

Obstructive Sleep Apnea (OSA) Severity Classification Using Tongue Ultrasound Images and YOLOv8 [†]

Rosezellynda D. Regular  and Cyrel O. Manlises * 

School of Electrical, Electronics, and Computer Engineering, Mapua University, Manila 1002, Philippines; rdregular@mymail.mapua.edu.ph

* Correspondence: ccontinmare@mapua.edu.ph

[†] Presented at the 7th Eurasia Conference on IoT, Communication and Engineering 2025 (ECICE 2025), Yunlin, Taiwan, 14–16 November 2025.

Abstract

Obstructive sleep apnea (OSA) is a widely known sleep disorder that leads to serious health problems and complications. The standard diagnosis method of OSA is polysomnography. However, the process is time-intensive, expensive, and not readily accessible. Machine learning (ML) has been increasingly applied in various medical imaging modalities; however, there is still a lack of research on applying ML to ultrasound imaging for OSA classification. Previous studies on ML applications in medical imaging adopt X-rays, Computed Tomography, and Magnetic Resonance Imaging, leaving ultrasound as an underexplored area. Using the You-Only-Look-Once version 8 algorithm and static tongue ultrasound images, we classified OSA severity: normal, mild, moderate, and severe. A total of 280 ultrasound images were augmented to 838 images using brightness scaling, which enhanced the training process of the model. The system was tested on 60 images, achieving an overall classification accuracy of 85%. The results demonstrate the possibility and potential of using machine learning and ultrasound imaging for classifying the severity of OSA, suggesting potential assistance to clinicians in diagnosing and intervening in this condition.

Keywords: YOLOv8; obstructive sleep apnea; ultrasound; image processing; tongue

1. Introduction

Difficulty in falling or staying asleep leads to poor-quality sleep, which increases the risk of serious illness and death [1]. A common sleep disorder that is characterized by airway obstruction resulting from throat muscle relaxation during sleep is known as obstructive sleep apnea (OSA) [2–4]. Many cases go undiagnosed, and if left untreated, OSA can lead to health complications [5]. Polysomnography (PSG) is a known standard for the diagnosis of OSA [2], as it records and evaluates the physiological effects of breathing episodes. However, the disadvantage of PSG is that the process is slow [6] and costly [7] as it requires patients to be admitted to a hospital that has a sleep laboratory, where they are under the supervision of a specialized technician. To address these limitations, wearable technologies have been developed to enable portability and convenience in classifying OSA [8–10]. Ultrasound imaging has also emerged, offering a non-invasive approach to visualize the upper airway and assess anatomical factors that may contribute to OSA [11,12].

Manlises et al. used AI to classify the severity of OSA based on ultrasound imaging, utilizing a gated recurrent unit model in determining the severity of OSA, with an overall



Academic Editors: Teen-Hang Meen, Chi-Ting Ho and Cheng-Fu Yang

Published: 23 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

performance of 43.49% [13,14]. They used a modified optical flow method (MOFM) to track tongue movements captured in ultrasound recordings. It is used to examine these movements during normal breathing and while performing the Müller maneuver. The extracted data from the MOFM were used as a vital parameter for model training. However, the system's performance was impacted by dataset imbalances due to the limited data, and augmentation techniques did not apply to this type of dataset [13]. The lack of standard compression force when placing the transducer on the submandibular area causes variations between patients, as pressure is adjusted to maintain clear ultrasound images. This inconsistency affects both ultrasound images and system performance [14].

Despite the increasing use of machine learning (ML) in medical imaging, research on applying ML to ultrasound images for diagnosing and classifying OSA remains limited. Imaging methods such as X-rays, Computed Tomography, and Magnetic Resonance Imaging have been extensively studied using ML, whereas ultrasound remains underexplored. You Only Look Once (YOLO) is a known object detection model and has demonstrated strong performance and adaptability across various fields. Researchers have successfully applied object detection in agriculture and healthcare [15–18]. Moreover, different YOLO versions, including YOLOv3 and YOLOv4, have been used for fish species detection, disease classification in goldfish, and fruit identification [19–21]. These applications present the versatility and continued relevance of YOLO. Hence, we classified OSA severity using tongue ultrasound images through machine learning techniques, without employing optical flow to analyze dynamic movements.

Specifically, we developed a prototype system using a Raspberry Pi 4 that captures ultrasound images of the tongue using a camera, and classified the severity of OSA based on tongue ultrasound images with the YOLOv8 algorithm. We adopted a confusion matrix to assess how well the system works. The system offers a significant advantage in the medical field, particularly in diagnosing OSA. Automating this process aids physicians and ultrasound technologists in the early detection and intervention for OSA. The results of this study serve as a foundation for future research.

2. Methodology

2.1. Conceptual Model

Figure 1 shows the conceptual framework of this study, outlining the process for classifying OSA severity using tongue ultrasound images. Image collection is performed using a camera integrated into the prototype device. The captured images were pre-processed to enhance quality. The YOLOv8 model then classifies the images into normal, mild, moderate, and severe levels. The predicted severity result is finally displayed through the system's graphical user interface.

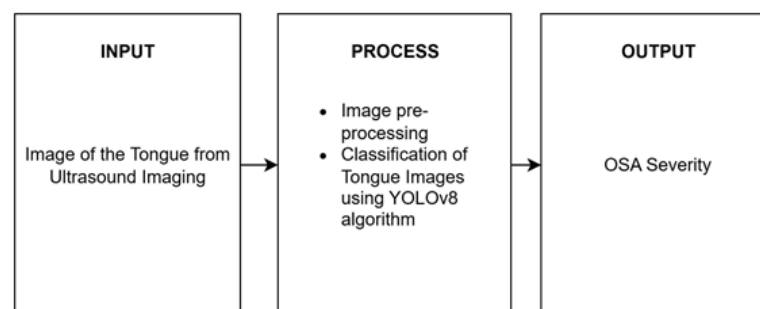


Figure 1. Conceptual model of this study.

2.2. Hardware Development

Figure 2 illustrates the design and implementation of the hardware prototype and its components. A 1080p USB camera (A4Tech, Taipei, Taiwan)) is used to capture the tongue ultrasound images. It features a 3.5-inch LCD touchscreen with a resolution of 320×480 , serving as the primary display interface. This screen is positioned adjacent to the camera holder within the casing, ensuring optimal visibility for user interaction. Additionally, the Raspberry Pi (Raspberry Pi Ltd., Cambridge, UK) is housed in the same housing as the LCD.



Figure 2. Prototype design of the system developed in this study.

2.3. System Workflow

The system workflow is shown in Figure 3. The system starts by receiving images captured by the camera as input. These images are pre-processed to enhance quality and optimize recognition accuracy for the YOLOv8 algorithm. The pre-processing steps include resizing, noise reduction, and normalization, as well as standardizing the images and improving their suitability for the training dataset. Once pre-processed, the images are analyzed to classify the severity of OSA. Finally, these results are displayed on an LCD screen for user interpretation.

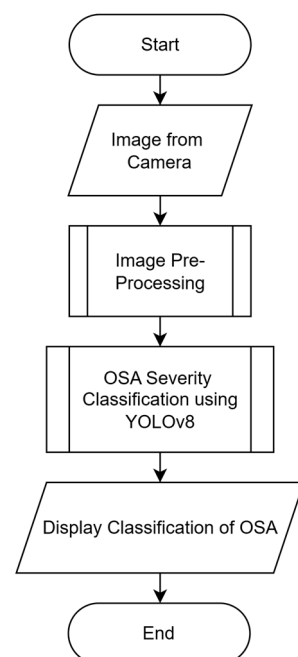


Figure 3. System workflow of this study.

2.4. Data Gathering

The dataset used in this study consisted of 280 submental ultrasound images from patients diagnosed with OSA. Submental ultrasound was selected because it captures internal tongue structures that regular cameras cannot. As the base dataset, a total of 280 static ultrasound images were selected for analysis. The dataset includes 60 images per class (normal, mild, moderate, and severe). These were extracted from ultrasound recordings using Adobe Premiere Pro version 25.2 and supplemented with images taken from the monitor using a USB camera. All images and severity classifications were verified by physicians and technicians at Cardinal Tien Hospital.

The tongue region was annotated using Roboflow, where bounding boxes were applied to create training labels. Brightness scaling from -30 to $+30\%$ was used for data augmentation, expanding the dataset to 838 images (Figure 4). These were grouped into 754 for training, 84 for validation, and 60 (15 per class) for testing. The YOLOv8 model was trained to detect the tongue and classify OSA severity, generating bounding boxes, class labels, and confidence scores on new images.

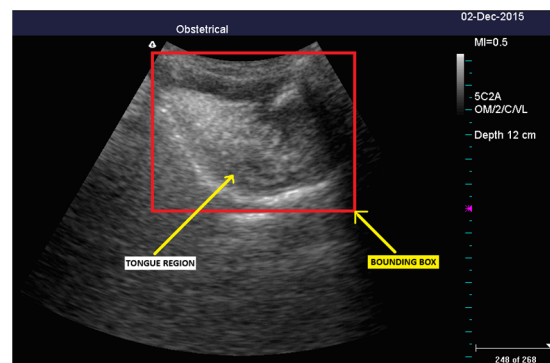


Figure 4. Sample dataset labelled using Roboflow.

3. Results and Discussion

A sample output from the system is shown in Figure 5. The test results showed a prediction of 91.2% for the severe classification. This indicates that the model of the system can accurately classify the severity of OSA.

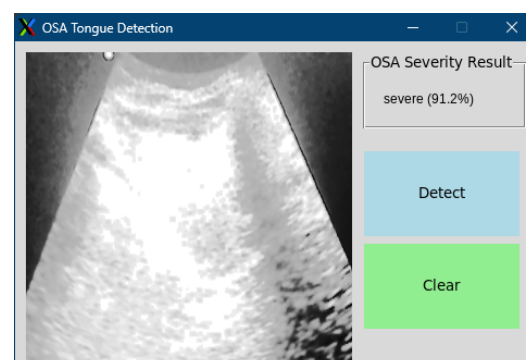


Figure 5. Sample test result (severe).

The confusion matrix of the system's prediction across the four severity levels is presented in Table 1. Diagonal entries represent the correct classification of obstructive sleep apnea severity. Out of 15 test images per class, the system achieved perfect classification for the normal severity level. In contrast, it correctly classified 10 images as mild, 12 as

moderate, and 14 as severe. The accuracy was calculated using Equation (1), resulting in an overall accuracy of 85%.

$$Accuracy = \frac{\sum_{n=1}^4 A_{nn}}{\sum_{i=1}^4 \sum_{j=1}^4 A_{ij}} \quad (1)$$

Table 1. Confusion matrix of OSA severity classification.

		Observation			
		Normal	Mild	Moderate	Severe
Predicted results	Normal	15	0	0	0
	Mild	3	10	2	0
	Moderate	3	0	12	0
	Severe	1	0	0	14

These findings indicate that the system performs best when distinguishing between extreme cases, normal and severe, where tongue ultrasound features are more distinct and identifiable. However, during testing, mild and moderate showed slightly lower accuracy, with mild being the most prone to misclassification.

The results showed that the system developed performed best when the visual difference between severities is more distinct. The system achieved an overall accuracy of 85%, but encountered some limitations. The small dataset size might limit the ability of the model to generalize the dataset. To address this, brightness scaling, an augmentation technique, was applied to increase data variety and enhance the model's training process. However, augmentation cannot substitute for a larger and more diverse set of real patient data. Since the model was trained and evaluated using static images rather than dynamic sequences, some anatomical cues available during real-time tongue movement may have been lost.

4. Conclusions and Recommendations

The study aims to classify the severity of OSA using static tongue ultrasound images through a machine learning approach. A system was created using a Raspberry Pi 4B, which included an integrated USB camera and LCD. An overall accuracy of 85.0% demonstrates satisfactory performance in classifying normal and severe OSA cases. However, mild and moderate levels were more challenging due to their similar visual features. It is necessary to expand the dataset to enhance model reliability and explore additional parameters, such as tongue area, thickness, or other non-invasive signals, to improve accuracy. Furthermore, testing alternative models or refining training techniques may help address misclassification between mild and moderate cases.

Author Contributions: Topic development, Research Conceptualization, Paper editing, Polishing, and Finalization, C.O.M.; Paper writing, Software development (programming), Hardware development (prototype creation), R.D.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The Institutional Review Board of the Cardinal Tien Hospital has approved this prospective observational study (CTH-105-3-5-057).

Informed Consent Statement: Written informed consent was obtained from each participant.

Data Availability Statement: The original data presented in the study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.19598218>.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Crowley, K. Sleep and Sleep Disorders in Older Adults. *Neuropsychol. Rev.* **2011**, *21*, 41–53. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Chen, J.W.; Lin, S.-T.; Wang, C.-Y.; Lin, C.-C.; Hsu, K.-C.; Yeh, C.-Y.; Hwang, S.-H. A Signal Segmentation-Free Model for Electrocardiogram-Based Obstructive Sleep Apnea Severity Classification. *Adv. Intell. Syst.* **2023**, *5*, 2200275. [\[CrossRef\]](#)
3. Jugé, L.; Olsza, I.; Knapman, F.L.; Burke, P.G.R.; Brown, E.C.; Stumbles, E.; de Frescheville, A.F.B.; Gandeia, S.C.; Eckert, D.J.; Butler, J.E.; et al. Effect of upper airway fat on tongue dilation during inspiration in awake people with obstructive sleep apnea. *Sleep* **2021**, *44*, zsab192. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Gulotta, G.; Iannella, G.; Vicini, C.; Polimeni, A.; Greco, A.; De Vincentiis, M.; Visconti, I.C.; Meccariello, G.; Cammaroto, G.; De Vito, A.; et al. Risk Factors for Obstructive Sleep Apnea Syndrome in Children: State of the Art. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3235. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Yang, W.; Fan, J.; Wang, X.; Liao, Q. Sleep Apnea and Hypopnea Events Detection Based on Airflow Signals Using LSTM Network. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 5366–5369. [\[CrossRef\]](#)
6. Mendonça, F.; Mostafa, S.S.; Ravelo-García, A.G.; Morgado-Dias, F.; Penzel, T. A Review of Obstructive Sleep Apnea Detection Approaches. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 825–837. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Kalkanis, A.; Testelmans, D.; Papadopoulos, D.; Van den Driessche, A.; Buyse, B. Insights into the Use of Point-of-Care Ultrasound for Diagnosing Obstructive Sleep Apnea. *Diagnostics* **2023**, *13*, 2262. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Loceo, A.N.G.; Brandan Lim, R.; Pellegrino, R.V. Monitoring of Breathing Effort and Oxygen Levels for Identification of Sleep Apnea. In Proceedings of the 2023 15th International Conference on Computer and Automation Engineering (ICCAE), Sydney, Australia, 3–5 March 2023; pp. 361–366. [\[CrossRef\]](#)
9. Wang, C.H.; Lu, S.J.; Chou, Y.M.; Kao, C.Y.; Lin, H.Y. A Wearable Solution for Obstructive Sleep Apnea Risk Evaluation Based on Optical Sensor. In Proceedings of the IEEE Sensors, Vienna, Austria, 29 October–1 November 2023. [\[CrossRef\]](#)
10. Sumona, S.A.; Aurthy, W.B.N. A Novel Low-Cost Monitoring System for Sleep Apnea Patients. In Proceedings of the 3rd International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 23–25 February 2023. [\[CrossRef\]](#)
11. Davidson, T.M. The Great Leap Forward: The anatomic basis for the acquisition of speech and obstructive sleep apnea. *Sleep Med.* **2003**, *4*, 185–194. [\[CrossRef\]](#) [\[PubMed\]](#)
12. You-Ten, K.E.; Siddiqui, N.; Teoh, W.H.; Kristensen, M.S. Point-of-care ultrasound (POCUS) of the upper airway. *Can. J. Anaesth.* **2018**, *65*, 473–484. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Manlises, C.O.; Chen, J.W.; Huang, C.C. A gated recurrent unit model based on ultrasound images of dynamic tongue movement for determining the severity of obstructive sleep apnea. *Ultrasonics* **2024**, *141*, 107320. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Manlises, C.O.; Chen, J.W.; Huang, C.C. Dynamic tongue area measurements in ultrasound images for adults with obstructive sleep apnea. *J. Sleep Res.* **2020**, *29*, e13032. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Legaspi, J.; Pangilinan, J.R.; Linsangan, N. Tomato Ripeness and Size Classification Using Image Processing. In Proceedings of the 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 8–9 December 2022; pp. 613–618. [\[CrossRef\]](#)
16. Martinez, M.R.B.; Dayrit, K.M.D.; Yumang, A.N. Classification of Red Watermelon Varieties Using Canny Edge Detection and CNN. In Proceedings of the 2024 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 29 June 2024; pp. 47–52. [\[CrossRef\]](#)
17. Bascara, F.C.A.; Yumang, A.N. Defect Detection and Classification of Soybean Using Convolutional Neural Network. In Proceedings of the 2024 7th International Conference on Information and Computer Technologies (ICICT), Honolulu, HI, USA, 15–17 March 2024; pp. 265–270. [\[CrossRef\]](#)
18. Macawile, M.J.; Quiñones, V.V.; Ballado, A.; Dela Cruz, J.; Caya, M.V. White blood cell classification and counting using convolutional neural network. In Proceedings of the 2018 3rd International Conference on Control and Robotics Engineering (ICCRE), Nagoya, Japan, 20–23 April 2018; pp. 259–263. [\[CrossRef\]](#)
19. Chan, C.J.L.; Reyes, E.J.A.; Linsangan, N.B.; Juanatas, R.A. Real-time Detection of Aquarium Fish Species Using YOLOv4-tiny on Raspberry Pi 4. In Proceedings of the 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAET), Kota Kinabalu, Malaysia, 13–15 September 2022; pp. 1–6. [\[CrossRef\]](#)

20. Migallon, M.F.; Santos, T.A.M.; Villaverde, J.F. Fruit Identification and Estimation of Calories Using YOLOv3 Algorithm. In Proceedings of the 2024 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Shah Alam, Malaysia, 29 June 2024; pp. 112–116. [[CrossRef](#)]
21. Medina, J.K.; Tribiana, P.J.P.; Villaverde, J.F. Disease Classification of Oranda Goldfish Using YOLO Object Detection Algorithm. In Proceedings of the 2023 15th International Conference on Computer and Automation Engineering (ICCAE), Sydney, Australia, 3–5 March 2023; pp. 249–254. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.