ResSpec: Classification of Respiratory Sounds Using Machine Learning Algorithms

Kylle Margauxe Abubo Hufalar, Jasper Vanjo Nastor Toribio, Daryll Matthews Arabe Zambrano, Johnny Flores Verzola

Lorma Colleges – College of Computer Studies and Engineering.
CLI Bldg. San Juan Campus, San Juan La Union
Johnny.verzola@lorma.edu, daryllmatthews.zambrano@lorma.edu

Abstract

This study develops a respiratory sound classification system to assist respiratory therapists in identifying and classifying respiratory sounds using machine learning. ResSpec categorizes sounds like crackles and wheezes, which are indicative of conditions such as asthma and pneumonia. Using annotated datasets, it supports clinical decision-making and facilitates the early identification of respiratory issues.

Evaluation through the System Usability Scale (SUS) survey, involving licensed respiratory therapists, yielded an SUS score of 82.91, indicating "excellent" usability. The system proved effective for clinical support and educational purposes, particularly in detecting crackles and wheezes with reasonable accuracy.

While promising, limitations include the need for a larger dataset to improve classification of other respiratory sounds. Future efforts will focus on expanding the dataset, optimizing the model, and addressing current limitations to broaden its utility in healthcare.

Keywords**:** *Respiratory Sound Classification, Machine Learning, Crackles, Wheezes, Usability, Respiratory Therapy*

1. Introduction

Respiratory sounds, particularly lung sounds, are critical indicators of pulmonary health and are routinely assessed in clinical practice. These sounds are generated by the movement of air through the respiratory tract and can be categorized into normal and abnormal sounds. Normal lung sounds, such as vesicular breath sounds, are typically heard in healthy individuals, whereas abnormal lung sounds, such as wheezes, crackles, and rhonchi, can indicate various respiratory pathologies including asthma, chronic obstructive pulmonary disease (COPD), and pneumonia (Pasterkamp, et al., 1997).

Traditional methods of respiratory sound pattern detection often involve contact means such as stethoscopes, spirometers, and wearable systems (Chen et al., 2018). These methods can sometimes lead to discomfort for the patients and insufficient accuracy in respiration. Moreover, these methods are often unable to provide continuous monitoring, which is crucial for detecting subtle changes in respiratory patterns (Tobin et al., 1988). Recognizing the limitations of these traditional methods, advancements in technology offer promising solutions. Many new techniques of respiratory monitoring have been made available for clinical use recently (Tobin et al., 1988). These include pulse oximetry, transcutaneous carbon dioxide monitoring, and various wearable and remote technologies. However, their place is not always well defined, and appropriate use of available monitoring techniques and correct interpretation of the data provided can help improve our understanding of the disease processes involved and the effects of clinical interventions.

Respiratory sound patterns are an important indicator of an individual's health (Tobin et al., 1988). Altered respiratory patterns can be symptoms of many diseases. Moreover, breathing acts as an alarm for emotional stress, brain and body balance, cold and high body temperature or overheating. Also, during exercise or any physical exertion, our breathing rate acts as the bio-marker of physical exertion and fatigue.

In the local context, communities in the Philippines, especially in rural areas, face significant healthcare challenges. Access to healthcare professionals and advanced diagnostic tools is limited, leading to delays in diagnosis and treatment. Respiratory diseases often go undiagnosed or misdiagnosed, exacerbating health outcomes.

Chronic respiratory diseases, including Chronic Obstructive Pulmonary Disease (COPD) and asthma, pose significant health challenges in the Philippines. Urban locations, which are marked by elevated levels of air pollution and exposure to indoor smoke from biomass fuels, worsen these circumstances. Current data suggests an increase in respiratory diseases caused by variables such as the relaxation of COVID-19 measures and the spread of viruses such as influenza, mycoplasma pneumoniae, respiratory syncytial virus (RSV), and SARS-CoV-2. Acute respiratory tract infections were the predominant ailments in the country in 2020, afflicting almost 710.2 thousand individuals. In order to tackle this public health concern, it is imperative to enhance knowledge on respiratory well-being and advocate for preventative actions.

In rural areas, where healthcare professionals are limited, access to even basic health assessments can be a challenge. Respiratory diseases, such as asthma and COPD, are common in these populations and can go undiagnosed for long periods. ResSpec achieved to bridge this gap by providing an automated solution for early detection of respiratory conditions. By combining AI-driven diagnostics with real-time monitoring, the system provides a user-friendly platform for both patients and healthcare providers to track respiratory health continuously. This is especially crucial for individuals in rural areas, where healthcare access is limited, and the need for preventive care is high.

The World Health Organization (WHO), as laid out by the CFR.org Editors, is the UN organization responsible for leading international public health activities. Over its nearly seventy-five (75) years of existence, the WHO has achieved important milestones, such as the elimination of smallpox, while also experiencing criticism for perceived failings, such as its delayed response to the Ebola outbreak in 2014.

The researchers adopted a pluralistic approach for respiratory sound classification to pursue a well-rounded and effective approach. The researchers adopted a pluralistic approach for respiratory sound classification to secure a well-rounded and effective approach. There were, among other things, clinical insights, technical tools, and iterative testing. Clinical expertise was gathered through interviews with licensed respiratory therapists, providing a foundation of real-world knowledge about respiratory sounds and their clinical significance. Simultaneously, the incorporation of advanced technical tools such as digital stethoscopes for recording sounds and software such as Audacity for preprocessing data enhanced the effectiveness of data collection, cleaning, and evaluation. By merging these disparate ingredients together, it was consistently intended to transmute the gap between clinical practice and technological solutions, creating systems that are not only accurate but are also practical solutions for the healthcare professional. This pluralistic approach enabled building the bridge between all stakeholders from the medical and technical domains, thus ensuring conformance of the project with clinical reality and alignment with cutting-edge developments in machine learning and software development.

2. Objectives

ResSpec achieved to develop an application for detecting and classifying different respiratory sounds using machine learning, ensuring an acceptable accuracy rate. The primary objective was to provide users with a tool that accurately identified lung sound patterns, enabling healthcare professionals to diagnose patients with confidence. The application is user-friendly, allowing healthcare professionals to make timely analysis.

Specifically, the study achieved to:

1. Utilized a comprehensive dataset of respiratory sounds, including various lung sound patterns such as wheezes, crackles, and normal respiratory sounds.
2. Applied the compiled dataset to train machine learning algorithms for classifying recorded respiratory sounds. The focus was ensuring the robustness and reliability of the model's performance.
3. Developed a user-friendly software application that empowers healthcare professionals to identify respiratory sounds more confidently and efficiently, aiding in accurate diagnosis and treatment decisions.
4. Assessed the performance of the machine learning model, ensuring it meets an acceptable accuracy rate for clinical use.

3. Materials and methods

The study employed a mixed method of qualitative and developmental research. The descriptive part of the study design gave opportunity to collect, analyze, and interpret data on lung sound patterns, which include wheezes, crackles, and normal respiratory sounds. It allowed the researchers to provide detailed information on the characteristics and clinical importance of these sound patterns and would also capture several needs and challenges that respiratory therapists and other healthcare professionals face. Therefore, a combination of descriptive character with qualitative interviews and secondary data sources created a well-optimized ground for understanding the requirements of the system to be developed.

The developmental part of ResSpec is associated with design, implementation, and evaluation of the ResSpec system. This phase included data preprocessing, model training utilizing machine learning algorithms, and development of user-friendly application tailored for clinical use. Iterative testing and validation were carried out that met the standards of accuracy, usability, and clinical applicability. With that, use of two designs in this way also continues to give the study a structured framework for addressing the specific study objectives while maintaining a proper balance between theoretical exploration and practical application.

The selection of respondents was based on their qualifications in the field of respiratory care and through judgment sampling, considering their knowledge and engagement. The seven (7) Licensed Respiratory Therapists and ten (10) Focus Group Participants were chosen because the researchers believed their insights would be helpful for the study and would contribute to improving the system. Judgment sampling was utilized due to the limited time and the small pool of individuals with the necessary competencies for this selection.

Figure 1. Software Methodology

Figure 1 above presented the methodology developed using Cross-Industry Standard Process for Data Mining (CRISP-DM) as its main methodology, a widely used methodology in data mining and data science projects. It is also used and is effective for machine learning projects. Providing the researchers, a structured approach in the development of ResSpec, with clear stages and tasks to guide the process. The adoption of the Cross-Industry Standard Process for Data Mining (CRISP-DM) as the main methodology in the development of ResSpec signifies the significance of a structured approach in the project. With the 6 distinct phases CRISP-DM provides a robust framework for guiding the project and the researchers throughout the development process. CRISP-DM ensures that ResSpec is developed with a focus on accuracy, reliability, and practical utility. Through this systematic approach, ResSpec contributes meaningfully to the field of healthcare, offering a promising tool for aiding medical professionals for a more accurate and timely diagnosis.

4. Results

The researchers focused on understanding the distinctive features of these respiratory sounds, with wheezes being high-pitched and continuous sounds often associated with asthma or COPD, and crackles being short, explosive, and discontinuous sounds commonly heard in pneumonia. In addition to the secondary data, seven (7) licensed respiratory therapists at Lorma Medical Center in San Fernando, La Union, were interviewed to provide clinical insights into the recognition of respiratory sounds in practice. The discussions focused on key auditory features, such as pitch, intensity, timing within the respiratory cycle (inspiration vs. expiration), and sound location on the chest. Therapists also considered patient-specific indicators like accessory muscle use or observable movements. To further support this investigation, the therapists examined an open-source respiratory sound dataset to assess its applicability to clinical scenarios. Their feedback ensured the dataset was consistent with real-world observations and useful for practical applications.

To further enhance the dataset's robustness, the researchers used a Stemoscope digital stethoscope to manually record respiratory sounds from patients. Informed consent was obtained from the patients before recording their respiratory sounds. This manual data collection aimed to support the validation and real-world evaluation of the system. To provide a comprehensive understanding of the dataset utilized in the study, an example of some of the audio files included in the dataset is presented. This figure illustrates some of the key samples in the dataset, which were used to train and evaluate the Convolutional Neural Network (CNN) model for respiratory sound classification. Each audio in the database has a corresponding annotation, a snippet of it which is shown in Figure 2.

start_time	end_time	crackles	wheezes
0.036	0.579	0	0
0.579	2.45	0	0
2.45	3.893	1	1
3.893	5.793	0	0
5.793	7.521	1	0
7.521	9.279	1	0
9.279	11.15	0	0
11.15	13.036	0	1
13.036	14.721	1	0
14.721	16.707	0	0
16.707	18.507	1	0
18.507	19.964	1	0

Figure 2. Audio Annotation

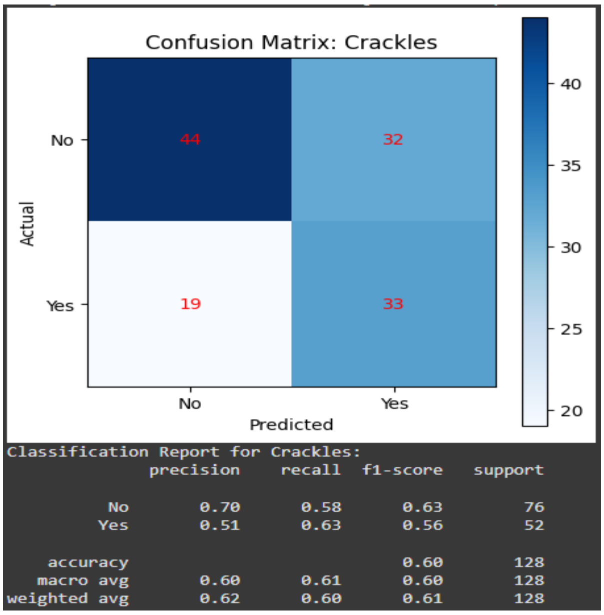
Specific and distinct steps in the process of training machine learning model for audio classification to analyze respiratory sounds were executed. The initial stage of data collection was called the pre-processing step at which time researchers collected and cleaned a dataset with respiratory sound recordings. It included denoising and artifact removal, audio segmentation into chunks that are meaningful for diagnosis followed by corresponding labeling like wheezes, crackles or normal. This is part of preprocessing which was used to standardize the data so that it could feed into the model.

The researchers then performed feature extraction by isolating the most important audio features that correlate to respiratory sound classification. These features were based on both time-domain (amplitude, pitch) and frequency domain spectral components that identify characteristic lung sound patterns. Subsequently these were converted to an image format that could be used for machine learning. Mel spectrograms provided detailed and separated parts of the audio signals

After extracting the features, the researchers entered into model selection and training phase. Using a Convolutional Neural Network (CNN) made the most sense as it had already achieved success with image and audio data analysis. The CNN model was trained on the spectrogram representations of respiratory sounds to determine what patterns within sound features delineate different acoustical categories. Supervised learning in the past, the researchers have had problems where the model was trained on labeled audio data.

Hyperparameter tuning was performed during the training of the model to get it tuned-up leading up to just the right performance. Hyperparameters such as learning rate, batch size and the layer number of convolution layers were tuned iteratively to find a good balance among accuracy computational efficiency. The training process also used metrics such as accuracy, precision, recall and F1-score to understand the model with each iteration.

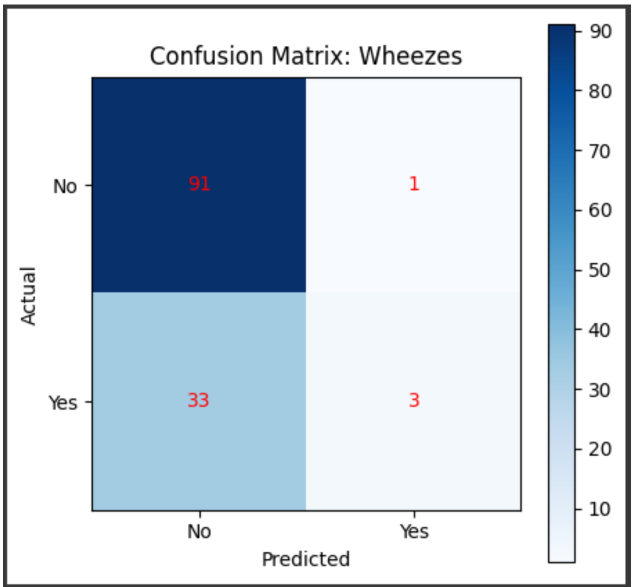
The confusion matrix in detail, which gives a complete perspective on the performance of a model when classifying respiratory sounds into different categories such as wheezes, crackles, or even normal sounds. It shows samples of correct classifications (the true positives) and misclassifications (false positives and false negatives), thus giving insights into the strengths and weaknesses of the model. Finally, the confusion matrix gives an understanding of the accuracy, precision, recall, and overall reliability of the model's predictions. From this thorough analysis of this matrix which pinpoints some areas where the model has performed quite well, as well as conditions or other patterns under which it's been known to make errors, tailor-made improvements may then be made, such as strengthening the dataset, improving feature extraction, or fine-tuning machine learning algorithms, to boost the robustness and dependability of systems towards clinical applications.



	precision	recall	f1-score	support
No	0.70	0.58	0.63	76
Yes	0.51	0.63	0.56	52
accuracy			0.60	128
macro avg	0.60	0.61	0.60	128
weighted avg	0.62	0.60	0.61	128

Figure 3. Confusion Matrix for Crackles

The confusion matrix and classification report for crackle detection in figure 8 reveal moderate model performance, with an overall accuracy of 60%. The model correctly identifies the absence of crackles (class “No”) 44 times but misclassifies it as present (class “Yes”) 32 times. For actual crackles, it correctly predicts 33 instances while incorrectly classifying 19. The precision for detecting crackles is 51%, indicating that nearly half of the positive predictions are false positives. The recall is 63%, suggesting the model captures a reasonable proportion of true crackle instances. The F1-score of 0.56 reflects a balance between precision and recall, indicating room for improvement. A macro-average F1-score of 0.60 highlights balanced performance across classes, though the lower recall for class “No” (58%) contributes to the modest weighted average F1-score of 0.61. These findings suggest the model is conservative in predicting “Yes” but requires further optimization to enhance crackle detection accuracy.



Classification Report for Wheezes:

	precision	recall	f1-score	support
No	0.73	0.99	0.84	92
Yes	0.75	0.08	0.15	36
accuracy			0.73	128
macro avg	0.74	0.54	0.50	128
weighted avg	0.74	0.73	0.65	128

Figure 4. Confusion Matrix for Wheezes

The classification report for wheeze detection in figure 10 shows an overall accuracy of 73%, reflecting moderate performance. While the model achieves high precision for wheeze detection (75%), its recall is only 8%, indicating a significant failure in identifying most actual wheeze cases. Out of 36 true wheeze instances, only 3 are correctly predicted, with 33 misclassified as “No.” In contrast, for non-wheeze cases, the model achieves a near-perfect recall of 99%, correctly predicting 91 out of 92 instances. The F1-score for wheeze detection is 0.15, highlighting a poor balance between precision and recall, whereas the “No” class achieves a much higher F1-score of 0.84. The macro-average F1-score of 0.50 underscores the imbalance in performance across classes, and the weighted average F1-score of 0.65 reflects the model’s bias toward “No” predictions. These results indicate a need for further optimization, particularly to improve recall for wheeze detection, possibly through rebalancing the dataset or adjusting decision thresholds.

The training performance graph illustrates how the model's accuracy improved over time during training. It plots the training accuracy versus the number of epochs or training cycles, showing how the model's ability to classify respiratory sounds becomes more refined as it learns from the data.

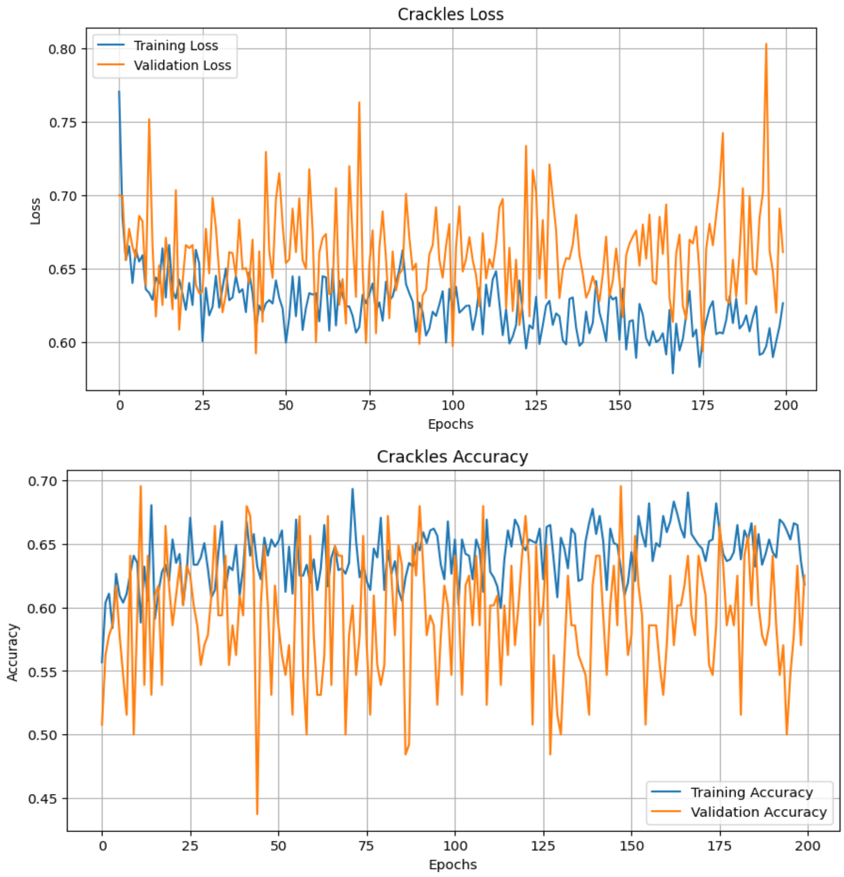


Figure 5. Training Loss and Validation for Crackles

The plot in Figure 5 demonstrates the training and validation performance of a neural network model over 200 epochs. It showcases both training and validation loss, as well as training and validation accuracy. The steady decrease in training loss indicates that the model is successfully learning from the data. However, after 150 epochs, the validation loss begins to rise, signaling overfitting as the model becomes too specialized to the training data. This is further supported by the trend in accuracy, where training accuracy continues to improve while validation accuracy starts to decline after the same point. Ideally, training and validation loss and accuracy should converge, indicating that the model generalizes well to unseen data. The divergence observed in this plot suggests the need for measures to prevent overfitting.

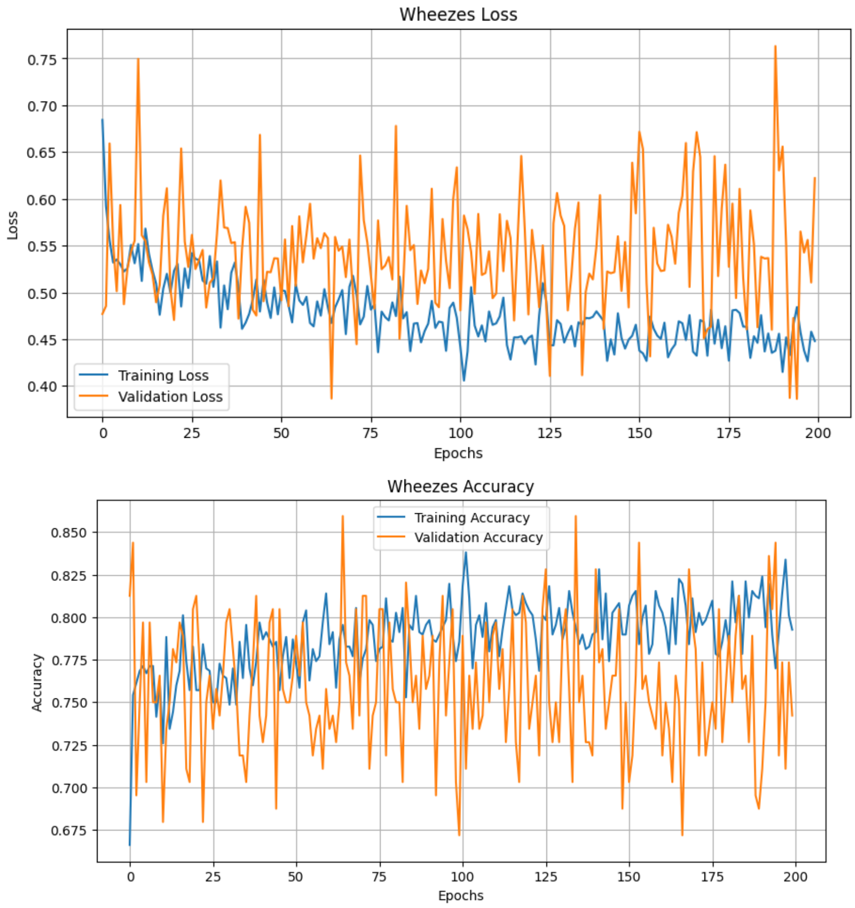


Figure 6. Training Loss and Validation for Wheezes

Figure 6 provides insight into the model's performance and learning behavior over 200 epochs. As training progresses, the training loss steadily decreases, indicating that the model is increasingly able to fit the training data. However, the validation loss remains elevated and fluctuates, particularly in the later epochs, further highlighting the issue of overfitting. This pattern is reinforced by the accuracy trends, where early improvements in both training and validation accuracy give way to a plateau and eventual decline in validation accuracy after approximately 100 epochs. The divergence between training and validation performance underscores the need for strategies to improve generalization.

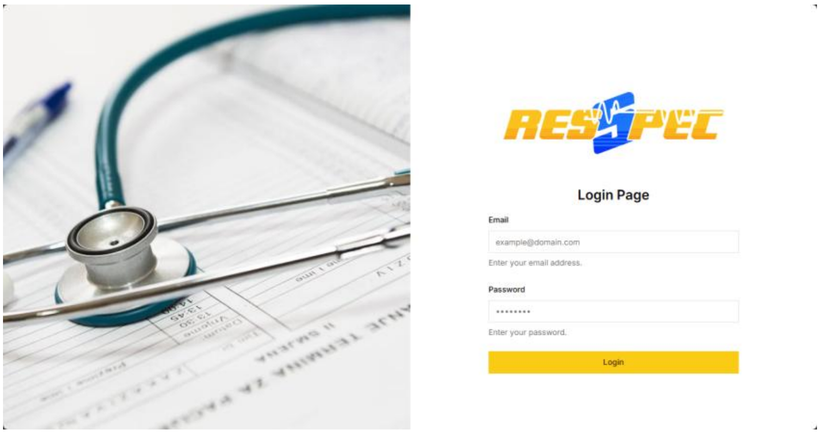


Figure 7. Login Page

On Figure 7, Upon launching the app, users are greeted by the login page, which serves as the gateway to access the system. User accounts can only be created and assigned by the super admin.

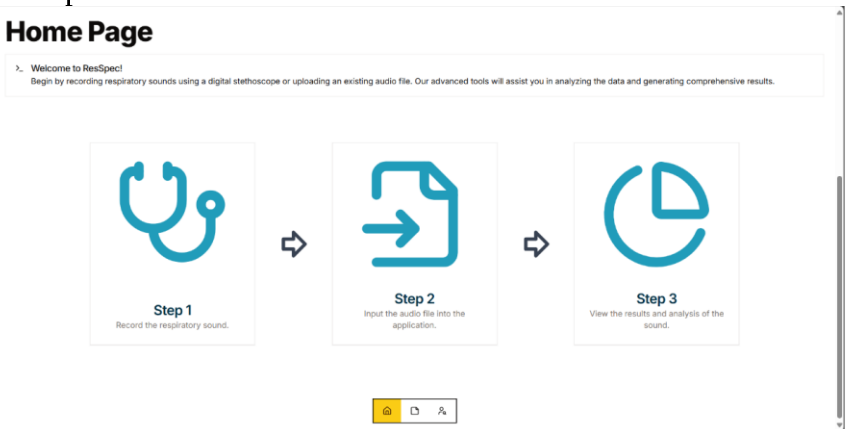


Figure 8. Tutorial Page

Figure 8 highlights the Tutorial Page is one of the key elements within the ResSpec application to assist users on how it can be used. Related Guide With a welcoming introductory page presented to the user with basic instructions here is further elaborated in form of guide outlining step by step process through which the user can avail the utmost working of this application. The tutorial is designed for both newbies and experts, so anyone with a basic understanding of computer technology can follow along.

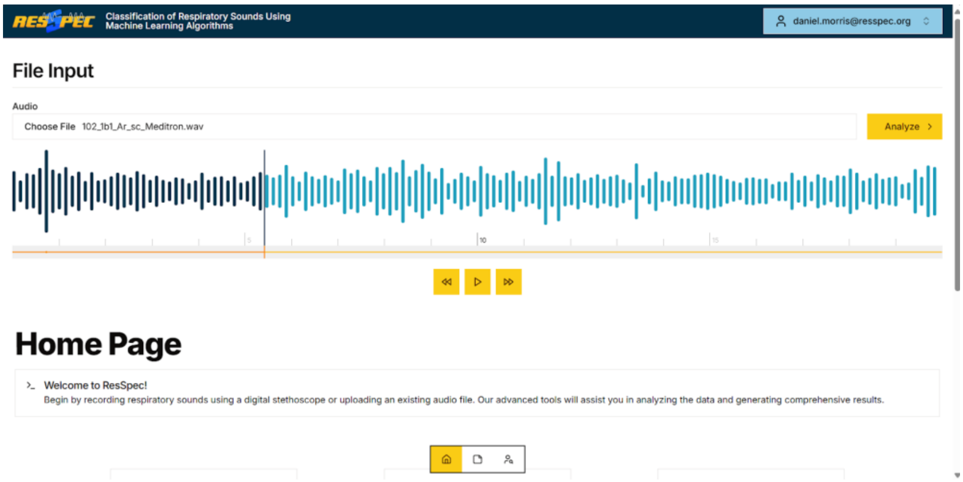


Figure 9. Audio Input

Figure 9 presents the audio input feature that forms the core part of the functionality of the application. It enables uploading sound recordings of respiration for analysis and classification. This can be achieved by providing users with simple options on the selection and uploading of audio files in the format that is compatible through an intuitive interface. After the audio file is uploaded, the application processes it to extract relevant features that are to be analyzed for the classification of the respiratory sound. It is capable of handling different types of respiratory sounds like wheezes, crackles, and normal respiratory sounds, therefore ensuring that categorization is done correctly and reliably.

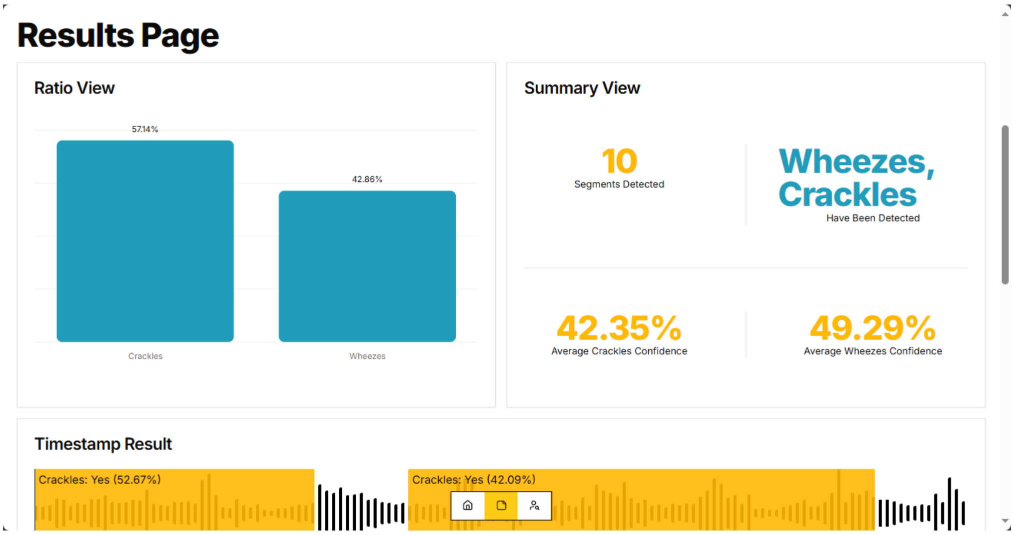


Figure 10. Results Page

Figure 10 presents the results page that acts as a consolidated presentation of the results obtained from the feature extraction of the uploaded respiratory sound. This page gives the user an intuitive and detailed summary of the processed data, thereby presenting the results in a comprehensible and user-friendly format.

Figure 11. Timestamp Result Page

Figure 11 presents the timestamp result page such as a diagram for the analysis performed on uploaded respiratory sound. It thus indicates at which specific timestamps the anomalies are detected. Besides, it has details of the kind of respiratory sounds that are above the results classified as crackles, wheezes, normal respiratory sounds, and at what point in the audio file are these patterns detected. Thus, present a thorough fine-grained timestamp result page for respiratory professionals to identify better irregularities for precise and timely diagnosis.

To determine the level of usability of the developed functional web-based system ResSpec, respiratory sound classification system, the researchers evaluated the system using the System Usability Scale (SUS) and conducted demonstrations with licensed respiratory therapists. This process involved hands-on interaction with the system, followed by interviews to gather qualitative insights and perspectives on the system's functionality and clinical relevance (see appendix D). The results specifically address the fifth objective of the study: assessing the usability and performance of the developed system.

The SUS results, presented in the table, reflect three (3) key dimensions of usability: relevance, usability and system performance, and overall satisfaction.

Based on the results of the SUS evaluation shown at figure 28, it demonstrated that the developed system is a user-friendly and clinically relevant tool for respiratory sound classification. With its current functionality, the system meets its intended purpose and holds promise for broader adoption in clinical and educational settings. However, future improvements, such as increasing the dataset and addressing identified limitations, would further enhance its utility and reliability.

Raw Score	X	Y	X0	Y0	SUS Score	Grade	Adjective	Acceptable
43	28	15	23	10	82.5	A	Excellent	Acceptable
45	30	15	25	10	87.5	A	Excellent	Acceptable
50	31	19	26	6	80	B	Good	Acceptable
46	31	15	26	10	90	A	Excellent	Acceptable
45	29	16	24	9	82.5	A	Excellent	Acceptable
44	27	17	22	8	75	B	Good	Acceptable
					82.91666667	A	Excellent	Acceptable

5. Conclusion

The researchers were able to successfully gathered detailed information from licensed respiratory therapists, which is valuable information to the research in relation to the understanding and differentiation of respiratory sounds such as the wheezes, crackles and normal sounds. Such a process made sure that the model development had its clinical relevance in practice as sound identification and classification criteria was shared by the respiratory therapists, hence making a robust foundation for the study.

For the second objective, the researchers had in their possession a comprehensive dataset of respiratory sounds that consisted of not only cumulative sounds of a single lung such as wheezes and crackles but also normal respiratory sounds. In addition to this, the said dataset was also accumulated through the combination of manually recording sounds using a digital stethoscope and obtaining information from a credible online repository. The data was subjected to preprocessing and validation to improve the quality of the dataset to the intended statement of objectives.

For the third objective, the researchers used the collected dataset to train machine learning algorithms with an emphasis on correctly identifying recorded respiratory sounds. The model's robustness and dependability were emphasized, and sophisticated techniques were employed to improve its classification capabilities and guarantee consistent outcomes over a range of data inputs.

For the fourth objective, the researchers developed a user-friendly application whose objective is to enable healthcare personnel to diagnose respiratory sounds accurately, confidently and in a less time-consuming manner. This application combines the trained model of machine learning and other functions such as sound analysis, visualization of results and their export that make considerable sense in assisting sound diagnosis and treatment. The user-friendly design combined with the full range of features makes it possible to effortlessly introduce the application into the clinical processes.

For the fifth objective, the usability of the developed system was evaluated using the System Usability Scale (SUS). Licensed respiratory therapists were engaged to test the system through live demonstrations and hands-on usage. The SUS results indicated that the system achieved an overall score of 82.91, categorized as "excellent" and "acceptable" for clinical use. This confirmed the system’s relevance, usability, and overall satisfaction among end-users. However, it was noted that the system's primary utility lies in educational and non-emergency scenarios, given the constraints of time-sensitive settings. Further suggestions was made as to the actual enhancement of the further availability and function of the system for possible use during emergency cases. Feedback from various users also emphasized the continuous improvement and upgrade according to the experiences of users to enable sustained reliability during clinical practice.

6. References

Géron, A. (2020). Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (2nd ed.). O'Reilly Media.

Tobin, M. J., Manthous, C. A., & Schmidt, G. A. (1998). Principles and Practice of Intensive Care Monitoring. McGraw-Hill.

Shivakumar, V. (2019). Classification of respiratory sounds. Stanford University CS230 Course Project Report. https://cs230.stanford.edu/projects_fall_2019/reports/26260886.pdf

Arar, M.E., & Sedef, H. (2023). An efficient lung sound classification technique based on MFCC and HDMR. SIViP, 17, 4385–4394. https://doi.org/10.1007/s11760-023-02672-2

Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. Journal of Electrical Systems and Information Technology, 10(44). https://www.researchsquare.com/article/rs-1885746/v2

Brochard, L., Martin, G. S., Blanch, L., Pelosi, P., Belda, F. J., Jubran, A., … & Vincent, J. L. (2012). Clinical review: Respiratory monitoring in the ICU - a consensus of 16. Critical Care, 16(219). https://ccforum.biomedcentral.com/articles/10.1186/cc11146

Chen, J., Hu, B., Moore, P., Zhang, X., & Ma, X. (2018). Non-contact respiration measurement using an ultrasonic proximity sensor for user identification. Sensors (Basel, Switzerland), 19(1), 85. https://doi.org/10.3390/s19010085

9. Author(s) Biodata (50 words)

KYLLE MARGAUXE A. HUFALAR is a Bachelor of Science in Information Technology student at Lorma Colleges. He is a consistent academic scholar with interests in artificial intelligence and digital innovation. Outside of academics, he enjoys creating music and playing video games, balancing creativity with technical pursuits. He can be contacted at email: kyllemargauxe.hufalar@lorma.edu.

	**JASPER VANJO N. TORIBIO** is pursuing a Bachelor of Science in Information Technology at Lorma Colleges. A consistent Dean's Lister, he is passionate about coding and exploring complex systems. He enjoys discovering new programming techniques and playing video games, merging curiosity with hands-on problem-solving. He can be contacted at email: jaspervanjo.toribio@lorma.edu.
	**DARYLL MATTHEWS A. ZAMBRANO** is a full scholar and senior Information Technology student at Lorma Colleges. With a strong foundation in digital literacy and leadership, he also holds PhilNITS certification. He enjoys editing in Photoshop and playing video games, blending artistic skill with technical expertise. He can be contacted at email: daryllmatthews.zambrano@lorma.edu.