SINTA Jurnal Sistem Informasi dan Teknologi Komputasi
Volume 3 Nomor 3 Juli 2026 [page: 258-266]
DOI: 10.61124/sinta.v3i4.398

KLASIFIKASI SUARA PARU – PARU YANG DITINGKATKAN MENGGUNAKAN SPECAUGMENTASI DAN MEL SPECTROGRAM

Jaenal Arifin*[1], Dony Hutabarat[2], Moh Nur Shodiq[3]
[1]Program Studi Teknik Elektro, Universitas Telkom Purwokerto, Indonesia
[2]Program Studi Fisika, Universitas Islam Negeri Sultan Maulana Hasanuddin Banten
[3]Program Studi Teknologi Rekayasa Komputer, Politeknik Negeri Banyuwangi, Banyuwangi.
*e-mail: jaenala@telkomuniversity.ac.id

Abstrak
Penyakit paru-paru merupakan kontributor utama morbiditas dan mortalitas di seluruh dunia. Suara napas pada individu dapat bervariasi dari keadaan normal hingga patologis. Studi ini bertujuan untuk meningkatkan klasifikasi menggunakan *SpecAugment* dan *Mel Spectrogram* dengan validasi silang 10 fold. Dataset dikumpulkan dari database Kaggle, dataset *International Conference on Biomedical and Health Informatics* (ICBHI), dan dataset Rumah Sakit Fortis India. Dataset suara paru-paru terdiri dari 1363 pasien. Pra-pemrosesan meliputi konversi data audio ke sensor float, frekuensi sampling 16 kHz, dan transformasi ke Mel Spectrogram. Augmentasi data dilakukan menggunakan *google brain specAugment*. Dalam studi ini, kami menggunakan validasi silang 10 fold. Hasil eksperimen memperoleh nilai akurasi terbaik sebesar 95.72%, presisi 96.06%, recall 95.71%, F1-Score 95.74%, dan AUC 99.51%. Studi ini berkontribusi pada penelitian AI di bidang elektromedis dan pemrosesan sinyal. Peningkatan dataset suara paru-paru dengan menambahkan spesifikasi meningkatkan akurasi dan keandalan diagnosis gangguan suara paru-paru.

Kata kunci: Klasifikasi Suara Paru-Paru; Mel Spectrogram; SpecAugmentasi; Validasi Silang 10 Fold.

Abstract
Lung diseases are major contributors to morbidity and mortality worldwide. Breath sounds in individuals can vary from normal to pathological states. This study aims to improve the classification using SpecAugment and Mel Spectrogram with 10-fold cross-validation. The datasets were collected from the Breath Sounds database (Kaggle), International Conference on Biomedical and Health Informatics (ICBHI) dataset, and Fortis Hospital India dataset. The lung sound dataset comprised 1363 patients. Pre-processing includes the conversion of audio data to a float sensor, 16 kHz sampling frequency, and transformation to Mel Spectrogram. Data augmentation was performed using the Google Brain SpecAugment. In this study, we used 10-fold cross-validation. The experimental results obtained the best accuracy value of 95.72%, precision of 96.06%, recall of 95.71%, F1-Score of 95.74%, and AUC of 99.51%. This study contributes to AI research in electromedicine and signal processing. Enhancing the lung sound dataset by adding specifications improves the accuracy and reliability of lung sound disorder diagnosis.

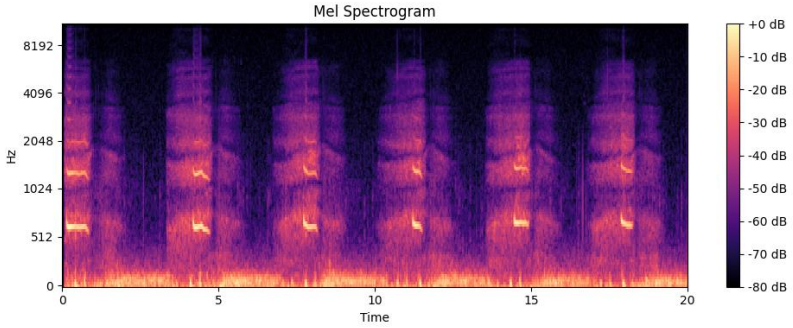
Keywords: Lung Sound Classification; Mel Spectrogram; SpecAugmentation; 10-Fold Cross Validation.

1. PENDAHULUAN

Penyakit paru-paru merupakan masalah kesehatan global dengan angka kematian meningkat setiap tahunnya. Investigasi kelainan suara paru-paru sangat penting. Suara yang dihasilkan oleh sistem pernapasan dan paru-paru memberikan informasi penting mengenai keadaan fisiologis dan patologis manusia. Organisasi Kesehatan Dunia (WHO) memperkirakan bahwa, berdasarkan data yang tersedia penyakit paru obstruktif kronis menjadi penyebab kematian terbesar ketiga pada tahun 2030 [1]. Stetoskop konvensional kesulitan untuk menangkap suara paru-paru untuk evaluasi diagnostik [2]. Stetoskop digital lebih baik kinerjanya dibanding model konvensional [3]. Stetoskop digital dapat menangkap dan menyimpan suara suara paru-paru [4]. Sinyal suara paru-paru dapat digunakan untuk konsultasi, pembelajaran dan kegiatan penelitian. Stetoskop elektronik nirkabel meningkatkan efektivitas pembelajaran pendidikan kedokteran. Kemampuan berbagi suara stetoskop mengurangi waktu yang dibutuhkan untuk diagnosis [5], [6]. Suara paru-paru, sebagai sinyal akustik, memiliki rentang frekuensi 100 Hz sampai 2 kHz [8,9]. Stetoskop ini mendukung *bluetooth* untuk terhubung ke aplikasi Eko yang tersedia di *smartphone* sistem IOS atau Android. Setelah terhubung di aplikasi Eko dapat merekam suara paru-paru yang diambil dari stetoskop [10,11]. Teknik pemrosesan sinyal memiliki peran penting dibidang medis karena dapat membantu mendapatkan kualitas sinyal asli tanpa kehilangan informasi. Ada beberapa alasan mengapa pengambilan sampel frekuensi 16 kHz dilakukan pada penelitian ini.

DOI: 10.61124/sinta.v3i4.398

Gambar 1. Stetoskop digital [7]



Gambar 2. Bentuk mel spektrogram suara paru-paru [8]

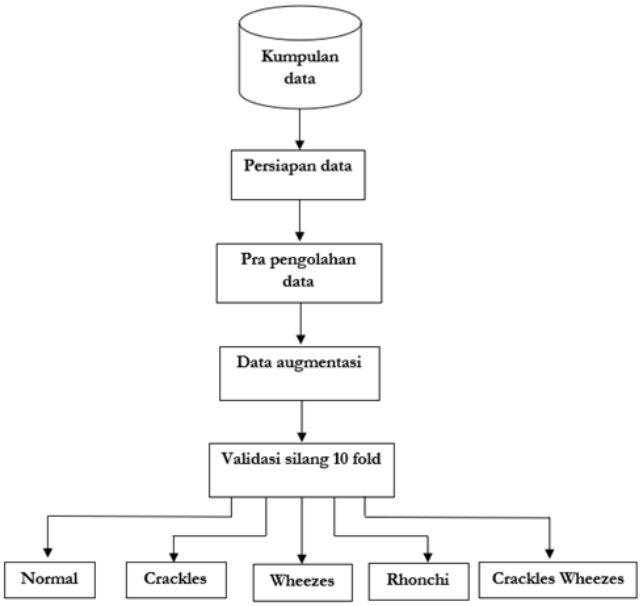
Frekuensi pengambilan sampel yang lebih tinggi menyediakan informasi temporal yang lebih rinci dalam proses penentuan resolusi temporal. Penggunaan frekuensi pengambilan sampel sebesar 16 kHz dapat meminimalkan cacat sinyal (distorsi) dan menghasilkan representasi sinyal yang baik. Frekuensi pengambilan sampel yang lebih tinggi dapat meningkatkan ketelitian representasi sinyal audio sehingga sinyal memiliki tingkat akurasi dan kualitas yang lebih baik. Beberapa peneliti telah mengembangkan mel spectrogram dengan tujuan meningkatkan nilai akurasi dalam mendeteksi dan mengklasifikasikan suara paru-paru [9]. *Mel Spectrogram* mempunyai peran penting dalam meningkatkan akurasi deteksi dan kategorisasi suara paru-paru, karena secara efektif menangkap perubahan spektrum dan informasi frekuensi dalam sinyal suara[10], [11]. *Mel Spectrogram* menawarkan keunggulan dibandingkan representasi spektral lainnya karena desainnya yang selaras dengan sensitivitas pendengaran manusia [12], [13].

Penelitian ini bertujuan untuk mengevaluasi efektivitas *SpecAugment* dalam meningkatkan kinerja klasifikasi suara paru-paru berbasis *mel spectrogram* melalui metode validasi silang 10 fold. Peneliti bertujuan untuk meningkatkan akurasi dan penerapan metode studi kasus untuk mengkategorikan kelainan suara paru-paru. Studi ini berkontribusi dengan menerapkan *specaugment* pada sinyal suara paru-paru untuk mengklasifikasikan lima kelas normal, *crackles*, *wheezes*, *rhonchi*, dan *crackles wheezes*. Penerapan validasi silang 10 fold bertujuan untuk meningkatkan keandalan dan akurasi. Dataset suara paru-paru baru yang dihasilkan dari augmentasi data diharapkan dapat memfasilitasi replikasi ulang dalam inisiatif penelitian di masa mendatang. Penelitian ini memiliki keterbatasan berupa jumlah data yang terbatas yaitu sebanyak 514 sampel pada setiap kelas yang mencakup normal, *crackles*, *wheezes*, *rhonchi*, dan crackles-*wheezes*. Penelitian ini dapat memberikan kontribusi positif pada bidang kecerdasan buatan (AI), elektromedik dan pemrosesan sinyal. Kontribusi utama penelitian ini dibandingkan dengan penelitian sebelumnya terletak pada penerapan *SpecAugment* sebagai metode augmentasi data untuk klasifikasi suara paru-paru ke dalam lima kelas. Penerapan metode tersebut masih relatif terbatas dalam bidang suara medis. *SpecAugment* pada awalnya dikembangkan untuk mendukung penelitian pengenalan ucapan, kemudian penelitian ini mengadaptasi metode tersebut secara inovatif untuk pengolahan dan klasifikasi sinyal suara medis.

DOI: 10.61124/sinta.v3i4.398

2. METODE

Penelitian ini Berikut flowchat metode penelitian yang digunakan seperti digambarkan pada gambar 3. Blok diagram alur penelitian terdiri dari kumpulan dataset suara paru-paru, persiapan pengolahan data, pra pengoalahan data, data augmentasi menggunakan *google brain's spec* augmentasi, validasi silang 10 fold dan klasifikasi lima kelas suara paru-paru seperti suara normal, *crackels*, *wheezes*, *rhonchi* dan *crackles wheezes*.



Gambar 3. Metode Penelitian

1. Pengumpulan Data

Penelitian ini menggunakan dataset suara paru-paru dari tiga sumber, yaitu *dataset International Conference on Biomedical and Health Informatics* (ICBHI) [14], dataset Fortis Hospital India [15], dan database Suara Pernapasan (Kaggle) [16]. Dataset suara paru-paru yang bersumber dari Fortis Hospital India direkam dengan menggunakan stetoskop digital yang terhubung ke laptop melalui *amplifier*. Rekaman suara paru-paru memiliki durasi 17 - 20 detik. Kumpulan data suara paru-paru ini tersedia secara online untuk tujuan penelitian. Dataset suara paru-paru yang bersumber dari Kaggle direkam dengan menggunakan stetoskop digital yang terhubung ke laptop melalui amplifier. Rekaman suara paru-paru memiliki durasi 17 - 25 detik dan disimpan dalam format file .wav. Kumpulan data suara paru-paru dari ICBHI 2017 Challenge terdiri dari berbagai data kelainan seperti bronkitis dan pneumonia. Kategori suara paru-paru sebagai subjek penelitian terdiri dari suara napas normal, *crackles*, *wheezes*, dan campuran suara *crackles* dan *wheezes*.

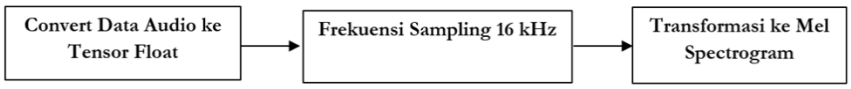
2. Persiapan Data

Kumpulan data suara paru-paru terdiri dari tiga sumber, dengan total 1.363 sampel audio dalam format file.wav. Kumpulan data suara paru-paru terdiri dari 449 sampel normal, 514 sampel suara *crackles*, 232 sampel suara *wheezes*, 52 sampel suara *rhonchi* dan 116 sampel suara *crackles wheezes*. Kumpulan data suara paru-paru berdasarkan tabel 1 memiliki data yang tidak seimbang dari jumlah masing-masing kelas.

Tabel 1. Kumpulan data suara paru-paru [14], [15], [16]

Suara Paru-Paru	Total Samples
Normal	449
Crackles	514
Wheezes	232
Rhonchi	52
Crackles Wheezes	116
Total	1363

3. Pra Pengolahan Data

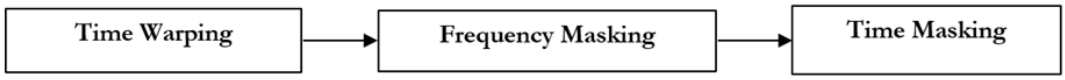


Gambar 4. Tahapan Pra Pengolahan Data [17]

Konversi data audio ke dalam format *tensor-float* merupakan bagian penting dari pengolahan data suara paru-paru untuk penelitian ini. Format ini memungkinkan pengolahan input audio sebagai data numerik. Data membantu dalam mengenali karakteristik suara seperti frekuensi dan amplitudo [17]. Pemilihan frekuensi sampling 16 kHz memiliki alasan kuat yaitu mendapatkan kualitas suara yang diperoleh memenuhi standar yang dapat diandalkan untuk mengidentifikasi karakteristik suara dan pola suara [18]. Penggunaan 16 kHz merupakan praktik yang baik antara kualitas perekaman dan efisiensi pengolahan [19]. Pada frekuensi ini, telah terbukti efektif dalam mengidentifikasi suara dan mengklasifikasikan suara paru-paru [20]. Penelitian ini menggunakan mel spectrogram untuk klasifikasi suara paru-paru karena kemampuannya untuk menangkap karakteristik penting sinyal suara dalam domain frekuensi, yang sesuai dengan persepsi pendengaran manusia[21], [22].

4. Data Augmentasi

SpecAugment google brain merupakan salah satu teknik augmentasi data [23]. Metode ini dirancang khusus untuk data berbasis spektrogram dan dapat digunakan dalam pengenalan sinyal suara, termasuk sinyal suara medis seperti suara paru-paru [24], [25]. Tujuan augmentasi data ini adalah untuk menambahkan variasi pada data pelatihan tanpa mengubah atau menghilangkan informasi penting dalam sinyal. Augmentasi data dalam penelitian ini meliputi *time warping*, *frequency masking*, dan *time masking*. Peneliti telah menggunakan *SpecAugment* sebagai teknik augmentasi data dari *google brain* [26], [27].



Gambar 5. Tahapan Data Augmentasi menggunakan Google Brain [25]

Time warping bekerja dengan mengubah sumbu waktu spektrogram secara acak memilih segmen sumbu waktu, dan menggesernya ke kiri atau kanan. Pendekatan ini mengeksekusi fluktuasi temporal dalam sinyal akustik [28]. *Masking frekuensi* melibatkan penutupan sejumlah pita frekuensi secara acak pada spektrogram. Tahapan ini diharapkan dapat meningkatkan kemampuan spektrogram untuk menangani variasi spektral dan keberadaan adanya *noise* [29], [30]. *Time Masking* melibatkan oklusi acak interval waktu tertentu pada spektrogram. Dengan mencakup segmen waktu cara ini diharapkan ketergantungan pada informasi temporal dapat diminimalkan. Dapat meningkatkan ketahanan model terhadap variasi dan gangguan temporal, termasuk celah waktu singkat dan *noise* yang ada pada sinyal suara paru-paru.

5. Validasi Silang 10 Fold

Teknik validasi silang 10-fold sering digunakan dalam domain *machine learning* dan *deep learning* [31], [32]. Pendekatan validasi silang 10 fold secara konseptual membagi dataset menjadi 10 segmen yang sama (fold), masing-masing berisi sampel yang secara kolektif mencerminkan kumpulan data lengkap [33], [34]. Dalam setiap iterasi, salah satu dari sepuluh segmen kumpulan data digunakan untuk pengujian, sementara segmen yang tersisa diatur dan digunakan sebagai pelatihan. Selama fase iterasi, model dilatih menggunakan data pelatihan dan dievaluasi dengan data pengujian pada sampel yang belum pernah dilihat, yaitu dengan validasi silang. Setiap iterasi didokumentasikan untuk menilai kinerja model, yang diwakili sebagai akurasi, presisi, *recall*, *F1-score*, dan AUC [37]. Setelah menyelesaikan 10 iterasi, nilai akurasi, presisi, recall, F1-score, dan *Area Under the Curve* (AUC) diperoleh sebagai referensi untuk kinerja yang lebih stabil dan andal. Studi ini mengevaluasi kinerja termasuk akurasi, presisi, recall, *F1-score* dan nilai AUC. Kinerja ini didefinisikan dalam Persamaan (1), (2), (3), (4) [38], [39].

DOI: 10.61124/sinta.v3i4.398

$$\text{Accuracy} = \frac{(TP)+(TN)}{(TP)+(TN)+(FP)+(FN)} \quad (1)$$

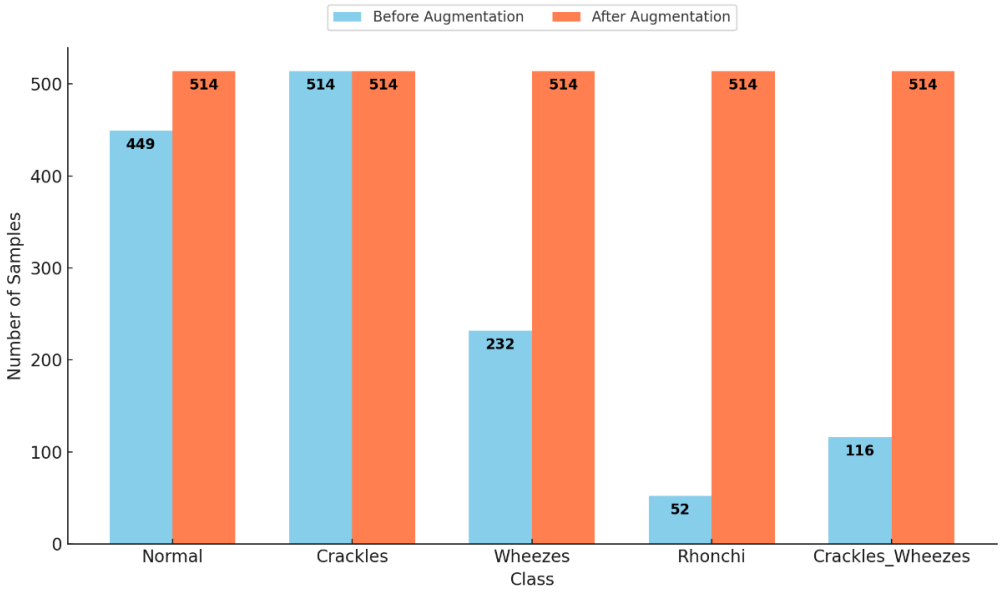
$$\text{Precision} = \frac{(True\ Positives\ (TP))}{(TP)+(FP)} \quad (2)$$

$$\text{Recall} = \frac{True\ Positives(TP)}{(TP)+(FN)} \quad (3)$$

$$\text{F1-Score} = 2X\ \frac{Precision\ X\ Recall}{Precision+Recall} \quad (4)$$

3. HASIL PENELITIAN

1. Hasil Data Augmentasi



Gambar 6. Distribusi Augmentasi Data

2. Hasil Validasi Silang 10 Fold

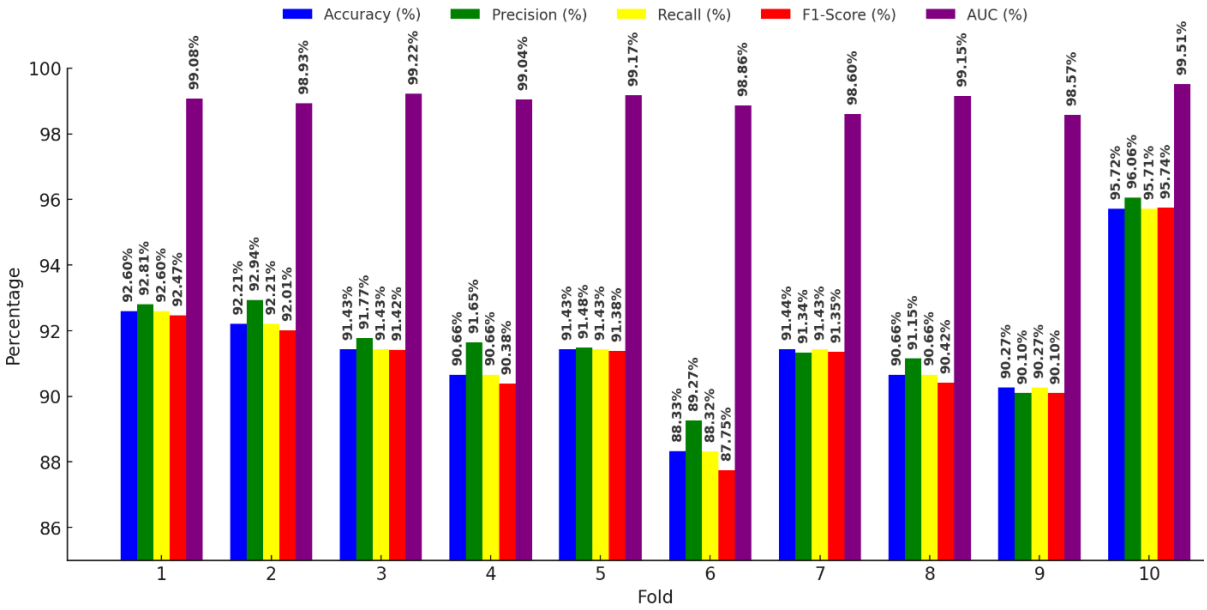
Percobaan yang dilakukan memperoleh hasil nilai matriks seperti akurasi, presisi, *recall*, F1-score, AUC untuk setiap fold. Matriks confusion terbaik dari sepuluh fold tersebut digunakan dalam eksperimen.

Tabel 2. Hasil Validasi Silang 10 Fold

Fold	Akurasi	Precision	Recall	F1-Score	AUC
1	92.60%	92.81%	92.60%	92.47%	99.08%
2	92.21%	92.94%	92.21%	92.01%	98.93%
3	91.43%	91.77%	91.43%	91.42%	99.22%
4	90.66%	91.65%	90.66%	90.38%	99.04%
5	91.43%	91.48%	91.43%	91.38%	99.17%
6	88.33%	89.27%	88.32%	87.75%	98.86%
7	91.44%	91.34%	91.43%	91.35%	98.60%
8	90.66%	91.15%	90.66%	90.42%	99.15%
9	90.27%	90.10%	90.27%	90.10%	98.57%
10	**95.72%**	**96.06%**	**95.71%**	**95.74%**	**99.51%**

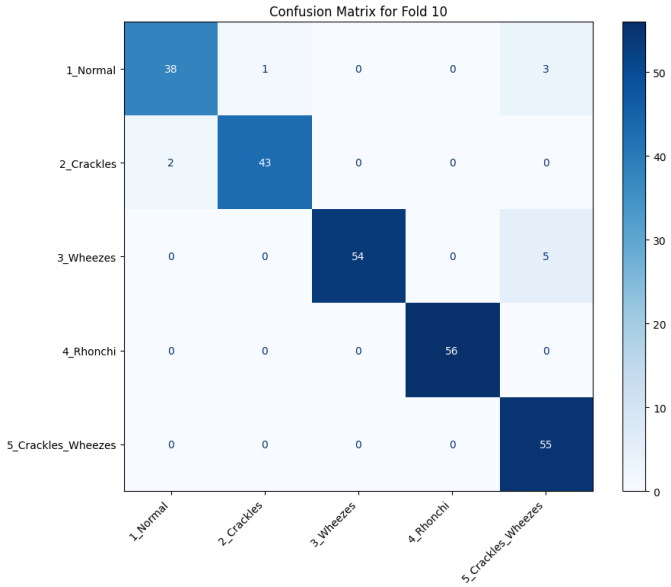
DOI: 10.61124/sinta.v3i4.398

Berdasarkan tabel 2, nilai akurasi sebesar 88.33% - 95.72%. Pada sebagian besar lipatan (fold), nilai akurasi secara konsisten di atas 90%; ini menunjukkan bahwa model stabil dan mengklasifikasikan data dengan benar. Nilai presisi adalah 89.27% - 96.06%. Sebagian besar lipatan konsisten dalam nilai presisi, hampir selalu 90%. Ini menunjukkan bahwa model dapat menghindari kesalahan dalam mendeteksi kelas positif. Nilai recall adalah 88.32% - 95.71%. Nilai recall pada sebagian besar lipatan menunjukkan bahwa model mampu mengenali semua data dalam kasus positif.



Gambar 7. Kinerja Metrik per Fold

Berdasarkan gambar 7, F1-Score terbaik berada di fold 10, dengan nilai 95.74%. F1-score terendah berada di fold 6, dengan nilai 87.75%. F1-score adalah metrik harmonisasi antara nilai recall dan precision dalam satu ukuran, sehingga menunjukkan keseimbangan antara keduanya. Sebagian besar nilai F1-score memiliki nilai di atas 90%, yang menunjukkan bahwa model memiliki kinerja yang solid dalam hal recall dan precision secara bersamaan. Nilai AUC sangat tinggi dan stabil, 98.57%-99.51%. Nilai AUC tertinggi sebesar 99.51% pada fold 10, diikuti 98.57% pada fold 9. Nilai AUC yang konsisten tinggi menunjukkan bahwa model memiliki kinerja yang sangat baik. Semua fold pada nilai AUC mendekati 100%, yang berarti bahwa model selalu tepat dalam memprediksi probabilitas keseluruhan kelas negatif dan positif.



Gambar 8. Confusion Matrik 10 Fold

DOI: 10.61124/sinta.v3i4.398

Berdasarkan gambar 8, terdapat lima kelas berlabel. Kelas suara crackles wheezes sangat akurat dalam mendeteksi, tetapi ada satu kesalahan dalam memprediksi suara crackles wheezes sebagai normal. Kelas rhonchi berkinerja sempurna; tidak ada kesalahan dalam prediksi dan identifikasi. Model dengan mudah membedakan kelas ronki dari kelas lain. Kelas mengi dikenali dengan cukup baik, tetapi ada kesalahan dalam memprediksi mengi sebagai suara berderak mengi. Ini berarti masih ada kemiripan antara kedua kelas; inilah yang membingungkan model. Model kelas suara berderak hampir selalu mendeteksi suara berderak dengan benar, tetapi ada kesalahan dalam memprediksi kelas suara berderak sebagai normal. Kelas suara berderak memiliki nilai recall dan presisi yang tinggi. Model kelas normal bagus dalam mendeteksi normal, tetapi ada sedikit kesalahan. Secara keseluruhan, model memiliki kinerja yang sangat baik, dengan akurasi tinggi di sebagian besar kelas. Pekerjaan selanjutnya perlu meningkatkan kelas normal, wheezes, dan suara crackles wheezes agar kinerja model lebih merata dan optimal.

4. DISKUSI

Akurasi tertinggi diperoleh pada fold ke-10, yaitu sebesar 95.72%, sedangkan akurasi terendah diperoleh pada fold ke-6, yaitu sebesar 88.33%. Selisih akurasi tertinggi dan terendah adalah 7.39 poin persentase. Perbedaan tersebut menunjukkan bahwa komposisi data pada setiap fold masih memberikan pengaruh terhadap performa model. Meskipun terdapat variasi antar-fold, standar deviasi akurasi hanya sebesar 1.90%. Nilai ini relatif kecil dibandingkan rata-rata akurasinya. Dengan demikian, model dinilai cukup stabil karena perubahan data pelatihan dan validasi pada setiap fold tidak menyebabkan perubahan performa yang sangat besar. Rendahnya performa pada fold ke-6 dapat disebabkan oleh beberapa kemungkinan, seperti adanya sampel dengan pola suara yang lebih sulit dibedakan, kualitas rekaman yang kurang baik, ketidakseimbangan jumlah sampel antarkelas. Precision tertinggi terdapat pada fold ke-10 sebesar 96.06%, sedangkan precision terendah terdapat pada fold ke-6 sebesar 89.27%. Standar deviasi precision sebesar 1.84% juga menunjukkan bahwa tingkat ketepatan prediksi model relatif konsisten pada seluruh fold. Kondisi ini menunjukkan bahwa model cenderung sedikit lebih baik dalam menjaga ketepatan prediksi dibandingkan menemukan seluruh sampel yang seharusnya termasuk dalam suatu kelas. Pada klasifikasi suara paru-paru, nilai precision yang tinggi penting menunjukkan bahwa model tidak terlalu sering menetapkan suatu jenis suara paru-paru pada sampel yang sebenarnya berasal dari kelas lain. Akan tetapi, interpretasi yang lebih rinci tetap memerlukan nilai precision masing-masing kelas.

Nilai AUC merupakan hasil yang paling tinggi dan paling stabil dalam pengujian ini. Rata-rata AUC mencapai 99.01%, dengan nilai minimum 98.57% dan nilai maksimum 99.51%. Standar deviasi AUC hanya sebesar 0.29%, sedangkan selisih antara nilai tertinggi dan terendah hanya sebesar 0.94 poin persentase. Nilai AUC yang mendekati 100% menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat baik. Artinya, berdasarkan skor probabilitas yang dihasilkan, model dapat memisahkan sampel dari kelas yang berbeda dengan baik pada berbagai nilai ambang klasifikasi. AUC tertinggi diperoleh pada fold ke-10 sebesar 99.51%, sedangkan AUC terendah diperoleh pada fold ke-9 sebesar 98.57%. Kondisi ini menunjukkan bahwa kemampuan model dalam membedakan sampel berdasarkan probabilitas masih sangat baik. Pekerjaan penelitian berikutnya dapat dilakukan membandingkan Mel Spectrogram dengan representasi sinyal audio lainnya untuk mengetahui representasi yang paling sesuai dari penelitian klasifikasi suara paru-paru. Representasi sinyal audio mencakup MFCC, chrogram, spectral contracst, wavelet transform, representasi sinyal audio mentah dan embedding sinyal audio yang telah dilatih sebelumnya.

5. KESIMPULAN

Berdasarkan hasil validasi silang 10 fold memiliki skor terbaik pada semua matriks penilaian. Model menghasilkan akurasi terbaik sebesar 95.72%, *precision* sebesar 96.06%, *recall* sebesar 95.71%, *F1-score* sebesar 95.74%, dan AUC sebesar 99.51%. Ini menunjukkan bahwa model sesuai harapan dan dapat diandalkan. Keseimbangan kinerja baik, dengan nilai *F1-score* yang tinggi dan nilai yang konsisten antara *recall* dan *precision*. Ini penting untuk implementasi aplikasi tertentu, yaitu keseimbangan antara deteksi positif dan penghindaran *false positive*. Nilai AUC memiliki skor yang sangat tinggi. Ini menunjukkan bahwa kinerjanya hampir sempurna. Sangat efektif untuk kebutuhan aplikasi yang memerlukan pembagian kelas yang akurat.

DAFTAR PUSTAKA

[1] World Health Organization, "The top 10 causes of death," https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[2] X. Liao *et al.*, "Automated detection of abnormal respiratory sound from electronic stethoscope and mobile phone using MobileNetV2," *Biocybern. Biomed. Eng.*, vol. 43, no. 4, pp. 763–775, Oct. 2023, doi: 10.1016/j.bbe.2023.11.001.

[3] Y. Y. Ang, L. R. Aw, V. Koh, and R. X. Tan, "Characterization and cross-comparison of digital stethoscopes for telehealth remote patient auscultation," *Med. Nov. Technol. Devices*, vol. 19, Sep. 2023, doi: 10.1016/j.medntd.2023.100256.

[4] N. Tharapecharat *et al.*, "Digital stethoscope with processing and recording based on cloud," in *BMEiCON 2023 - 15th Biomedical Engineering International Conference*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/BMEiCON60347.2023.10322053.

[5] X. Li, Y. Shang, J. Wei, and Y. Zhou, "Research on electronic stethoscope system and signal processing algorithm," in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2634/1/012037.

[6] Y. S. Rao, B. Mehta, A. Kathiriya, A. Dharvia, D. Pancholi, and H. Patel, "Smart Stethoscope for Remote Healthcare Monitoring Using IoT: A Cost-Effective Solution," in *2023 3rd Asian Conference on Innovation in Technology, ASIANCON 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ASIANCON58793.2023.10270283.

[7] J. J. Seah, J. Zhao, D. Y. Wang, and H. P. Lee, "Review on the Advancements of Stethoscope Types in Chest Auscultation," *Diagnostics*, vol. 13, no. 9, pp. 1–19, May 2023, doi: 10.3390/diagnostics13091545.

[8] J. S. Park, K. Kim, J. H. Kim, Y. J. Choi, K. Kim, and D. I. Suh, "A machine learning approach to the development and prospective evaluation of a pediatric lung sound classification model," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-27399-5.

[9] A. Roy and U. Satija, "ILDNet: A Novel Deep Learning Framework for Interstitial Lung Disease Identification Using Respiratory Sounds," in *2024 International Conference on Signal Processing and Communications, SPCOM 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/SPCOM60851.2024.10631581.

[10] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Bandara, and I. D. L. T. Diez, "Lung Sound Classification with Multi-Feature Integration Utilizing Lightweight CNN Model," *IEEE Access*, vol. 12, pp. 21262–21276, Feb. 2024, doi: 10.1109/ACCESS.2024.3361943.

[11] C. Wu, D. Lei, and Z. Xu, "Respiratory Disease Classification Model Based on Feature Fusion," in *2023 4th International Conference on Intelligent Computing and Human-Computer Interaction, ICHCI 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 148–155. doi: 10.1109/ICHCI58871.2023.10277774.

[12] M. Chaiani, S. A. Selouani, and M. Boudraa, "Voice Disorder Detection Using Enhanced Auditory Perception-Scaled Spectrograms," in *2022 45th International Conference on Telecommunications and Signal Processing, TSP 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 49–54. doi: 10.1109/TSP55681.2022.9851379.

[13] R. Islam and M. Tarique, "Spectrogram and Mel-Spectrogram Based Dysphonic Voice Detection Using Convolutional Neural Network," in *International Conference on Electrical, Computer, and Energy Technologies, ICECET 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICECET61485.2024.10698112.

[14] ICBHI Challenge, "ICBHI 2017 Challenge - Respiratory Sound Database." Accessed: Aug. 22, 2024. [Online]. Available: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge

[15] N. Baghel, V. Nangia, and M. K. Dutta, "ALSD-Net: Automatic lung sounds diagnosis network from pulmonary signals," *Neural Comput. Appl.*, vol. 33, no. 24, pp. 17103–17118, Dec. 2021, doi: 10.1007/s00521-021-06302-1.

[16] https://www.kaggle.com/datasets/vbookshelf/respiratory-sound-database, "Respiratory Sound Database."

[17] L. D. Yang, R. B. Yue, J. Wang, and M. Liu, "Neural Network Model Based on the Tensor Network for Audio Tagging of Domestic Activities," *Front. Phys.*, vol. 10, Apr. 2022, doi: 10.3389/fphy.2022.863291.

[18] R. Mahum, A. Irtaza, A. Javed, H. A. Mahmoud, and H. Hassan, "DeepDet: YAMNet with BottleNeck Attention Module (BAM) TTS synthesis detection," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1, pp. 1–16, Dec. 2024, doi: 10.1186/s13636-024-00335-9.

[19] N. H. Valliappan, S. D. Pande, and S. Reddy Vinta, "Enhancing Gun Detection With Transfer Learning and YAMNet Audio Classification," *IEEE Access*, vol. 12, pp. 58940–58949, 2024, doi: 10.1109/ACCESS.2024.3392649.

[20] A. Fava *et al.*, "Pre-processing techniques to enhance the classification of lung sounds based on deep learning," *Biomed. Signal Process. Control*, vol. 92, Jun. 2024, doi: 10.1016/j.bspc.2024.106009.

[21] R. Gupta, R. Singh, C. M. Travieso-González, R. Burget, and M. Kishore Dutta, "DeepRespNet: A deep neural network for classification of respiratory sounds," *Biomed. Signal Process. Control*, vol. 93, Jul. 2024, doi: 10.1016/j.bspc.2024.106191.

[22] K. N. Lal, "A lung sound recognition model to diagnoses the respiratory diseases by using transfer learning," *Multimed. Tools Appl.*, vol. 82, no. 23, pp. 36615–36631, Sep. 2023, doi: 10.1007/s11042-023-14727-0.

[23] H. Wang, Y. Zou, and W. Wang, "Specaugment++: A hidden space data augmentation method for acoustic scene classification," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, International Speech Communication Association, 2021, pp. 21–25. doi: 10.21437/Interspeech.2021-140.

[24] G. Shanthakumari and E. Priya, "Interpretation of Lung Sounds Using Spectrogram-Based Statistical Features," in *Futuristic Communication and Network Technologies*, A. Sivasubramanian, P. N. Shastry, and P. C. Hong, Eds., Singapore: Springer Nature Singapore, 2022, pp. 815–823.

[25] J. Ramonaite and G. Korvel, "Noisy Phoneme Recognition Using 2D Convolution Neural Network," in *2023 IEEE 10th Jubilee Workshop on Advances in Information, Electronic and Electrical Engineering, AIEEE 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/AIEEE58915.2023.10134866.

[26] Y. Zhang, A. Herygers, T. Patel, Z. Yue, and O. Scharenborg, "Exploring Data Augmentation in Bias Mitigation Against Non-Native-Accented Speech," in *2023 IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ASRU57964.2023.10389756.

[27] J. Han, M. Matuszewski, O. Sikorski, H. Sung, and H. Cho, "Randmasking Augment: A Simple and Randomized Data Augmentation For Acoustic Scene Classification," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICASSP49357.2023.10095001.

[28] D. S. Park *et al.*, "Specaugment: A simple data augmentation method for automatic speech recognition," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, Graz, Austria: INTERSPEECH, Sep. 2019, pp. 2613–2617. doi: 10.21437/Interspeech.2019-2680.

[29] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data Augmentation and Deep Learning Methods in Sound Classification: A Systematic Review," *Electronics (Basel).*, vol. 11, no. 22, Nov. 2022, doi: 10.3390/electronics11223795.

[30] A. Y. Chang *et al.*, "GAP-AUG: GAMMA PATCH-WISE CORRECTION AUGMENTATION METHOD FOR RESPIRATORY SOUND CLASSIFICATION," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, Seoul: Institute of Electrical and Electronics Engineers Inc., Apr. 2024, pp. 551–555. doi: 10.1109/ICASSP48485.2024.10447967.

[31] T. T. Wong and P. Y. Yeh, "Reliable Accuracy Estimates from k-Fold Cross Validation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, Aug. 2020, doi: 10.1109/TKDE.2019.2912815.

[32] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction," *Clean. Eng. Technol.*, vol. 15, Aug. 2023, doi: 10.1016/j.clet.2023.100664.

[33] E. Bergman, L. Purucker, and F. Hutter, "Don't Waste Your Time: Early Stopping Cross-Validation," in *AutoML 2024 (International Conference on Automated Machine Learning)*, Paris, France: Proceedings of Machine Learning Research (PMLR), Volume 256, Sep. 2024, pp. 1–32. [Online]. Available: https://github.com/automl/DontWasteYourTime-early-stopping

[34] K. S. Gill and R. Gupta, "Chronic Kidney Disease Detection Using GridSearchCV Cross Validation Method," in *2023 International Conference on Recent Advances in Electrical, Electronics and Digital Healthcare Technologies, REEDCON 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 318–322. doi: 10.1109/REEDCON57544.2023.10151392.