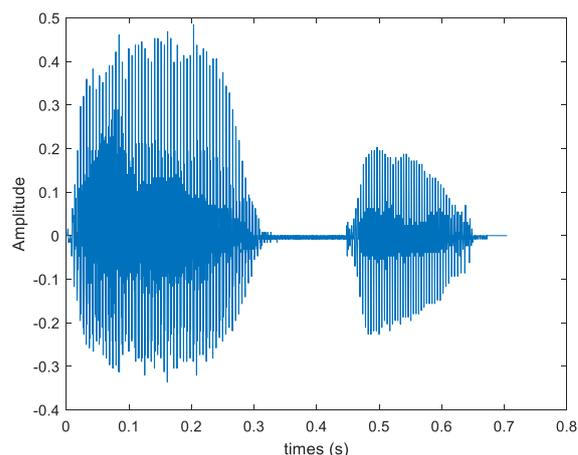
K-Nearest Neighbor adalah teknik klasifikasi yang diawasi. Metode ini mengklasifikasikan objek berdasarkan hasil perbandingan. Algoritma mencari k sampel di antara sampel terdekat. Dalam pendekatan ini, instance baru ditetapkan ke kelas yang menerima suara terbanyak dari K tetangga terdekatnya (Dujaili,2021). KNN memiliki beberapa metrik untuk mengukur jarak antar benda, termasuk jarak Euclidean. Persamaan 2

adalah bentuk matematis dari jarak Euclidean (Venkata dkk,2022).

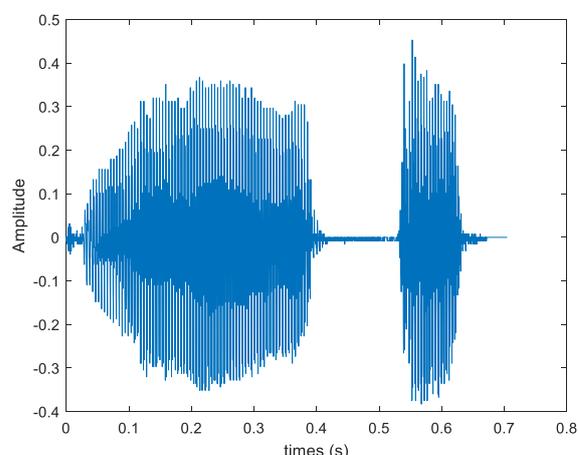
$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2} \tag{2}$$

III. HASIL DAN PEMBAHASAN

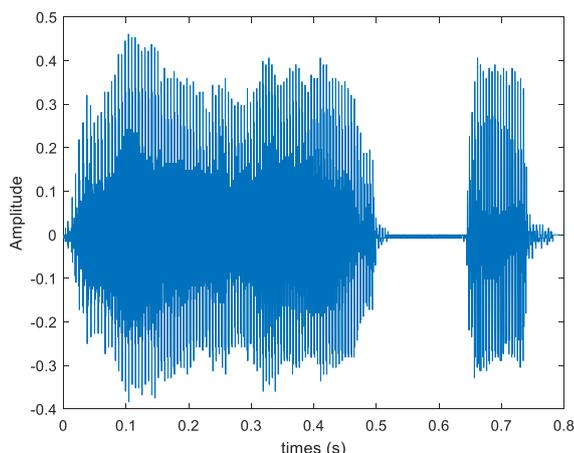
Dataset diperoleh dengan bantuan 7 relawan dengan kondisi sumbing, baik CL, CP, atau CL/P. Penelitian ini menggunakan tiga kata: 'paku', 'kapak', dan 'atap'. Setiap sukarelawan mengulangi kata-kata tersebut lima belas kali, menghasilkan 315 kumpulan data. Tahap selanjutnya adalah meningkatkan kualitas setiap sinyal ucapan, untuk mencapai tujuan ini, kami menggunakan pre-stress dengan  $\alpha = 0,97$ . Metode ini akan meningkatkan kualitas sinyal pada frekuensi tinggi. Selanjutnya, sinyal yang ditingkatkan dinormalisasi untuk memastikan amplitudo maksimum adalah 1. Sinyal input dari setiap kata diilustrasikan pada gambar 4.



Gambar 4. Sinyal Kata Paku



Gambar 5. Sinyal Kata Kapak



Gambar 6. Sinyal Kata Atap

Sinyal kemudian diuraikan menjadi dekomposisi empat tingkat menggunakan wavelet Reverse Biorthogonal 1.1. Selanjutnya, energi diekstraksi dari lima sub-band di setiap sinyal dari setiap wavelet induk. Hasil dekomposisi setiap induk wavelet ditunjukkan pada Tabel 1.

Table 1. The Energy of Reverse Biorthogonal 1.1 decomposition

Samples	energy				
	A4	D1	D2	D3	D4
Paku	0.7161422729	0.0005622085184	0.003393413765	0.01787455423	0.09806796483
Paku	0.6149955047	0.0003130549173	0.002127421334	0.01396396442	0.08442529879
Kapak	0.3183470986	0.0001437235624	0.001071194085	0.007649641687	0.05192271146
Kapak	0.3066396077	0.0001444710915	0.0009045658509	0.005749656757	0.03767786026
Atap	0.2934662725	0.0002838276379	0.00201322621	0.01313840503	0.07322545119
Atap	0.2934662725	0.0002838276379	0.00201322621	0.01313840503	0.07322545119

Tabel 1 menunjukkan distribusi energi sampel yang berbeda menggunakan dua jenis dekomposisi wavelet: Reverse Biorthogonal 1.1. Dekomposisi wavelet adalah metode yang digunakan untuk menganalisis komponen frekuensi suatu sinyal. Setiap tingkat dekomposisi (A4, D1, D2, D3, D4) sesuai dengan pita frekuensi yang berbeda, dengan A4 biasanya mewakili koefisien perkiraan (frekuensi rendah) dan D1 hingga D4 mewakili koefisien detail (frekuensi lebih tinggi).

Tabel menunjukkan distribusi energi sinyal suara berdasarkan menggunakan metode Reverse Biorthogonal 1.1 pada kata-kata "Paku," "Kapak," dan "Atap." Hasil eksperimen menunjukkan bahwa energi terbesar terkonsentrasi pada koefisien aproksimasi (A4), yang merepresentasikan komponen frekuensi rendah dari sinyal. Energi pada level detail (D1 hingga D4) menurun secara bertahap, mencerminkan kontribusi frekuensi tinggi yang semakin kecil. Pola ini konsisten di semua sampel, menunjukkan bahwa informasi dominan dari sinyal suara manusia cenderung berada pada frekuensi rendah.

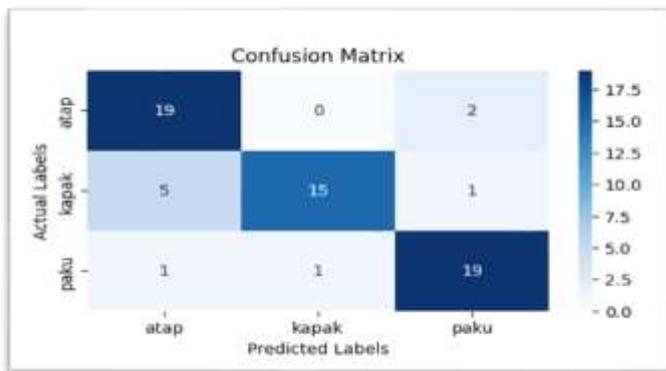
Meskipun pola distribusi energi serupa di antara kata-kata, terdapat perbedaan nilai energi pada tiap level. Kata "Paku" memiliki energi lebih tinggi dibandingkan "Kapak" dan "Atap," khususnya pada D1 hingga D4, yang dapat mengindikasikan perbedaan karakteristik akustik akibat

artikulasi suara. Sebaliknya, distribusi energi pada kata "Atap" terlihat lebih stabil di semua level dekomposisi, mencerminkan pola artikulasi yang lebih seragam.

Pada fase pengenalan, fitur energi yang dihitung dimasukkan ke KNN, di mana K=1. Berdasarkan percobaan, hasil yang diperoleh ditunjukkan pada Tabel 3 dan 4, sedangkan matriks kebingungan dari setiap hasil pengenalan ditunjukkan pada gambar 5 dan 6. Hasil ini dievaluasi menggunakan metrik presisi, penarikan, skor F1, dan akurasi.

Laporan klasifikasi:				
	precision	recall	f1-score	support
atap	0.76	0.90	0.83	21
kapak	0.94	0.71	0.81	21
paku	0.86	0.90	0.88	21
accuracy			0.84	63
macro avg	0.85	0.84	0.84	63
weighted avg	0.85	0.84	0.84	63
Akurasinya untuk kelas atap: 0.90				
Akurasinya untuk kelas kapak: 0.71				
Akurasinya untuk kelas paku: 0.90				

Gambar 7. Confusion matrix on Reverse Biorthogonal 1.1 Decomposition Recognition



**Gambar 8.** Recognition results on Reverse Biorthogonal 1.1 Decomposition

#### IV. KESIMPULAN

Pengabdian Masyarakat ini mengeksplorasi efektivitas transformasi wavelet dalam ekstraksi fitur untuk pengenalan sinyal ucapan sumbing. Kumpulan data, yang terdiri dari 315 sampel dari tiga kata ('paku', 'kapak', dan 'atap') yang diucapkan oleh 7 sukarelawan dengan kondisi sumbing, diproses menggunakan dekomposisi Reverse Biorthogonal 1.1. Hasilnya menunjukkan bahwa distribusi energi di berbagai tingkat dekomposisi bervariasi secara signifikan, menyoroti komponen frekuensi berbeda yang ditangkap oleh masing-masing metode. Fase pengenalan, menggunakan KNN, menunjukkan bahwa fitur yang dihasilkan oleh Reverse Biorthogonal 1.1 menunjukkan hasil memuaskan, dengan akurasi 90% pada kata atap dan paku, sedangkan pada kata kapak akurasi hanya mencapai 70%. Hal ini dapat disebabkan karena berbagai faktor, salah satunya adalah kualitas data pelatihan untuk "kapak," yang kemungkinan memiliki variasi akustik yang lebih kecil, sehingga menyebabkan model kurang mampu mengenali karakteristik spesifik kata ini. Pengabdian Masyarakat ini dapat diterapkan di daerah lain sehingga dapat membantu mempermudah penderita dalam melatih pengucapan kata dengan Fonem lainnya. Terakhir, memperluas kumpulan data untuk memasukkan sampel ucapan yang lebih beragam dan melibatkan kelompok sukarelawan yang lebih besar dapat membantu menggeneralisasi temuan dan meningkatkan penerapan metode yang diusulkan dalam pengaturan klinis untuk diagnosis dan pengobatan gangguan bicara yang lebih baik yang terkait dengan kondisi sumbing.

#### DAFTAR PUSTAKA

- Alfatwa, Watermarking Pada Citra Digital Menggunakan Discrete Wavelet Transform. Bandung: Institute Teknologi Bandung, 2009.
- Al Dujaili MJ, Ebrahimi-Moghadam A, Fatlawi A. Speech emotion recognition based on SVM and KNN classifications fusion. *IJECE*. 2021 Apr 1;11(2):1259.
- Bender PL. Genetics of cleft lip and palate. *J Pediatr Nurs*. 2000;15: 242–249. [://web-api.bps.go.id/download](http://web-api.bps.go.id/download). Hal 51
- Bisono. Sumbing bibir/langitan. Dalam. 2018: Reksoprodjo S, editor. Kumpulan kuliah ilmu bedah. Tangerang: Binarupa Aksara. Hal 393-6.
- Hidayat S, Tajuddin M, Yusuf SAA, Qudsi J, Jaya NN. Wavelet Detail Coefficient As a Novel Wavelet-Mfcc Features in Text-Dependent Speaker Recognition System. *IIUM Engineering Journal*. 2022;23(1):68–81.
- KABUPATEN SUMBAWA DALAM ANGKA SUMBAWA REGENCY IN FIGURES 2021. Badan Pusat Statistik Kabupaten Sumbawa. Hal xlv
- Kemendes RI. (2018). InfoDATIN: Kelainan bawaan. Pusat Data Dan Informasi Kementerian Kesehatan RI, 1–6.
- J. Levis and R. Suvorov, "Automatic speech recognition," *The encyclopedia of applied linguistics*, pp. 1–8, 2012.
- Nelson. Celah bibir dan palatum. Dalam: Wahab AS, editor. Ilmu kesehatan anak. Jakarta: EGC; 2000;2(15):1282
- N. F. I. Prayoga, "Analisis Speaker Recognition Menggunakan Metode Dynamic Time Warping (DTW) Berbasis Matlab," *AVITEC*, vol. 1, no. 1, Agu 2019, doi: 10.28989/avitec.v1i1.492
- Sjamsuhidayat R, de Jong W. Buku ajar ilmu bedah. Ed ke-2. Jakarta: EGC; 2005. Hal 344-5.
- Setyawan Y, Implementasi Speech Recognition untuk Asisten Virtual dengan Python *JURNAL ALGOR - VOL. IV NO. 1 (2022)*
- Supandi, A. ANGKA KEJADIAN SUMBING BIBIR DI RSUP Prof. Dr. R. D. KANDOU MANADO PERIODE 2011-2013. *E-Journal Unsrat*. Hal.6
- Venkata Subbarao M, Terlupu SK, Geethika N, Harika KD. Speech Emotion Recognition Using K-Nearest Neighbor Classifiers. In: Shetty D. P, Shetty S, editors. *Recent Advances in Artificial Intelligence and Data Engineering [Internet]*. Singapore: Springer Singapore; 2022 [cited 2024 Jun 18]. p. 123–31. (*Advances in Intelligent Systems and Computing*; vol. 1386). Available from: [https://link.springer.com/10.1007/978-981-16-3342-3\\_10](https://link.springer.com/10.1007/978-981-16-3342-3_10)