Original Investigation

Radiologists are Human Too: Addressing the Impacts of Occupational Exhaustion, Travel History, and Hunger on Mammography Reading Performance

Abdulaziz S. Alshabibi, PhD, Sultan F. Alhujaili, PhD, Basel Qenam, PhD, Areej Aloufi, PhD, Salman M. Albeshan, PhD, Meaad M. Almusined, PhD, Abdulmajeed Alotabibi, PhD, Nuha A. Khoumais, MD, MBA-IHM

Rationale and Objectives: To investigate how occupational exhaustion, travel, and hunger affect mammography reading performance among radiologists and radiology trainees.

Methods: Thirty participants (22 radiologists, eight radiology trainees) completed mammography reading assessments using the DetectedX platform during two radiology conferences in Saudi Arabia. Each participant interpreted 30 de-identified mammographic cases (15 abnormal, 15 normal) under standardized conditions. Performance was measured using jackknife alternative free-response receiver operating characteristic (JAFROC), lesion sensitivity, area under the receiver operating characteristic curve (ROC AUC), sensitivity, and specificity. Three independent variables were self-reported using a questionnaire completed immediately before the mammography reading session: occupational exhaustion (low to moderate or high, assessed by the Emotional Exhaustion subscale of the Maslach Burnout Inventory–General Survey), recent travel (traveler or non-traveler), and hunger status (hungry or not hungry). Mann–Whitney U tests were used to examine differences in reading performance associated with each variable.

Results: Participants with high occupational exhaustion had significantly lower JAFROC scores compared to those with low to moderate exhaustion (0.213 vs 0.383; $p = 0.041$). Recently traveled participants had significantly lower ROC AUC scores (0.681 vs. 0.772; $p = 0.03$) and lower sensitivity (70.0% vs. 80.0%; $p = 0.04$) than non-travelers. Hungry participants exhibited higher sensitivity (85.0% vs. 70.0%; $p = 0.04$) but lower specificity (40.0% vs. 65.0%; $p = 0.02$) compared to non-hungry participants.

Conclusion: Occupational exhaustion was associated with poorer JAFROC scores, recent travel lowered sensitivity and overall diagnostic accuracy, and hunger increased sensitivity at the cost of specificity. These findings highlight the importance of addressing radiologists' well-being by optimizing workloads, allowing recovery periods after travel, and ensuring structured meal schedules. Future research should explore real-world clinical settings and targeted interventions, including AI integration.

Acad Radiol 2025; 32:4408–4420

From the Radiological Sciences Department, College of Applied Medical Sciences, King Saud University, 145111, Riyadh, Saudi Arabia (A.S.A., B.Q., A.A., S.M.A., M.M.A.); Radiation Sciences and Medical Imaging Department, College of Applied Medical Sciences, Jouf University, Aljouf 72388, Saudi Arabia (S.F.A.); Department of Radiological Sciences, College of Applied Medical Sciences, King Saud bin, Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia (A.A.); King Abdullah International Medical Research Centre, National Guard Health Affairs, Riyadh, Saudi Arabia (A.A.); King Faisal Specialist Hospital and Research Center, Department of Radiology, Riyadh, Saudi Arabia (N.A.K.). Received January 14, 2025; revised April 25, 2025; accepted April 28, 2025. **Address correspondence to:** A.S.A. e-mail: abdalshabibi@ksu.edu.sa

https://doi.org/10.1016/j.acra.2025.04.067

INTRODUCTION

Despite the digital evolution of mammography screening, its accuracy remains dependent on human interpretation. Since Garland's landmark 1949 study, retrospective analyses have consistently revealed substantial error rates of ~30% (1–3), with rates of 3–5% observed in daily practice (2,4), mainly from cognitive errors and perceptual or search errors (5,6). Perceptual errors occur when abnormalities are initially overlooked but detected upon review, such as missing a subtle mass due to fatigue; cognitive errors arise when abnormalities are detected but their significance is misinterpreted, such as dismissing an architectural distortion (8,9).

Previous research on errors has focused on visual search patterns. The Nodine–Kundel three-phase model describes how radiologists analyze 2D images as follows: pattern recognition, focused attention, and decision-making (7,8). In the first phase, expert radiologists can identify abnormalities within ~0.25 s through holistic viewing; this is followed by detailed examination of potential targets and a comprehensive re-scan and concluded with diagnostic decision-making (9). Eye fixation experiments have revealed three main types of errors: search (failure to fixate on lesions; 30%), recognition (insufficient fixation time; 25%), and decision (adequate fixation but incorrect interpretation; 45%) (13).

Research has also found significant variations in performance between radiologists, with numerous contributing factors (10–12), including annual mammogram volumes (13), training quality, experience (14), fatigue (15,16), sleep patterns, duration of wakefulness (17), time of day (18,19), lesion characteristics (20), and even social networking among radiologists (21). Further potential influences therefore warrant investigation, including occupational exhaustion, travel, and hunger, which have been shown to impact competence in other fields (22–24) but are thus far under-investigated in radiological practice.

Occupational exhaustion, for instance, is one of the critical components and initiators of burnout (25–29). It often results from prolonged stress and heavy workloads and is therefore particularly relevant in healthcare settings, where mental and emotional depletion can lead to diminished cognitive resources and increased risk of errors (25–29). The demands–resources theory provides a framework for understanding occupational exhaustion, positing that it results from an imbalance between job demands and available resources (27). In radiology, these demands may include excessive workload, emotional labor, and interpersonal conflicts, which all require sustained effort and can incur physiological and psychological costs. Insufficient recovery from such demands can trigger exhaustion, and when demands consistently outweigh resources, chronic fatigue may develop, leading to occupational exhaustion.

Similarly, the effects of travel, such as disrupted routines and jet lag, can also influence alertness and attention to detail (24). Radiologists are increasingly required to travel for various reasons, including conferences, specialized consultations, and global health initiatives. Though essential for professional development and collaboration, this can disrupt work schedules and potentially affect cognitive function. Travel fatigue, characterized by general fatigue, disorientation, and weariness, often occurs during or immediately after travel (30) and can be exacerbated by other travel stressors, such as long queues, delays, and security checks. Irregular meals, limited access to preferred food, and disrupted sleep due to schedules or sleeping environments, can also strain health and impair cognitive function (30).

Furthermore, physiological hunger has been shown in other fields to alter decision-making processes and risk assessments (18,19,23,31,32). For example, studies of parole hearings have found that judges tend to make more lenient decisions early in the day and shortly after meals than later in the day or just before eating, indicating increased risk aversion. Similar risk-averse trends have been observed among audiologists evaluating patients and individuals engaged in gambling, suggesting a broader phenomenon of risk-averse decision-making in response to hunger. In radiology, this could manifest as a tendency to interpret ambiguous findings as potentially malignant, thus reducing the risk of missing tumors—an error that impacts both patients and litigation risks.

Several theoretical frameworks offer explanations for this phenomenon. One is the hot–cold empathy gap theory, which posits that individuals struggle to accurately predict or empathize with decision-making processes across different psychological states (33); thus, radiologists in a "hot" state, such as hunger, would find it harder to replicate the rational decision-making of a "cold" or calm—in this case, satiated—state. Hungry radiologists would instead respond to an instinct to avoid risks under perceived scarcity, making less logical and more cautious decisions. This shift in decision-making may be unconscious and difficult to mitigate, even for decisions unrelated to hunger.

Alternatively, ego-depletion theory (34,35) argues that cognitive resources are limited and can be depleted by states like hunger (35). For example, hungry participants with low glucose levels may experience negative mood and fatigue (36–38), even while the hunger hormone ghrelin increases stress and anxiety through cortisol elevation (39–43). These physiological changes would then reduce available cognitive resources, leading to greater reliance on emotional cues and heuristics and thus to intuitive rather than deliberate decision-making (41). This means a more cautious approach, resulting in more cases recalled for further investigation rather than risking missed cancers.

Given the potential impact of occupational exhaustion, travel, and hunger on radiologists, this study aimed to investigate how they influence performance in mammography interpretation.

MATERIALS AND METHODS

This study used self-reported and performance data of radiologists from Saudi Arabia and few other neighboring countries during the 13th Annual Radiological Society of Saudi Arabia (RSSA) conference in Jeddah (February 9–11, 2023) and the 14th RSSA conference in Riyadh (December 14–16, 2023). Participants were practicing radiologists and trainees who also completed a survey on factors such as occupational exhaustion, travel, and hunger. Data collection was integrated into the conference activities, enabling efficient gathering of real-world performance data from practicing radiologists at various career stages and experience levels in a standardized setting. The online platform DetectedX (44) was used, which provides standardized sets of pre-diagnosed, de-identified mammography images to allow direct comparisons of radiologists' performance metrics and performance-influencing factors.

Ethical Approval, Informed Consent, and Confidentiality

Ethical approval was secured from the Human Research Ethics Committee, project number E-24–9079. Participants consented to the use of de-identified data for research purposes; the data from DetectedX were anonymized, with radiologists identified only by reader numbers.

DetectedX Overview

DetectedX is a digital screen-reading program designed to evaluate radiologists' performance in detecting breast cancer and cancer-free cases. The test set consisted of 30 de-identified mammographic cases (15 abnormal and 15 normal), each comprising the following two views: cranio-caudal and medio-lateral oblique. Normal cases were those not recalled at screening or recalled and verified as benign. Abnormal cases were those recalled at screening and confirmed histopathologically as cancerous. Figure1 presents an example of the mammography images on the DetectedX platform.

Test Set Reading and Environment

The mammogram readings were conducted between 8:00 a.m. and 6:00 p.m. The participants selected time slots based on their availability, so the session time allocation was not randomized. Prior to their session, participants viewed a web-based instructional video on using the DetectedX platform, covering the mammogram review process, rating system, and image manipulation features like zoom, pan, and brightness/contrast adjustments. No clinical histories or information regarding the distribution of normal versus abnormal cases were provided to the participants. To simulate clinical conditions, the reading environment was standardized, with ambient light levels maintained at 30–40 lux and images displayed on an LED backlight monitor with a resolution of 3840×2160 pixels.

Each radiologist was allotted around 2 h to evaluate the test set. They were instructed to mark identifiable lesions and to assign ratings using a four-point confidence scale as follows: 2 (benign), 3 (equivocal), 4 (suspicious), and 5 (unquestioned malignancy). For cases deemed to have no significant abnormalities, the radiologists proceeded to the next case by selecting the "NEXT" button. For binary classification analysis (calculating sensitivity and specificity), the responses were categorized as follows: selecting "NEXT" or assigning a score of 2 (benign lesion) was considered a negative finding (normal), while scores of 3 to 5 were considered positive findings (abnormal–breast cancer). A response was deemed correct when either a negative finding matched a normal case or a positive finding matched an abnormal case. The software allowed retrospective review and for any decision to be changed until the final submission.

Performance Metrics

Upon completion of the test set, the system immediately analyzed the data and provided instant feedback for the participants.

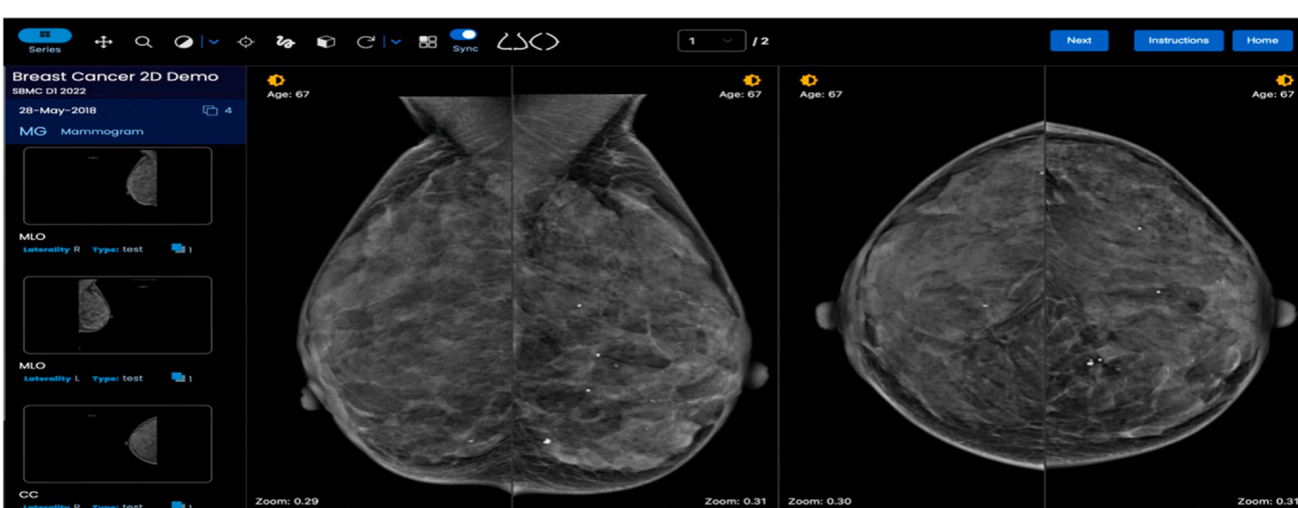

Figure 1. DetectedX platform interface showing standard two-view digital mammogram: CC and MLO views with image manipulation tools and navigation controls. CC, cranio-caudal; MLO, medio-lateral oblique.

The performance metrics that were generated and are included in this study were the jackknife alternative free-response receiver operating characteristic (JAFROC) curve, lesion sensitivity, area under the receiver operating characteristic curve (ROC AUC), sensitivity, and specificity. These are defined as follows:

Sensitivity: The proportion of actual positive cases in the test set that are correctly identified.

Specificity: The proportion of actual negative cases in the test set that are correctly identified.

ROC AUC: A single comprehensive measure of the overall ability to discriminate between normal and abnormal cases. This metric combines sensitivity and specificity by plotting the true-positive rate against the false-positive rate across different confidence thresholds. The resulting value ranges from 0.5 (performance at the level of chance) to 1.0 (perfect diagnostic accuracy).

Lesion sensitivity: The proportion of actual malignant breast lesions in the test set that are correctly identified and localized within a 50-pixel radius of the true lesion center.

JAFROC: A measure of performance in detecting and localizing lesions and identifying negative cases. It is calculated from the alternative free-response receiver operating characteristic curve, which plots the lesion localization fraction (LLF) against the false-positive fraction (FPF).

- **LLF:** The fraction of true-positive decisions with the correct localization.
- **FPF:** The false-positive fraction, which is *1−specificity*.

DetectedX also collected data on the participants' professional roles (radiologists vs. radiology trainees) and their years of experience in those roles.

Independent Variables: Occupational Exhaustion, Travel Status, and Hunger

Three independent variables were self-reported using a questionnaire completed immediately before the mammography reading session: occupational exhaustion, recent travel, and hunger status. Occupational exhaustion was assessed using the Emotional Exhaustion subscale of the Maslach Burnout Inventory–General Survey (MBI-GS). The MBI-GS has demonstrated strong cross-cultural validity, with a reliability coefficient of 0.89 for the Emotional Exhaustion subscale (45). This subscale includes seven items aimed at evaluating feelings of fatigue, emotional depletion, and burnout. The specific items are the following:

1. I feel emotionally drained by my work.
2. Working with people all day long requires a great deal of effort.
3. I feel like my work is breaking me down.
4. I feel frustrated by my work.
5. I feel I work too hard at my job.
6. It stresses me too much to work in direct contact with people.
7. I feel like I am at the end of my rope.

The participants responded to each item using a seven-point Likert scale, ranging from 0 (never) to 6 (every day). The total score across these seven items was calculated for each participant, with higher scores reflecting greater levels of occupational exhaustion. Participants scoring 30 or less were classified as experiencing low to moderate exhaustion, while those scoring above 30 were considered to have high exhaustion levels (46). In addition to the MBI-GS, two supplementary questions were included in the assessment. First, the participants were asked a binary (yes/no) question regarding whether they had traveled far to attend the conference, capturing their travel status. Second, immediately prior to the reading assessment, the participants were asked to indicate (yes/no) whether they felt hungry at that moment.

Statistical Analysis

Data analysis was conducted using IBM-SPSS Statistics version 29 to evaluate the potential impact of readers' occupational exhaustion, hunger, and travel on mammography reading performance using various performance metrics, including sensitivity, specificity, lesion sensitivity, ROC AUC, and JAFROC. The statistical significance level was 0.05 for all tests. Non-parametric tests were applied after initial assessment with the Kolmogorov-Smirnov test revealed a non-normal data distribution.

The preliminary analysis began with Mann–Whitney U tests of differences between the readers' professional roles (categorical variable: radiologists and radiology trainees) and their performance metrics. We also used Spearman's rank correlation coefficient to explore potential significant correlations between the readers' experience levels in their professional roles, measured in years (continuous variable), and their performance metrics. These two examinations aimed to identify statistically significant differences or correlations and inform the selection or exclusion of potential covariates.

The primary analysis evaluated whether reading performance varied according to occupational exhaustion (low to moderate exhaustion = 0, high exhaustion = 1), hunger status (0 = not hungry, 1 = hungry), and long-distance travel to attend the conference (0 = non-traveler, 1 = traveler) using Mann–Whitney U tests. This non-parametric test was selected for its robustness; its ability to compare independent groups effectively, even with small or unequal sample sizes; and its suitability for data that do not meet parametric test assumptions, such as normal distribution (47–51).

RESULTS

Participant Characteristics

Initially, 32 participants were recruited (22 radiologists, eight radiology trainees, and two radiographers). The radiographers were excluded from the analysis because the study focused specifically on radiologists and radiology trainees. Of the 30 remaining participants, one radiologist's data was only partially analyzed due to their selection of the screening-only option in DetectedX, which allows for binary (positive/negative) case classifications without lesion localization or the four-point rating scale. Thus, while this radiologist's data was included in the analyses of sensitivity and specificity ($n = 30$), it was excluded from ROC, JAFROC, and lesion sensitivity ($n = 29$).

The median experience of the 30 participants was 5.50 years (interquartile range [IQR]: 2.75–10.75 years), within which the eight radiology trainees had a median experience of 3.50 years (IQR: 1.25–4.00 years).

Preliminary Analyses

Performance Comparison of Radiologists and Radiology Trainees

The results of Mann–Whitney U tests examining the differences in performance metrics between radiologists and radiology trainees are presented in Table 1. No significant differences were found.

Performance Correlations with Reader Experience

Spearman correlation analyses between the participants' experience (measured in years of service in their professional role as either radiologist or radiology trainee) and the performance metrics are presented in Table 2. No significant correlations were found when analyzing the participants collectively or separating them by professional role.

The preliminary analyses thus found no significant differences in performance between radiologists and radiology trainees (Mann–Whitney U tests) nor any significant correlations between years of professional experience and the performance metrics (Spearman's rank correlations). Consequently, neither professional role nor experience level were included as covariates in the primary analysis.

Participant Distributions Across Independent Variables

The proportions of radiologists in each category were unbalanced (Table 3), which can reduce statistical power. Under such conditions, near-significant findings may indicate obscured effects that should be tested in future research.

Primary Analyses: Effects of Occupational Exhaustion, Travel, and Hunger on Performance Metrics

Mann–Whitney U tests were performed to examine differences in performance metrics according to occupational exhaustion, travel status, and hunger (Table 4).

Occupational Exhaustion

Occupational exhaustion significantly affected JAFROC scores (U = 31.0, p = 0.041), with participants experiencing high exhaustion demonstrating lower JAFROC scores than those with low-to-middle levels of exhaustion (Fig 2). No significant differences were observed for lesion sensitivity, sensitivity, or specificity.

Travel

Travel status significantly impacted ROC (U = 40.0, p = 0.03) and sensitivity (U = 45.0, p = 0.04), with participants who traveled exhibiting lower ROC scores (Fig 3) and lower sensitivity (Fig 4) than those who did not. No significant differences were found for JAFROC, lesion sensitivity, or specificity.

Hunger

Hunger significantly affected sensitivity (U = 131.0, p = 0.04) and specificity (U = 39.5, p = 0.02), with hungry participants demonstrating higher sensitivity (Fig 5) but lower specificity (Fig 6) than non-hungry participants. No significant differences were observed for JAFROC, lesion sensitivity, or ROC.

DISCUSSION

This study aimed to assess the influence of three human factors—occupational exhaustion, recent travel, and hunger—on the mammogram reading performance of radiologists and trainees. The results show that occupational exhaustion negatively impacts JAFROC scores, travel decreases sensitivity and ROC scores, and hunger creates a

TABLE 1. Comparison of Diagnostic Performance Metrics Between Radiologists (*n* = 22) and Radiology Trainees (*n* = 8): Mann–Whitney U Test Results for JAFROC, Lesion Sensitivity, ROC, Sensitivity, and Specificity

	Radiology Trainees Median (IQR)	Radiologists Median (IQR)	Mann–Whitney U	p*
JAFROC	0.33 (0.21–0.41)	0.41 (0.24–0.55)	104.0	0.349
Lesion sensitivity	31.81 (18.18–52.27)	54.55 (31.81–59.09)	113.0	0.168
ROC	0.75 (0.645–0.771)	0.77 (0.68–0.84)	101.0	0.429
Sensitivity	80.0 (70.0–87.50)	80.0 (70.0–90.0)	85.0	0.909
Specificity	52.5 (31.25–65.0)	60.0 (45.0–75.0)	114.5	0.219

ROC, receiver operating characteristic; JAFROC, jackknife alternative free-response ROC; IQR, interquartile range. Sample size for radiology trainees is eight; for radiologists, it is 22 for sensitivity and specificity and 21 for JAFROC, lesion sensitivity, and ROC due to missing data for one radiologist for these metrics.

* Two-tailed test.

TABLE 2. Spearman Correlation Analysis of Reader Experience and Performance Metrics: Examining Relationships Across All Readers (n = 30), Radiologists (n = 22), and Trainees (n = 8)

Reader experience	JAFROC		Lesion Sensitivity		ROC		Sensitivity		Specificity	
	r	p	r	p	r	p	r	p	r	p
All readers	0.005	0.980	0.293	0.123	0.119	0.540	0.106	0.578	0.119	0.532
Radiologists	−0.047	0.840	0.221	0.336	0.098	0.673	0.123	0.587	0.024	0.915
Radiology trainees	−0.434	0.282	0.262	0.531	0.0001	1.000	0.092	0.828	0.231	0.582

JAFROC, jackknife alternative free-response ROC; ROC, receiver operating characteristic; 'r' represents the Spearman correlation coefficient (ranges from −1 to +1), and 'p' represents the p-value for statistical significance.

Note: Sample size for radiology trainees is eight; for radiologists, it is 22 for sensitivity and specificity and 21 for JAFROC, lesion sensitivity, and ROC due to missing data for one radiologist for these metrics.

TABLE 3. Sample Distribution by Human Factors: Participant Numbers Across Occupational Exhaustion Levels (Low to Middle or High), Travel Status (Travelers or Non-Travelers), and Hunger Status (Hungry or Not Hungry)

		Number of Readers
Occupational exhaustion	Low to middle	24
	High	6
Travel	Not traveled	22
	Traveled	8
Hunger	Not hungry	22
	Hungry	8

trade-off by increasing sensitivity while decreasing specificity.

Highly exhausted readers exhibited significantly lower JAFROC scores than their low and moderately exhausted counterparts, which is consistent with literature that links occupational exhaustion—a primary symptom and key element of burnout (27,45,52,53)—to adverse outcomes for physicians, including increased medical errors (25,26). Exhaustion may thus negatively impact both the accuracy of lesion localization and the ability to identify normal mammographic findings, systematically reducing the quality and value of readings. This finding adds to prior general observations of the prevalence of occupational exhaustion among breast radiologists (54,55).

Breast imaging presents unique challenges among the radiological subspecialties, including excessive work demands, fear of litigation, and stressful interactions with administrators, colleagues, and patients (56). In the broader context of radiology, occupational exhaustion stems from professional challenges, including long hours in static positions, high caseloads, and constant adaptation to evolving technologies. Increasing volumes and non-reporting responsibilities, such as multidisciplinary tumor conferences, further strain the limited pool of radiologists, leading to excessive and irregular working hours (57,58).

These factors contribute to occupational exhaustion and may compromise both the well-being and professional performance of radiologists. The lower JAFROC scores among exhausted radiologists in the current study suggest they may expend increased emotional and cognitive effort to maintain performance, potentially leading to errors or oversights. This is consistent with demands–resources theory (27,29), which predicts that the balance between job demands and available resources directly influences diagnostic accuracy. The multifaceted landscape of stressors points to an urgent need for organizational interventions to address workload management, provide adequate resources, and promote radiologist well-being in order to reduce occupational exhaustion and maintain high standards of patient care.

Radiologists who had recently traveled were also less accurate and more likely to miss diagnoses; this extends the travel fatigue literature (24,30) by providing specific evidence of its impact on radiological diagnosis. The decreases in ROC and sensitivity also suggest that the effect of travel is more complex than previously understood. While sleep disruptions and circadian rhythm problems are well-known (18,59), the current findings from a national conference suggest that the stresses of travel itself—queues, security checks, disrupted meals, trekking through large airports with luggage—may have separate and measurable effects on diagnostic accuracy. Appropriate adaptations may include modifying work schedules, optimizing travel arrangements, or providing recovery periods. Travel stressors may also include psychological strains, such as prolonged immobility, limited social interaction, and crowded vehicles. Though less studied than circadian disruptions, all these stressors warrant further investigation (24,60).

Although we asked the participants if they traveled to attend the conference, we did not query the distance, direction, or time zone. The participants came from locations across Saudi Arabia and nearby countries, with minimal time zone differences, yet even those small variations may have produced sleep disruptions or jet lag, which were not included in the analysis. Nevertheless, lower performance among primarily domestic travelers highlights the impact of travel-related factors, suggesting that even short-distance travel can affect diagnostic accuracy and raising further questions about the impact of longer distances, multiple time zones, and international travel. Future studies should

TABLE 4. Impact of Human Factors on Diagnostic Performance: Mann–Whitney U Test Results Comparing Performance Metrics Across Occupational Exhaustion, Travel Status, and Hunger Status Groups

Occupational exhaustion				
	Low to Moderate Median (IQR)	High Median (IQR)	Mann–Whitney U	p
JAFROC	0.383 (0.287–0.552)	0.213 (0.163–0.422)	31.0	0.041*
Lesion sensitivity	45.45 (27.27–54.55)	40.91 (18.18–63.64)	65.5	0.854
ROC	0.772 (0.685–0.835)	0.717 (0.670–0.790)	53.5	0.414
Sensitivity	80.0 (70.0–90.0)	85.0 (75.0–100.0)	100.0	0.158
Specificity	60.0 (45.0–75.0)	40.0 (30.0–65.0)	35.0	0.057
Travel				
	Not traveled Median (IQR)	Traveled Median (IQR)	Mann–Whitney U	p
JAFROC	0.365 (0.246–0.551)	0.335 (0.209–0.537)	76.0	0.72
Lesion sensitivity	54.55 (36.36–63.64)	27.27 (18.18–54.55)	47.5	0.07
ROC	0.772 (0.721–.842)	0.681 (0.631–0.785)	40.0	0.03*
Sensitivity	80.0 (70.0–90.0)	70.0 (62.5–77.5)	45.0	0.04*
Specificity	55.0 (45.0–72.5)	60.0 (43.7–72.5)	88.0	1.0
Hunger				
	Not hungry Median (IQR)	Hungry Median (IQR)	Mann–Whitney U	p
JAFROC	0.360 (0.267–0.551)	0.425 (0.187–0.509)	79.0	0.83
Lesion sensitivity	45.45 (22.72–54.55)	45.45 (27.27–63.64)	94.0	0.64
ROC	0.755 (0.687–0.842)	0.753 (0.641–0.821)	77.5	0.75
Sensitivity	70.0 (65.0–90.0)	85.0 (80.0–95.0)	131.0	0.04*
Specificity	65.0 (45.0–75.0)	40.0 (32.5–60.0)	39.5	0.02*

IQR, interquartile range; JAFROC, jackknife alternative free-response ROC; ROC, receiver operating characteristic.

Note: Sample size for radiology trainees is eight; for radiologists, it is 22 for sensitivity and specificity and 21 for JAFROC, lesion sensitivity, and ROC due to missing data for one radiologist for these metrics

* Significant at the.05 level (two-tailed).

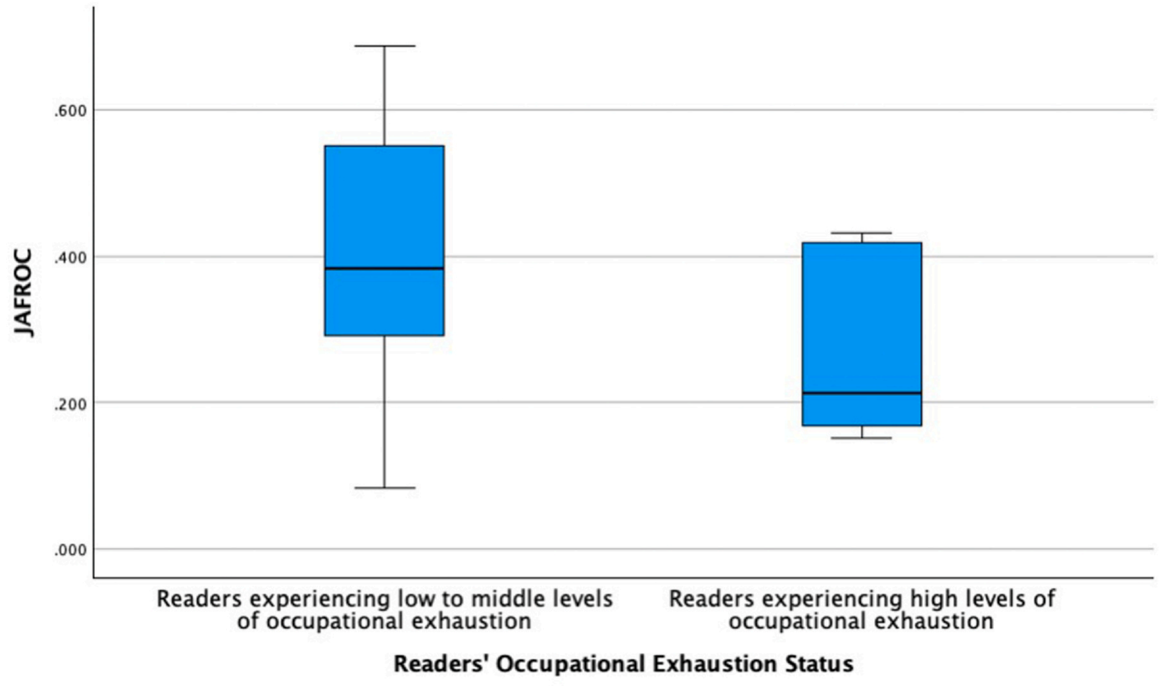

Figure 2. Impact of occupational exhaustion on JAFROC scores: box plot comparing low-to-middle (n = 21, median = 0.383, IQR = 0.287–0.552) and high (n = 8, median = 0.213, IQR = 0.163–0.422; p = 0.041) exhaustion groups.

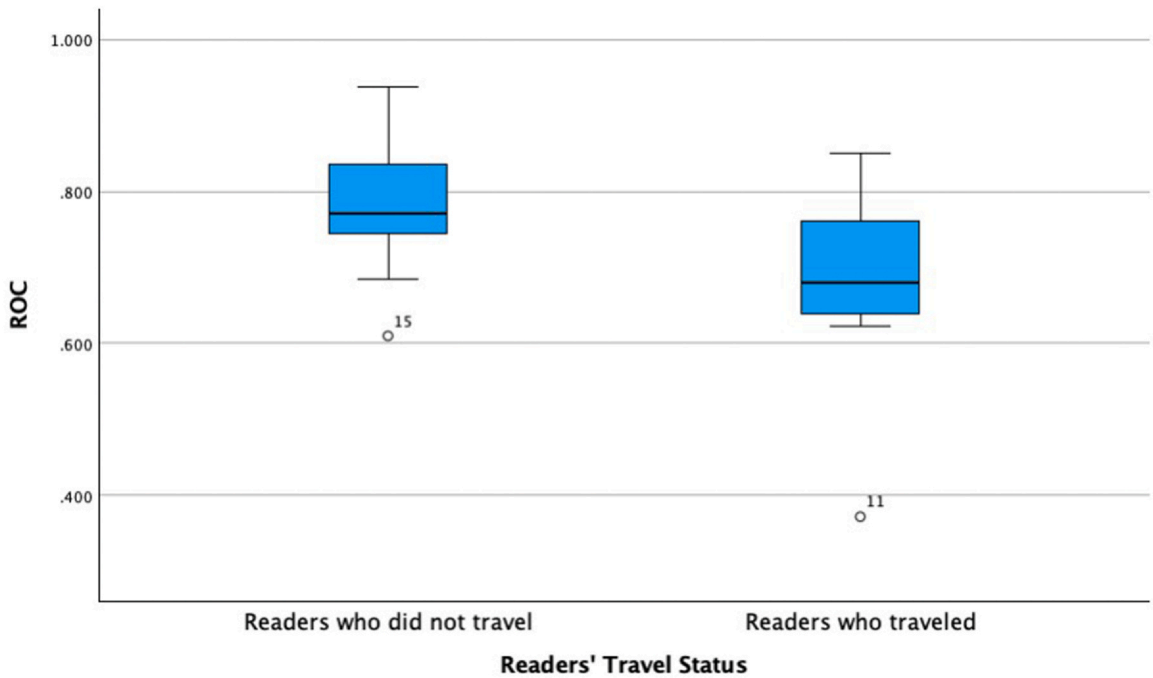

Figure 3. Effect of travel on ROC performance: box plot comparing non-travelers (*n* = 21, median = 0.772, IQR = 0.721–0.842) to Travelers (*n* = 8, median = 0.681, IQR = 0.631–0.785; p = 0.03).

therefore investigate individual travel parameters (e.g., duration, direction, time zones).

The participants' physiological hunger was also significant and was associated with higher sensitivity but lower specificity, suggesting a hunger-driven shift in readers' decision thresholds and hence a trade-off that did not impact overall accuracy. These effects are consistent with other fields in which decision-makers have displayed increased risk aversion when hungry (18,19,23,31,32), as described in the introduction. In radiology, this would manifest as a lower threshold for considering potential abnormalities as cancerous and a preference for recalling patients for further

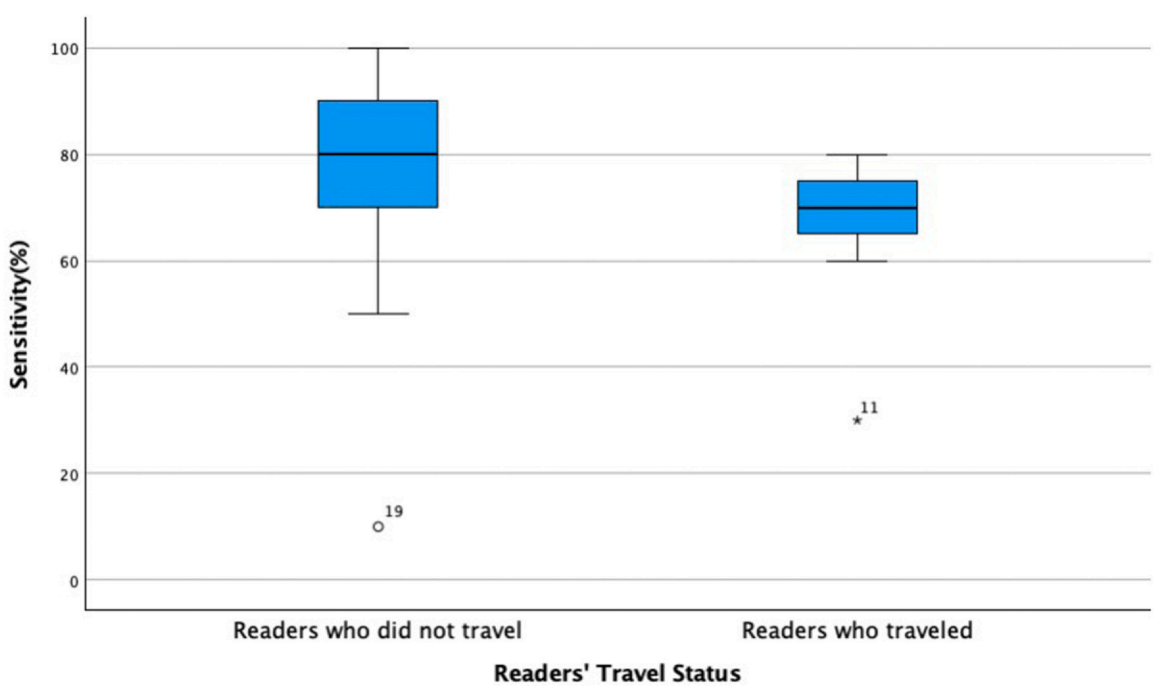

Figure 4. Impact of travel on sensitivity: box plot comparing non-travelers (*n* = 22, median = 80.0%, IQR = 70.0–90.0%) to Travelers (*n* = 8, median = 70.0%, IQR = 62.5–77.5%; p = 0.04).

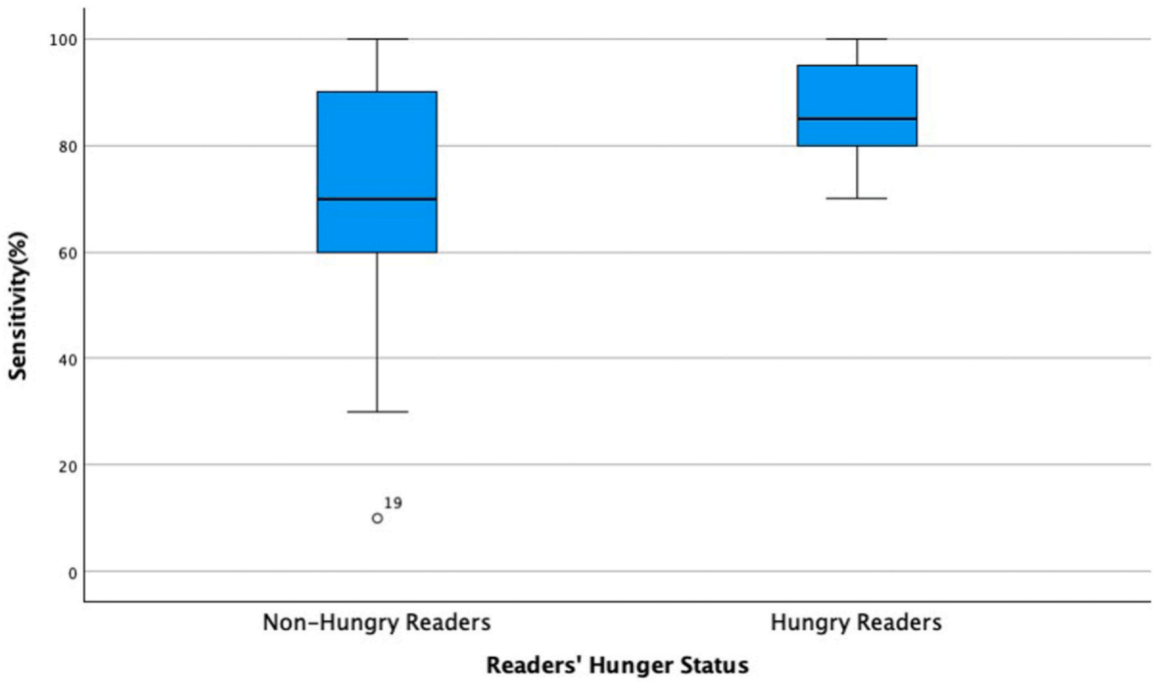

Figure 5. Effect of hunger on sensitivity: box plot comparing non-hungry (*n* = 22, median = 70.0%, IQR = 65.0–90.0%) to Hungry (*n* = 8, median = 85.0%, IQR = 80.0–95.0%; p = 0.04) Readers.

investigations to reduce negative consequences for the patient and the chance of malpractice litigation. However, although higher sensitivity while hungry could reduce false negatives—thus preventing delayed diagnoses and poorer patient outcomes—the concurrent decrease in specificity may increase false positives, leading to overdiagnosis, heightened patient anxiety, and additional healthcare costs from follow-up imaging and biopsies.

Overall, the findings highlight the importance of physiological factors in radiological practice and decision-making protocols. The demands of radiology work often disrupt mealtimes, which can be compounded by dietary trends like

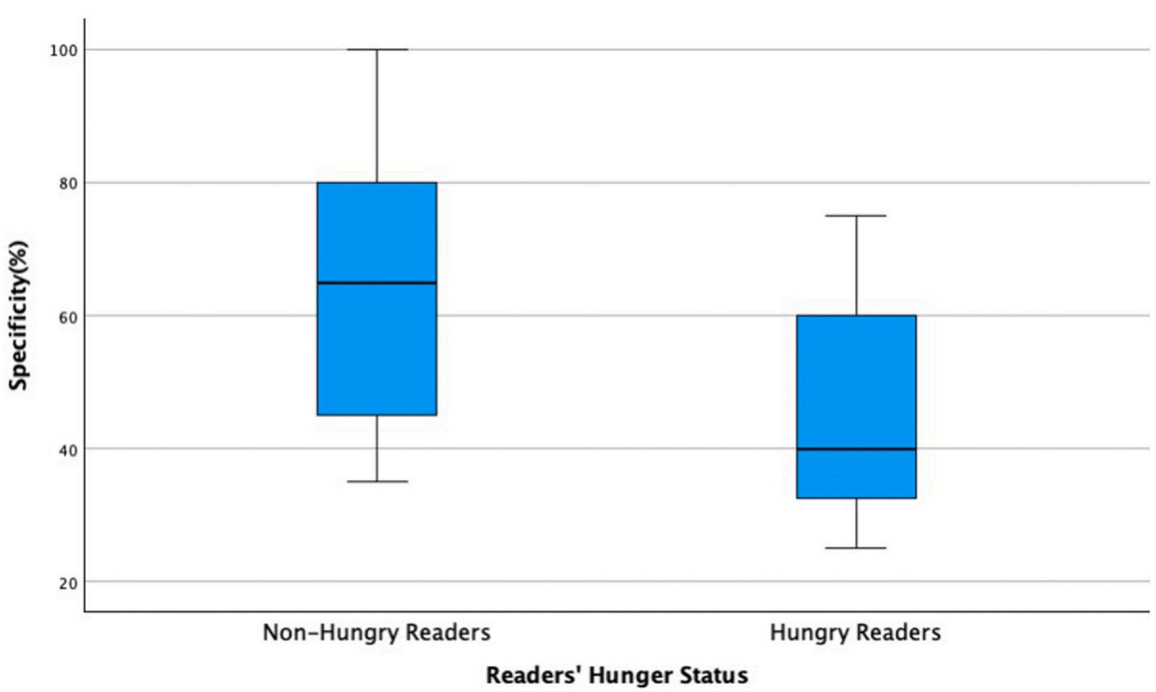

Figure 6. Impact of hunger on specificity: box plot comparing non-hungry (*n* = 22, median = 65.0%, IQR = 45.0–75.0%) to Hungry (*n* = 8, median = 40.0%, IQR = 32.5–60.0%; p = 0.02) Readers.

intermittent fasting (61). Future research could explore interventions to mitigate these effects, such as optimizing meal schedules or providing nutritional support during long reading sessions.

Interpreting mammograms involves multiple processes, including pattern recognition to identify potential abnormalities, detailed examinations of flagged areas, and deciding if abnormalities are present (1), and the differential effects of the stressors on performance metrics suggest they may selectively impact individual process. For example, hunger has previously been shown to influence decision-making and risk assessment (2), which are part of the final decision phase, in contrast to occupational exhaustion and travel, which may impact visual search and attention (22,24). Future research using eye tracking, neuroimaging, and cognitive testing (62–64) could better elucidate these distinct pathways and inform targeted interventions. The findings also suggest that other, seemingly mundane physiological factors may have unexpected impacts on clinical decision-making; we should therefore consider the effects of such stressors, explore innovative solutions, and seek insights from other high-stakes decision-making professions.

By identifying the impact of occupational exhaustion, hunger, and travel on diagnostic accuracy, we have highlighted the importance of a more holistic approach to radiology practice and the need for a paradigm shift in how radiology departments and healthcare systems manage working conditions and environments for their staff. By researching, acknowledging, and addressing human factors, we can enhance not only the quality of diagnoses but also the sustainability of the radiology profession in the face of growing challenges, such as burnout and cognitive fatigue.

LIMITATIONS AND RECOMMENDATIONS

This study was conducted in a controlled environment that differed from typical clinical settings, particularly in the lack of real-world variables present in daily clinical work, such as interruptions, established workflows, and typical workplace stressors. The prevalence of cancer cases in our sample was also higher than in routine screening, where cancer detection rates are typically 4–5 per 1000 examinations, which may have influenced reading patterns and the readers' thresholds for recall. Although previous research has found strong correlations between laboratory and clinical readings (65), the differences in cancer prevalence and reading environment require careful consideration when generalizing the findings to clinical practice.

The structured two-hour interpretation sessions did not replicate a full clinical workday but did enable us to isolate and examine the impacts of occupational exhaustion, travel, and hunger independently of the fatigue of extended shifts (15). However, the study design did not control for several potential confounding variables related to fatigue, including participants' total time awake before reading, the quality and duration of their sleep, and their circadian phase.

While conducting this study during conferences provided access to a large group of radiologists, conference attendees may differ systematically from the broader radiological workforce in terms of academic engagement, institutional support for professional development, and motivation for performance improvement. The conference setting is also distinct from routine clinical practice, with varied session scheduling, different workstation setups, and distractions from conference activities. Motivation and performance awareness, in the form of heightened attention and professional pride, might also have influenced reading behavior.

Because the conferences were held in Saudi Arabia, some confounding factors common at international conferences, such as alcohol consumption, were not present, but others, like unfamiliar hotel accommodations and conference commitments, were. Future studies should therefore examine settings that more closely mirror real-world clinical conditions to validate and expand upon these findings.

The sample size warrants careful consideration of statistical power and generalizability; while comparable to similar studies in radiological research, the relatively small sample and uneven distribution across conditions (Table 3) may have prevented the detection of subtle effects, which were suggested by the near-significant impacts of occupational exhaustion on specificity and of travel on lesion sensitivity. The inconsistent representation across training levels may also have prevented us from distinguishing experience-related effects from those directly associated with occupational exhaustion, hunger, or travel—indeed, the analysis of experience as a confounding variable yielded no significant effects, contradicting existing literature. This unexpected finding might also be attributable to methodological factors, such as variations in annual reading volumes and non-local test cases that eliminated institutional familiarity advantages. That is, while non-local cases enhanced internal validity by controlling for institutional familiarity bias and created a more controlled environment to isolate the variables of interest—which was a strength—they may also have reduced the performance differential between experienced and less experienced radiologists by removing the context in which experienced radiologists' pattern recognition would typically manifest, thus representing a limitation.

Mann–Whitney U tests were chosen for their ability to handle small and uneven sample sizes, and they identified significant effects across multiple metrics (47–51) with effect sizes at the low end of the range conventionally considered large (Exhaustion JAFROC: Cohen's d = 0.82; Travel ROC: d = 0.87; Travel sensitivity: d = 0.81; Hunger sensitivity: d = 0.81; Hunger specificity: d = 0.92). This suggests that the sample size was sufficient to capture large effects but that medium or small effects might have been subject to Type II errors (66,67). Future research should employ larger, balanced cohorts with controlled distributions of stressors across experience levels to enhance statistical power and generalizability. Randomized study designs in controlled settings would also allow for more precise examination of these relationships and a stronger assessment of causality. Such studies, with larger, balanced samples, could also take advantage of more sophisticated statistical approaches than were justifiable in the current study, such as mixed-effects models, which could account for both fixed and random effects and handle repeated measures and nested data structures.

Although the MBI-GS Emotional Exhaustion subscale is a well-validated tool, the travel and hunger questions were binary self-reports without prior validation, introducing concerns about reliability, subjectivity, and recall bias. Future studies should therefore consider standardized and validated tools for assessing physiological factors; they could also incorporate a broader range of physiological and psychological measures, such as blood glucose levels, body temperature, fatigue scales, sleep patterns, visual function tests, mood assessments, and concentration metrics; they could even correlate functional neuroimaging with occupational exhaustion and performance. It would also be valuable to examine how these factors interact with age, gender, and experience, which are known to influence susceptibility to fatigue and performance fluctuations (16,68). Long-term studies tracking radiologists through various work schedules and conditions could provide even more comprehensive insights into the impact of human factors on diagnostic accuracy.

Additionally, investigating interventions to mitigate the effects of exhaustion, hunger, and travel on performance would be beneficial for developing evidence-based radiology workplace policies. A significant advance could be the integration of AI technology, a major strength of which is its ability to produce near-instantaneous interpretations of radiographs, neither constrained by traditional working hours nor susceptible to human fallibilities, such as fatigue or hunger. Incorporating AI could complement radiologists' work by providing consistent interpretations with predictable strengths and weaknesses that would help to counterbalance the natural fluctuations in human performance caused by physiological and psychological factors. However, potential over-reliance on AI systems and the risk of training biases require careful consideration (69). Furthermore, major challenges remain in validating AI systems across diverse patient populations and imaging equipment and in developing strategies to mitigate algorithmic gender, ethnicity, and socioeconomic biases. Future implementations should focus on robust validation processes and appropriate human oversight to ensure AI integration enhances rather than compromises radiological care.

CONCLUSIONS

This study found that occupational exhaustion, travel, and hunger significantly impact mammographic interpretation. Key findings include decreased JAFROC scores with occupational exhaustion, reduced overall diagnostic performance (ROC and sensitivity) following travel, and a sensitivity–specificity trade-off when hungry, revealing the distinct impacts of physiological and psychological states on different aspects of diagnostic performance. The evidence suggests the possible need for new workplace policies and interventions for radiology departments, such as structured recovery periods after travel, optimized work schedules to manage exhaustion, and regular meal breaks, although validation of the findings in real-world clinical settings is needed to confirm the practical value of such measures. Future research with larger samples should focus on developing targeted interventions, integrating AI technologies to complement human expertise, and creating more supportive work environments.

FUNDING INFORMATION

None.

ALL AUTHORS CONTRIBUTED TO THE FOLLOWING

- Conception or design of the study
- Drafting and/or revision of the manuscript.

All authors have read and approved the final version of the manuscript and are solely responsible for its content.

STATEMENT OF DATA ACCESS AND INTEGRITY

The authors declare that they had full access to all of the data in this study and take complete responsibility for the integrity of the data and the accuracy of the data analysis.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. Garland LH. On the scientific evaluation of diagnostic procedures. Radiology 1949; 52(3):309–328.
2. Degnan AJ, Ghobadi EH, Hardy P, et al. Perceptual and interpretive error in diagnostic radiology-causes and potential solutions. Acad Radiol 2019; 26(6):833–845.
3. Garland LH, Miller ER, Zwerling HB, et al. Studies on the value of serial films in estimating the progress of pulmonary disease. Radiology 1952; 58(2):161–177.
4. Berlin L. Radiologic errors, past, present and future. Diagnosis 2014; 1(1):79–84.
5. Waite S, Scott J, Gale B, et al. Interpretive error in radiology. AJR Am J Roentgenol 2017; 208(4):739–749.
6. Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. RadioGraphics 2015; 35(6):1668–1676.

7. Nodine CF, Kundel HL. Using eye movements to study visual search and to improve tumor detection. Radiographics 1987; 7(6):1241–1250.
8. Kundel HL, Nodine CF, Carmody D. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol 1978; 13(3):175–181.
9. Kundel HL, Nodine CF, Conant EF, et al. Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology 2007; 242(2):396–402.
10. Elmore JG, Wells CK, Lee CH, et al. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994; 331(22):1493–1499.
11. A.S.A. Alshabibi. Factors Causing Variability When Reading Mammograms. 2022.
12. Clerkin N, Ski CF, Brennan PC, et al. Identification of factors associated with diagnostic performance variation in reporting of mammograms: a review. Radiography 2023; 29(2):340–346.
13. Hoff SR, Myklebust T-Å, Lee CI, et al. Influence of mammography volume on radiologists' performance: results from BreastScreen Norway. Radiology 2019; 292(2):289–296.
14. Nodine CF, Kundel HL, Mello-Thoms C, et al. How experience and training influence mammography expertise. Acad Radiol 1999; 6(10):575–585.
15. Krupinski E, Berbaum K, Caldwell R, et al. Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol JACR 2010; 7:698–704.
16. Taylor-Phillips S, Stinton C. Fatigue in radiology: a fertile area for future research. Br J Radiol 2019; 92(1099):20190043.
17. Alshabibi AS, Mo'ayyad ES, Tapia KA, et al. Impact of hours awake and hours slept at night on radiologists' mammogram interpretations. J Am Coll Radiol 2021; 18(5):730–738.
18. Alshabibi AS, Suleiman ME, Tapia KA, et al. Impact of time of day on radiology image interpretations. Clin Radiol 2020; 75(10):746–756.
19. Becker AS, Woo S, Leithner D, et al. The "Hungry Judge" effect on prostate MRI reporting: chronobiological trends from 35′004 radiologist interpretations. Eur J Radiol 2024; 179:111665.
20. Trieu PD, Puslednik L, Colley B, et al. Interpretative characteristics and case features associated with the performances of radiologists in reading mammograms: a study from a non-screening population in Asia. Asia-Pacific J Clin Oncol 2021; 17(1):139–148.
21. Applying a social network analysis (SNA) approach to understanding radiologists' performance in reading mammograms. In: Taba ST, Hossain L, Heard R, Brennan P, Lee W, Lewis S, editors. Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment. SPIE; 2017.
22. Lemonaki R, Xanthopoulou D, Bardos AN, et al. Burnout and job performance: a two-wave study on the mediating role of employee cognitive functioning. Eur J Work Organ Psychol 2021; 30(5):692–704.
23. Danziger S, Levav J, Avnaim-Pesso L. Extraneous factors in judicial decisions. Proc Natl Acad Sci 2011; 108(17):6889–6892.
24. Leatherwood W, Dragoo J. Effect of airline travel on performance: a review of the literature. Br J Sports Med 2013; 47(9):561–567.
25. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. Ann Surg 2010; 251(6):995–1000.
26. Tawfik DS, Profit J, Morgenthaler TI, et al. Physician burnout, well-being, and work unit safety grades in relationship to reported medical errors. Mayo Clin Proc 2018; 93(11):1571–1580.
27. Edú-Valsania S, Laguía A, Moriano JA. Burnout: a review of theory and measurement. Int J Environ Res Public Health 2022; 19(3):1780.
28. Soroush F, Zargham-Boroujeni A, Namnabati M. The relationship between nurses' clinical competence and burnout in neonatal intensive care units. Iran J Nurs Midwifery Res 2016; 21(4):424–429.
29. Bes I, Shoman Y, Al-Gobari M, et al. Organizational interventions and occupational burnout: a meta-analysis with focus on exhaustion. Int Arch Occup Environ Health 2023; 96(9):1211–1223.
30. Rossiter A, Warrington G, Comyns T. Effects of long-haul travel on recovery and performance in elite athletes: a systematic review. J Strength Cond Res 2021; 36(11):3234–3245.
31. Singh G, Launer S. Time of day and hearing aid adoption. Trends Hear 2018; 22:2331216518769789.
32. De Ridder D, Kroese F, Adriaanse M, et al. Always gamble on an empty stomach: Hunger is associated with advantageous decision making. PloS One 2014; 9(10):e111081.
33. Loewenstein G. Hot-cold empathy gaps and medical decision making. Health Psychol 2005; 24(4s):S49–S56.
34. Baumeister RF, Vohs KD. Self-regulation, ego depletion, and motivation. Soc Personality Psychol Compass 2007; 1(1):115–128.
35. Baumeister RF, Bratslavsky E, Muraven M, et al. Ego depletion: is the active self a limited resource? J Pers Soc Psychol 1998; 74(5):1252–1265.
36. Gailliot M. Hunger and reduced self-control in the laboratory and across the world: reducing hunger as a self-control panacea. Psychology 2013; 04:59–66.
37. Gailliot MT, Baumeister RF, DeWall CN, et al. Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. J Pers Soc Psychol 2007; 92(2):325–336.
38. Orquin JL, Kurzban R. A meta-analysis of blood glucose effects on human decision making. Psychol Bull 2016; 142(5):546–567.
39. Korbonits M, Goldstone AP, Gueorguiev M, et al. Ghrelin—a hormone with multiple functions. Front Neuroendocrinol 2004; 25(1):27–68.
40. Sarker MR, Franks S, Caffrey J. Direction of post-prandial ghrelin response associated with cortisol response, perceived stress and anxiety, and self-reported coping and hunger in obese women. Behav Brain Res 2013; 257:197–200.
41. Brown H, Proulx MJ, Stanton Fraser D. Hunger bias or gut instinct? Responses to judgments of harm depending on visceral state versus intuitive decision-making. Front Psychol 2020; 11:2261.
42. Currie PJ, Khelemsky R, Rigsbee EM, et al. Ghrelin is an orexigenic peptide and elicits anxiety-like behaviors following administration into discrete regions of the hypothalamus. Behav Brain Res 2012; 226(1):96–105.
43. Kluge M, Schüssler P, Zuber V, et al. Ghrelin enhances the nocturnal secretion of cortisol and growth hormone in young females without influencing sleep. Psychoneuroendocrinology 2007; 32(8-10):1079–1085.
44. Suleiman ME, Rickard M, Brennan PC. Perfecting detection through education. Radiography 2020; 26(2):S49–s53.
45. Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behav 1981; 2(2):99–113.
46. Maslach C., Jackson S., Leiter M. The Maslach Burnout Inventory Manual. 31997. p. 191–218.
47. MacFarland T, Yates J. Mann–Whitney U Test 2016:103–132.
48. Bürkner P.-C., Doebler P., Holling H. Optimal design of the Wilcoxon-Mann-Whitney-test 2022.
49. Tai KY, Dhaliwal J, Rmt Balasubramaniam V. Leveraging Mann–Whitney U test on large-scale genetic variation data for analysing malaria genetic markers. Malar J 2022; 21:79.
50. Dehaene H, De Neve J, Rosseel Y. A Wilcoxon–Mann–Whitney test for latent variables. Front Psychol 2021; 12:754898.
51. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1947; 18(1):50–60.
52. Deng F. The meaning of personal accomplishment and burnout in radiology. Am J Roentgenol 2018; 210(1):W41.
53. Fawzy N, Tahir MJ, Saeed A, et al. Incidence and factors associated with burnout in radiologists: a systematic review. Eur J Radiol Open 2023; 11:100530.
54. Parikh JR, Sun J, Mainiero MB. Prevalence of burnout in breast imaging radiologists. J Breast Imaging 2020; 2(2):112–118.
55. Parikh JR, Baird GL, Mainiero MB. A pre-post study of stressors and burnout affecting breast radiologists before and during the COVID-19 pandemic. Eur J Radiol Open 2023; 11:100507.
56. Parikh JR, Sun J, Mainiero MB. What causes the most stress in breast radiology practice? A survey of members of the society of breast imaging. J Breast Imaging 2021; 3(3):332–342.
57. Chetlen AL, Chan TL, Ballard DH, et al. Addressing burnout in radiologists. Acad Radiol 2019; 26(4):526–533.
58. Gransjøen AM. Prevalence, impact and prevention of stress and burnout, and promoting resilience among radiological personnel: a systematized review. Discov Psychol 2024; 4(1):11.

59. Alshabibi AS, Suleiman MaE, Tapia KA, et al. Impact of hours awake and hours slept at night on radiologists' mammogram interpretations. J Am Coll Radiol 2021; 18(5):730–738.
60. Bruno MA. Radiology errors across the diurnal cycle. Radiology 2020; 297(2):380–381.
61. Janaswamy R, Yelne P. A narrative review on intermittent fasting as an approachable measure for weight reduction and obesity management. Cureus 2022; 14(10):e30372.
62. Krupinski EA. Visual scanning patterns of radiologists searching mammograms. Acad Radiol 1996; 3(2):137–144.
63. van der Gijp A, Ravesloot CJ, Jarodzka H, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theory Pract 2017; 22(3):765–787.
64. Kaichi Y, Okada G, Takamura M, et al. Effect of fatigue on the brain activity of radiologists while reading images. Hiroshima J Med Sci 2023; 72(3-4):33–39.
65. Soh BP, Lee W, McEntee MF, et al. Screening mammography: test set data can reasonably describe actual clinical reporting. Radiology 2013; 268(1):46–53.
66. Cohen J. Statistical power analysis for the behavioral sciences. L. Erlbaum Associates,; 1988.
67. Cohen J. A power primer. Psychol Bull 1992; 112(1):155–159.
68. Wylie GR, Pra Sisto AJ, Genova HM, et al. Fatigue across the lifespan in men and women: state vs. trait. Front Hum Neurosci 2022; 16:790006.
69. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics 2023; 13(17):2760.