Comparison and Calibration of Axillary and Esophageal Temperature Measurements in Critically Ill ICU Patients: A Prospective Observational Study.

Satyajit Choudhury[1], Samir Samal [2], Pritam Chhotray[3], Abhilash Dash[4*], E. Shantanu Kumar Patra[5], Sagarika Panda[6], Shakti Bedanta Mishra[7]

[1]Assistant Professor, Department of Critical Care Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India.

[2]Professor, Department of Critical Care Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India.

[3]Associate Professor, Department of Pulmonary Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India.

[4]Associate Professor, Department of Critical Care Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India.

[5]Senior Resident, Department of Critical Care Medicine, IMS & SUM Hospital, Campus 3, Sitalapalli, Berhampur, Odisha, India.

[6]Associate Professor, Department of Critical Care Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India.

[7] Professor, Department of Critical Care Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India.

***Corresponding author:** Dr. Abhilash Dash
Associate Professor, Department of Critical Care Medicine, IMS & SUM Hospital, Bhubaneswar, Odisha, India, Email: abhilashdash@soa.ac.in

ABSTRACT

BACKGROUND: *Accurate temperature monitoring is critical in Intensive Care Unit (ICU) settings given its influence on treatment decisions. While esophageal temperature measurement is the gold standard, its invasive nature limits application. This study investigates the correlation and agreement between axillary (non-invasive) and esophageal temperature measurements in critically ill patients.*

METHODS: *A prospective observational study was conducted over six months, including adult ICU patients requiring continuous temperature monitoring. Both temperature measurements were taken simultaneously, yielding 1,109 pairs for analysis using Bland-Altman, correlation coefficients, and regression modeling.*

RESULTS: *The mean axillary temperature was 99.03°F, while esophageal averaged 98.93°F. Moderate correlation was observed (Pearson's r = 0.388; Spearman's ρ = 0.568). Bland-Altman analysis indicated a mean bias of 0.154°F, mitigated by a calibration model. Hyperthermia detection models showed improved sensitivity (55%) using machine learning approaches.*

CONCLUSION: *Calibrated axillary temperature measurements can reliably approximate core temperatures in ICU patients, enhancing patient comfort and safety while maintaining accuracy, suggesting their potential integration into standard practice.*

KEYWORDS: *Temperature management, Esophageal temperature, Axillary temperature, Intensive care unit, Concordance correlation coefficient.*

INTRODUCTION

Accurate temperature monitoring is essential for critically ill intensive care unit (ICU) patients. Even minor temperature deviations can significantly influence diagnostic and therapeutic decisions. Esophageal temperature measurement is widely considered the gold standard for assessing core body temperature owing to its precision and reliability. However, their invasive nature often results in patient discomfort and potential procedural risks, limiting their application in specific clinical contexts. This modality is particularly emphasized in post-cardiac arrest management, in which maintaining therapeutic hypothermia or normothermia is crucial for optimizing neurological outcomes and survival rates. (1,2).

In contrast, axillary temperature measurement is non-invasive and comparatively more convenient; however, its accuracy relative to esophageal measurement remains debatable. A nationwide survey of ICU practices in China demonstrated a preference for noninvasive methods, such as axillary temperature monitoring, despite ongoing concerns regarding precision (3). Previous investigations, including those involving post-cardiac surgery patients, have observed considerable variability in noninvasive temperature measurements, often influenced by ambient conditions and patient activity (4,5,6).

This ongoing debate regarding the capacity of axillary temperature measurements to consistently reflect core body temperature with the requisite accuracy for critical care forms the foundation of the present study. We aimed to ascertain whether axillary temperature measurement can be a reliable substitute for the esophageal method, offering a less invasive alternative that still provides clinically relevant data for decision-making.

The primary objective was to assess the correlation and agreement between axillary and esophageal temperature measurements in intensive care unit (ICU) patients under diverse clinical conditions. The secondary aim was to investigate the feasibility of calibrating or adjusting axillary readings to correspond more closely with esophageal measurements. These findings could establish axillary temperature measurement as a viable, non-invasive alternative that enhances patient comfort and safety without compromising the accuracy necessary for optimal care.

MATERIALS AND METHODS

Study Design

This prospective observational study was conducted in a tertiary care hospital's Intensive Care Unit (ICU). This study aimed to evaluate the correlation and agreement between axillary and esophageal temperature measurements in critically ill patients. The study period encompassed six months, from January 2023 to June 2023, encompassing various clinical conditions and temperature variations typically observed in an ICU setting. The hospital's Institutional Review Board (IRB) approved the study, and the Declaration of Helsinki conducted all procedures. Informed consent was obtained from all the participants or their legal representatives before inclusion in the study. Patient data confidentiality was strictly maintained throughout the study, and all data were anonymized and securely stored.

Participants

Participants were selected based on inclusion and exclusion criteria to ensure robust and generalizable results.

InclusionCriteria

- *Adult patients (≥ 18 years old) were admitted to the ICU.*

- *Patients requiring continuous temperature monitoring as part of their clinical management.*
- *Patients who provided informed consent or whose legal representatives consented to their behalf.*

Exclusion Criteria

- *Patients with contraindications to esophageal temperature monitoring, such as esophageal pathology or recent esophageal surgery.*
- *Patients in whom axillary temperature measurements were unfeasible or unreliable owing to factors such as significant arm injuries or deformities.*
- *Patients with conditions that could substantially alter peripheral blood flow, such as severe sepsis or circulatory shock, potentially affecting axillary temperature readings.*

Temperature Measurement Procedures

Esophageal Temperature Measurement

Esophageal temperature was measured using a calibrated esophageal temperature probe (400 Series, Smiths Medical, India) inserted into the lower third of the esophagus. This site was selected because of its proximity to the heart and capacity to accurately reflect the core body temperature. The probe was positioned by trained ICU personnel following standardized protocols to ensure correct placement and minimize patient discomfort. Continuous temperature data were recorded at 4-hourly intervals via the ICU's central monitoring system.

Axillary Temperature Measurements

Axillary temperature was measured using a high-precision digital thermometer (MC-246, Omron Healthcare, India) positioned centrally in the axilla with the arm adducted to maintain consistent readings. Each measurement was allowed to stabilize for a minimum of five minutes prior to recording. These temperatures were documented at the same 4-hourly intervals as the esophageal measurements to facilitate direct comparison.

Data Collection

Synchronous Collection*: Temperature measurements from both axillary and esophageal sites were obtained simultaneously. This concurrent collection methodology ensured that each pair of temperature readings corresponded to the same temporal point, thereby providing a robust dataset for precise comparison. The 1109 paired measurements were acquired across a heterogeneous patient population and at various time points, encompassing a comprehensive range of temperature data.*

Demographic and Clinical Data*: Supplementary data were collected for each patient, including age, sex, diagnosis, underlying medical conditions, and any interventions that could potentially influence temperature regulation (e.g., administration of antipyretics and utilization of external cooling or warming devices).*

Data Cleaning*: Temperature values outside the human physiological range were identified and excluded to maintain dataset integrity. This process was essential for providing an accurate representation of the temperature measurements and for mitigating potential biases in subsequent analyses.*

Statistical Analysis

Descriptive statistics (mean, median, standard deviation, and range) were used to summarize axillary and esophageal temperature measurements. Hyperthermia was defined as a temperature ≥ 101°F. Pearson's and Spearman's coefficients were calculated to assess the correlation between axillary and esophageal readings, while Bland-Altman analysis quantified systematic bias and limits of agreement. A paired t-test was used to determine whether differences in mean temperatures were statistically significant, and linear regression was performed to predict esophageal temperature from axillary values, with R^2 indicating explanatory power. The Concordance Correlation Coefficient (CCC) provides a comprehensive measure of both the accuracy and precision.

Multiple modeling strategies were evaluated to detect hyperthermia cases.

1. *Formula-Based Models (Original, Optimized, and Threshold-Adjusted) were assessed using accuracy, recall, and precision, given the clinical priority of identifying elevated temperatures (≥ 101°F).*

2. *Stacking Models (combining XGBoost and Random Forest) were implemented with and without SMOTE oversampling to address class imbalance, and the classification thresholds (e.g., 0.4) were tuned to optimize recall and precision.*

Calibration of axillary temperatures was conducted by applying the mean difference from the Bland-Altman analysis as a bias correction factor, supplemented by a linear regression–based calibration model. This model was validated using an independent dataset, enhancing the alignment between axillary and esophageal measurements and ensuring more reliable hyperthermia detection.

RESULTS

Demographics and Baseline Characteristics

*The study cohort comprised 1109 patients with a mean age of 58.7 ±13.2 years (range 18–92). The gender distribution was 58% (n = 643) male and 42% (n = 466) female. Sepsis constituted the primary diagnosis for 40% (n = 444) of the patients, while the remaining 60% (n = 665) were admitted for other critical conditions. Comorbidities, including hypertension (45 %, n = 499), diabetes mellitus (32 %, n = 355), chronic kidney disease (15 %, n = 166), and COPD (10 %, n = 111), were prevalent. An additional 20% (n = 222) of patients presented with other diagnoses, such as cardiac disease or malignancy (**Table 1**).*

***Table 1:** Demographic and Baseline Characteristics.*

Characteristic	***Value***
Age, years	*58.7 ± 13.2 (range 18-92)*
Sex, n (%)	*Male: 643 (58%)*
	Female: 466 (42%)
Primary Diagnosis, n (%)	*Sepsis: 444 (40%)*
	Other ICU Admissions: 665 (60%)
Comorbidities, n (%)	*Diabetes Mellitus: 355 (32%)*
	Hypertension: 499 (45%)
	Chronic Kidney Disease: 166 (15%)
	COPD: 111 (10%)
	Other (e.g., cardiac disease): 222 (20%)

Temperature Measurements and Descriptive Statistics

For the 1109 patients, 1109 paired axillary and esophageal temperature measurements were obtained. The mean axillary temperature was 99.03 °F (SD 2.49 °F), with a range of 37.3 °F to 106.34 °F, whereas esophageal measurements exhibited a mean of 98.93 °F (SD 2.03 °F), spanning 93.0–108.0 °F. Axillary readings were marginally higher on average and displayed a broader distribution than esophageal values.

Correlation and Agreement Analyses

Pearson's correlation coefficient between axillary and esophageal temperatures was 0.388, indicating a moderate linear association. Spearman's rank correlation was higher at 0.568, suggesting a stronger monotonic relationship. Bland-Altman analysis revealed a mean difference (bias) of 0.154 °F, with 95% limits of agreement from −3.099 °F to 3.407 °F. A paired t-test (t = 3.093, p = 0.0020) confirmed a statistically significant difference between the axillary and esophageal temperature means. Linear regression (Esophageal Temperature = 0.752 × Axillary Temperature + 24.392) accounted for 37.5% of the variance (R^2 = 0.375), and the Concordance Correlation Coefficient (CCC) was 0.546, indicating moderate agreement ***(Figure 1).***

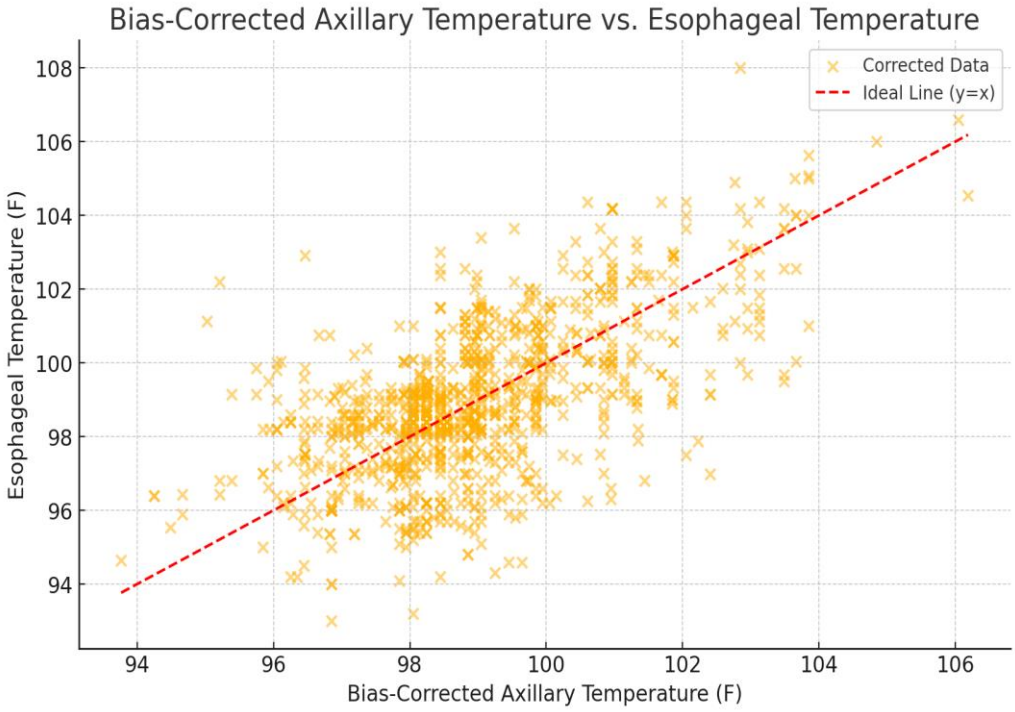

Figure 1:

Regression plot between Axillary vs Esophageal Temperatures.

Bias Correction and Calibration

*To enhance the accuracy of axillary readings, a bias correction factor of 0.154 °F, derived from the Bland-Altman analysis, was subtracted from the axillary measurements. This adjustment reduced the postcorrection bias to approximately zero. A linear regression–based calibration model validated on an independent dataset further improved the alignment between axillary and esophageal temperatures (****Figure 2****).*

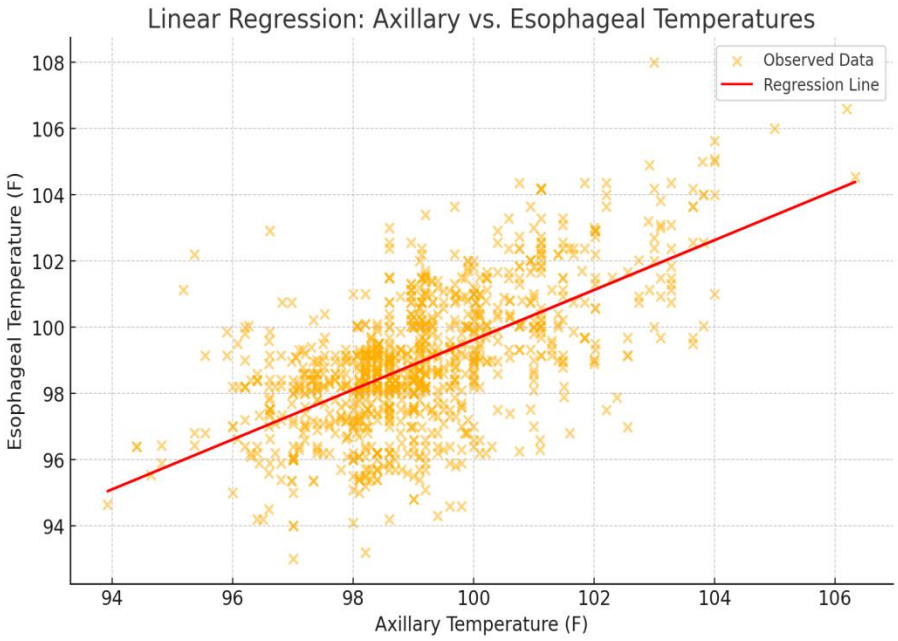

***Figure 2**: Bias-Corrected plot between Axillary vs Esophageal Temperature.*

Hyperthermia Detection (≥ 101°F)

Prevalence of Hyperthermia

Out of 1109 total measurements, 232 (15%) were classified as hyperthermia cases (≥ 101 °F) and 934 (85%) as non-hyperthermia cases.

Formula-Based Models

*The original formula achieved 84.25% accuracy but exhibited a low recall (29%), resulting in 40 undetected hyperthermia cases. An optimized version improved recall to 41% (33 undetected hyperthermia cases), and a threshold adjustment (100.4 °F) further increased recall to 46%. However, these incremental improvements did not achieve the level of sensitivity deemed sufficient for all clinical scenarios, as evidenced by the limited F1-scores (ranging from 41% to 54%)**(Table 2).***

***Table 2:** Performance Metrics Across Models*

Model	Accuracy	Recall (Hyperthermia)	Precision (Hyperthermia)	Specificity	F1-Score (Hyperthermia)
Original Formula	84.25%	29%	63%	97%	41%
Optimized Formula	85.27%	41%	70%	96%	52%
Formula (Threshold = 100.4˚F)	84.59%	46%	63%	94%	54%
Stacking (Threshold = 0.4)	84.59%	55%	61%	92%	58%
Stacking (No SMOTE)	84.95%	46%	65%	94%	54%
XGBoost (Pre-Stacking)	83.90%	48%	62%	93%	54%

Stacking Models

Stacking models that combined XGBoost and Random Forest demonstrated a higher recall (up to 55%) than formula-based approaches. The introduction of SMOTE oversampling addressed class imbalance but elevated the rate of false positives, reducing the specificity from 94% to 92%. A threshold was set at 0.4 balanced recall (55%), and precision (61%), yielding an F1-score of 58%. Without SMOTE, stacking achieved a slightly higher specificity (94%) and precision (65%), but the recall decreased to 46%.

XGBoost (Pre-Stacking)

XGBoost alone attained an accuracy of 83.90% and a recall of 48%. Although these metrics were comparable to those of certain formula-based models, the stacking framework ultimately enhanced the performance for hyperthermia detection.

DISCUSSION

In this study, axillary and esophageal temperatures were measured in 1109 critically ill patients, revealing a moderate correlation (Pearson's r = 0.388; Spearman's ρ = 0.568) and a mean bias of 0.154 °F in Bland-Altman analysis. A bias correction factor reduced this discrepancy to nearly 0 °F, and a linear regression–based calibration model further strengthened the alignment between the two measurement sites. These adjustments, along with ensemble machine-learning approaches (e.g., stacking XGBoost and Random Forest), improved the detection of hyperthermia (≥ 101 °F), demonstrating that calibrated axillary measurements can reliably approximate core temperature in the ICU setting.

Our finding of a moderate correlation between axillary and esophageal temperatures, along with the reduced bias achieved through calibration, aligns with previous investigations demonstrating that axillary measurements can serve as reliable proxies for core temperature when appropriate statistical corrections are applied. (7) Using Bland-Altman analysis to assess agreement is consistent with established practices. (8) It has been applied across various contexts, from temperature comparisons in veterinary settings (9) to non-invasive monitoring in human patients. (10) Although some studies have focused on pediatric populations or alternate measurement sites (e.g., rectal temperatures in newborns) (11,12), the principle remains that robust calibration models can significantly narrow the gap between noninvasive and core-temperature readings.

Recent research has also highlighted the potential of machine learning methods, such as stacking XGBoost and Random Forest, to enhance diagnostic accuracy. (7,13) This approach parallels our results, where integrating advanced modeling techniques further refined hyperthermia detection (≥ 101 °F). In ICU-based temperature management surveys (14) and broader investigations of fever thresholds (15), non-invasive measurement strategies frequently appear more practical for routine monitoring, underscoring the clinical value of improved axillary calibration. These studies support the hypothesis that the judicious application of bias correction alongside modern analytical methods can render axillary temperatures a more viable alternative to esophageal monitoring in critically ill patients.

Predictability and Clinical Application In the intensive care unit (ICU), accurate assessment of body temperature is critical for guiding clinical interventions and ensuring patient safety. The moderate correlation (R-squared =

0.375; CCC = 0.546) between axillary and esophageal measurements indicates that axillary readings account for a substantial proportion of the variance in esophageal temperature, yet do not fully correspond to core temperature values. This underscores the necessity of calibration to enhance predictive accuracy. Despite these limitations, axillary monitoring presents significant advantages in terms of patient comfort and reduced procedural risks. It may be particularly advantageous for patients who are unable to tolerate invasive methods or those at an elevated risk for complications, such as esophageal trauma or infection.

From a procedural perspective, the implementation of axillary monitoring can enhance the efficiency of temperature assessment by reducing the time and expertise required for probe placement, which may be particularly beneficial in high-acuity ICUs that manage frequent admissions and complex interventions. Kahn et al. (2020) and others have demonstrated that noninvasive monitoring strategies can improve patient outcomes by minimizing discomfort and procedure-related risks, thereby enabling clinicians to focus on other critical aspects of care. When precise core temperature data are essential, such as during active temperature management in sepsis or post-cardiac arrest care, esophageal measurement may still be warranted owing to its superior accuracy. In these scenarios, axillary monitoring can function as a complementary method, with calibrated readings providing a supportive perspective on temperature trends. (16)

Developing and integrating calibration models, as demonstrated by Hsieh et al. (2021), can mitigate discrepancies between axillary and esophageal measurements, resulting in a more accurate detection of fever or hypothermia. (17) Subsequently, these models can be refined using larger, more diverse datasets and potentially enhanced by machine learning techniques. (18,19) Such advancements hold promise for further improving the consistency of axillary monitoring, rendering it an increasingly viable tool for a range of clinical conditions while preserving patient comfort and optimizing the ICU workflow.

CONCLUSION

The present study's findings indicate that appropriately calibrated axillary temperature measurements can function as a viable noninvasive alternative to esophageal monitoring in intensive care unit patients. Although axillary readings inherently exhibit greater variability and a slight bias compared with esophageal methods, the application of a calibration model significantly improves their concordance with core temperature measurements. This approach has the potential to optimize temperature monitoring processes, minimize patient discomfort, and mitigate the risks associated with more invasive procedures. Subsequent research should focus on refining these calibration techniques and evaluating their efficacy across diverse clinical settings and patient populations, with the ultimate objective of incorporating robust calibrated axillary measurements into standard critical care practice.

Acknowledgement

We are grateful to the Dean IMS, and SUM Hospital, Bhubaneswar for the extended research facility atSUMHospital. The authors acknowledge Dr. Debasmita Dubey, MRL, IMS and SUM Hospital, Siksha 'O' Anusandhan, Deemed to be University for providing necessary facilities and supports.

Author contributions

SC, AD, SS, PC, SP *and* ***SBM*** *contributed to study design. SC wrote the manuscript. SBM and SKP analyzed the data and revised the manuscript. This study was being conducted under supervision of AD. Finally, the manuscript was being reviewed and approved by all authors.*

Data availability

The data sets analyzed in this study are accessible from corresponding author on valedictory reasonable request.

Funding

This study was being conducted without funding.

Ethics approval

Ethical approval was being obtained from Institutional Review Board (IRB) of the hospital before starting of the study. Informed consents were also obtained from each participant or legal representatives before inclusion to this study.

Conflict of interest

The authors have declared no conflict of interest.

REFERENCES:

1. *Scirica BM. Therapeutic hypothermia after cardiac arrest. Circulation. 2013;127(2):244-50.*
2. *Kim F, Nichol G, Maynard C, Hallstrom A, Kudenchuk PJ, Rea T, Copass MK, Carlbom D, Deem S, Longstreth WT, Olsufka M. Effect of prehospital induction of mild hypothermia on survival and neurological status among adults with cardiac arrest: a randomized clinical trial. Jama. 1;311(1):45-52.*
3. *Meng L, Wang C, Liu X, Bi Y, Zhu K, Yue Y, Wang C, Song X. Temperature management in the intensive care unit: a practical survey from China. Libyan Journal of Medicine. 2023;18(1):2275416.*
4. *Engelbart G, Brandt S, Scheeren T, Tzabazis A, Kimberger O, Kellner P. Accuracy of non-invasive sensors measuring core body temperature in cardiac surgery ICU patients–results from a monocentric prospective observational study. Journal of Clinical Monitoring and Computing. 2023;37(6):1619-26.*
5. *van der Kooi AW, Kappen TH, Raijmakers RJ, Zaal IJ, Slooter AJ. Temperature variability during delirium in ICU patients: an observational study. PloS one. 2013 Oct 23;8(10):e78923.*
6. *Insler SR, O'Connor MS, Leventhal MJ, Nelson DR. Association between perioperative hypothermia and adverse outcomes after cardiac surgery. Anesthesiology. 2011;114(2):376-384.*
7. *Gerke, O. Reporting standards for a bland–altman agreement analysis: a review of methodological reviews. Diagnostics. 2020;10(5),334.*
8. *Timbol, J. Determining loa between left ventricular ef estimated from abarquez formula and left ventricular ef measured by 2d echo. Asian Journal of Medical Sciences. 2020;11(5), 12-17.*
9. *Zakari FO, Omontese BO. Comparison of axillary and inguinal temperature with rectal temperature in dogs at a veterinary teaching hospital. Journal of Small Animal Practice. 2023 May;64(5):330-6.*

10. *Sang B., Lee C., & Lee D. Prospective comparative analysis of noninvasive body temperature monitoring using zero heat flux technology (spoton sensor) compared with esophageal temperature monitoring during pediatric surgery. Plos One 2022;17(8):e0272720.*
11. *Müller P., Berkel L., & Beaufort A. Axillary and rectal temperature measurements poorly agree in newborn infants. Neonatology. 2008;94(1):31-34.*
12. *Hoffman L., Santos M., Tucker R., &Laptook A. Neonatal oesophageal and axilla temperatures in the neonatal intensive care unit care. Acta Paediatrica. 2015;104(12).*
13. *Talwar A, Willaime JM, Pickup LC, Enescu M, Boukerroui D, Hickes W, Rahman NM, Gooding MJ, Kadir T, Gleeson FV. Pulmonary nodules: assessing the imaging biomarkers of malignancy in a "coffee-break". European Journal of Radiology. 2018 Apr 1;101:82-6.*
14. *Meng L., Temperature management in the intensive care unit: a practical survey from china. Libyan Journal of Medicine. 2023;18(1).*
15. *Lai F, Li X, Liu T, Wang X, Wang Q, Chen S, Wei S, Xiong Y, Hou Q, Zeng X, Yang Y. Optimal diagnostic fever thresholds using non-contact infrared thermometers under COVID-19. Frontiers in Public Health. 2022;10:985553.*
16. *Madison JM, Irwin RS. Chronic cough and COPD. Chest. 2020 ;157(6):1399-400.*
17. *Ferré-Bergadà M, Valls A, Raigal-Aran L, Lorca-Cabrera J, Albacar-Riobóo N, Lluch-Canut T, Ferré-Grau C. A method to determine a personalized set of online exercises for improving the positive mental health of a caregiver of a chronically ill patient. BMC Medical Informatics and Decision Making. 2021;21:1-3.*
18. *Johnson AE, Ghassemi MM, Nemati S, Niehaus KE, Clifton DA, Clifford GD. Machine learning and decision support in critical care. Proceedings of the IEEE. 2016 Jan 25;104(2):444-66.*
19. *Dong JZ, Wang LP, Zhu JH. Appropriate Antibiotic Therapy and WBC in Sepsis: Factors to Consider. Critical Care Medicine. 2016 1;44(4):e231.*