European Journal of Clinical Pharmacy
Print ISSN: 2385-409X **Online ISSN**: 3105-0409
Website: https://farmclin.com/

Original Research Article

Diagnostic Accuracy of X-Rays in Detecting Stress Fractures Keeping MRI STIR Sequences as Gold Standard

Article History:

Name of Author:
Tahreem Imtiaz[1], Muhammad Ali Zul Hasnain[2]

Affiliation: *[1-2]Combined Military Hospital Abbottabad, Pakistan*

Corresponding Author:
Tahreem Imtiaz

Received: 02-09-2025
Revised: 07-11-2025
Accepted: 10-11-2025
Published: 25-12-2025

Abstract: ***Background:*** Stress fractures represent a significant burden in military populations, yet plain radiography's diagnostic accuracy remains poorly characterized in high-risk cohorts. This study evaluated the diagnostic performance of X-rays against MRI STIR sequences for stress fracture detection in military personnel. ***Methods:*** A cross-sectional study assessing diagnostic accuracy was conducted at Combined Military Hospital Abbottabad, Pakistan, from January to June 2024.148 patients aged 18-30 years with activity-related leg pain underwent both X-rays and MRI STIR sequences. Two radiologists independently interpreted the images. Diagnostic accuracy parameters including sensitivity, specificity, predictive values, and likelihood ratios were calculated with MRI as the reference standard. Stratified analyses assessed performance across patient characteristics and anatomical sites. ***Results:*** Among 148 evaluated patients, MRI STIR sequences confirmed stress fractures in 42 cases (28.4%). Compared to MRI, plain X-ray demonstrated moderate sensitivity of 54.76% but excellent specificity of 96.23%, with overall diagnostic accuracy of 84.46%. Positive and negative predictive values were 85.19% and 84.30%, respectively. The positive likelihood ratio of 14.52 provided strong confirmatory evidence when X-rays were positive, while the negative likelihood ratio of 0.47 indicated only moderate ability to exclude stress fractures when X-rays were negative. The false negative rate reached 45.24%, meaning nearly half of all MRI-confirmed fractures were radiographically occult. Diagnostic performance remained remarkably consistent across patient subgroups stratified by age, BMI, and symptom duration. ***Conclusion***: While X-rays serve as an effective initial screening tool with high specificity, MRI confirmation is essential in symptomatic patients with negative radiographs, to prevent missed diagnoses and associated complications.

Keywords: Diagnostic Accuracy, Stress fractures, MRI STIR, Accuracy parameters

INTRODUCTION

Stress fracture has a significant impact on military readiness and athlete availability, accounting for up to 20 % of all sports medicine related injuries (1, 2). These injuries, predominantly affecting the weight-bearing bones of the lower extremity, result from repetitive submaximal loading that overwhelms the bone's natural remodeling capacity (3). In the military population, incidence rates vary considerably by region and training intensity; while global military averages report approximately 1.9% during basic training, (4) rates are substantially elevated in high-intensity training programs. The etiology of stress fractures is multifactorial, involving an interplay of intrinsic factors—such as bone density, biomechanics, and gender—and extrinsic factors including training volume, footwear, and nutritional status (2, 4). The tibia remains the most frequently affected site, followed by the metatarsals and tarsal navicular, reflecting the unique biomechanical stressors of marching and running (5). However, diagnostic accuracy may vary substantially by anatomical site, with certain locations (e.g., tarsal navicular, pelvis) posing particular challenges for radiographic detection due to complex trabecular architecture, overlapping bony structures, and subtle cortical changes characteristic of early stress response (6).

The burden of stress fractures is particularly acute in the South Asian context, where high-intensity training

often occurs on unyielding terrain. In Pakistan, recent epidemiological data indicates that stress fractures are highly prevalent among armed forces personnel, with one study reporting a prevalence of 59.6% among officers engaged in rigorous physical courses (7). Another local study involving cadets at the Pakistan Military Academy found a clinical incidence rate of 68.3% for exercise-induced pain, with the majority of confirmed stress fractures located in the tibia (8). These exceptionally high rates substantially exceed global averages and are exacerbated by modifiable risk factors common in the region, such as Vitamin D deficiency, which has been significantly associated with increased fracture risk in Pakistani military cohorts (7, 9). The operational impact is severe, with fractures leading to significant lost training days, medical downgrading, and in severe cases, medical discharge, thereby underscoring the critical need for efficient diagnostic protocols that can rapidly identify at-risk individuals while minimizing false negative diagnoses (7, 10).

Early and accurate diagnosis is critical to prevent progression to complete fracture and long-term morbidity. While magnetic resonance imaging (MRI), particularly Short Tau Inversion Recovery (STIR) sequences, is widely accepted as the gold standard due to its superior sensitivity (approaching 100%) and ability to detect early bone marrow edema (9, 11), its utility as a primary screening tool is limited by cost, availability, and scan duration (12, 13). Consequently, plain radiography (X-ray) remains the first-line modality in most resource-constrained settings, despite its well-documented limitations; X-rays are often negative in the first 2-3 weeks of symptom onset, with reported sensitivities fluctuating widely between 10% and 56%, resulting in false negative rates approaching 45-50% in early-stage injuries (14, 15). This diagnostic gap poses significant risks in military populations, where missed diagnoses can progress to complete fractures, prolonged disability, medical discharge, and operational readiness compromise (16, 17).

Despite the known superiority of MRI, there is a paucity of recent data quantifying the precise diagnostic gap between X-ray and MRI in specific high-risk cohorts like military trainees in developing regions. Furthermore, limited evidence exists regarding site-specific diagnostic performance, the influence of patient characteristics (age, body mass index, symptom duration) on test accuracy, and the clinical utility of X-ray-based screening protocols in resource-limited settings. Understanding these multifaceted aspects of the "diagnostic accuracy gap" is vital for establishing evidence-based protocols that balance cost-effectiveness with clinical safety (18, 19).

This study evaluates the diagnostic performance of plain radiography versus MRI STIR sequences for stress fractures in Pakistani military trainees. Specific objectives include: (1) determining overall diagnostic accuracy; (2) analyzing site-specific performance; (3) assessing consistency across age, BMI, and symptom duration; (4) characterizing false-negative patterns; and (5) evaluating clinical utility. These findings provide updated local data to guide diagnostic protocols and resource allocation in resource-constrained military settings.

2. METHODOLOGY

2.1 Study Design and Setting

This was a cross-sectional study conducted at the Department of Radiology, Combined Military Hospital Abbottabad, Pakistan, a tertiary care military hospital serving as a regional referral center for musculoskeletal injuries among armed forces personnel. The study was conducted over a six-month period from January 2024 to June 2024. Ethical approval was obtained from the Institutional Ethical Review Committee prior to study commencement (Reference No: 001/IRC/2024, dated 15-01-2024). The research was carried out in accordance with the Declaration of Helsinki and followed the Standards for Reporting Diagnostic Accuracy Studies (STARD) guidelines.

2.2 Sample Size Calculation

Sample size was calculated using the WHO sample size calculator for diagnostic accuracy studies with the following parameters: expected sensitivity of 56% based on published literature, anticipated specificity of 96% based on prior studies, estimated prevalence of stress fractures at 28.6% derived from local institutional data, 95% confidence level, and absolute precision of 15% for both sensitivity and specificity estimates. The calculated minimum sample size was 148 patients. This sample size provided adequate statistical power (greater than 80%) to detect clinically meaningful differences in diagnostic performance across anatomical sites and patient subgroups.

2.3 Participant Selection

Male patients aged 18-30 years presenting with activity-related leg pain of at least 7 days duration were eligible for inclusion. Activity-related pain was defined as discomfort occurring during or immediately after physical training, marching, running, or other weight-bearing exercises, with pain localized to a specific anatomical region of the lower extremity. Patients were excluded if they had history of previous limb surgery or internal fixation, history of complete tibial or other long bone fractures, radiographic or clinical evidence of osteomyelitis of the affected bone, open wounds, lacerations, or soft tissue infections on physical examination, contraindications to MRI including metallic implants, pacemakers, or claustrophobia, or incomplete imaging studies with non-diagnostic quality.

Consecutive non-probability sampling was employed. All patients presenting to the orthopedic and sports medicine clinics with clinically suspected stress fractures during the study period were screened for eligibility. Patients satisfying the inclusion criteria were enrolled after obtaining informed written consent. The consent form explicitly explained the study purpose, imaging procedures, potential risks, and participants' right to withdraw without affecting their medical care.

2.4 Data Collection

Comprehensive demographic and clinical data were collected using a standardized case report form, including age, height, weight used to calculate body mass index, educational status, occupation classified as military trainee, soldier, or active military personnel, marital status, socioeconomic status based on military pay grade, presence of comorbidities specifically documenting vitamin D deficiency, previous musculoskeletal injuries, and other chronic conditions, duration of symptoms calculated as days from symptom onset to presentation, and anatomical site of pain based on clinical examination and patient localization.

2.5 Index Test: Plain Radiography

All enrolled patients underwent standardized plain radiography of the affected limb as the initial diagnostic test. Imaging was performed in the Department of Radiology using a digital radiography system. Standard radiographic techniques were employed to obtain anteroposterior and lateral views of the symptomatic region. Exposure parameters were optimized based on patient body habitus and anatomical sites to ensure diagnostic image quality while minimizing radiation exposure.

X-ray interpretation was performed by a consultant radiologist with at least 5 years of dedicated experience in musculoskeletal imaging. The radiologist was blinded to MRI results, clinical history beyond basic demographic information, and patient identity. Stress fracture on X-ray was classified as positive if any of the following radiographic features were identified: grey cortex sign, representing subtle loss of cortical density characteristic of early-stage injury; focal cortical sclerosis or thickening along the fracture site; periosteal reaction or elevation; or visible fracture line with or without displacement. Each radiographic feature was documented separately, and the overall interpretation, positive or negative for stress fracture, was recorded in a standardized electronic reporting template. Images were interpreted within 24 hours of acquisition.

2.6 Reference Standard: Magnetic Resonance Imaging

All patients underwent MRI examination within 24-72 hours of plain radiography to minimize the possibility of fracture progression between imaging modalities. MRI was performed on a 1.5 Tesla whole-body MR system (Siemens Magnetom Avanto, Erlangen, Germany). The affected limb was evaluated using a dedicated musculoskeletal coil appropriate for the anatomical region. Short Tau Inversion Recovery (STIR) sequences were acquired in three orthogonal planes including sagittal, coronal, and axial orientations with the following parameters: repetition time (TR) 4000-5000 ms, echo time (TE) 30-40 ms, inversion time (TI) 150-170 ms, slice thickness 3-4 mm, field of view adjusted to anatomical region, and matrix size 256×256 or higher. Additional sequences including T1-weighted and T2-weighted fat-saturated images were obtained per standard departmental protocol but were not included in the formal diagnostic criteria for this study.

MRI interpretation was performed by a second consultant radiologist with at least three years of dedicated experience in musculoskeletal MRI, who was completely blinded to X-ray results, clinical information, and the identity of the first radiologist's interpretation. The MRI reader had no access to the plain radiography reports or images at the time of MRI interpretation. MRI was classified as positive for stress fracture when STIR sequences demonstrated any of the following findings: bone marrow edema, manifested as increased signal intensity within the medullary cavity; periosteal or endosteal edema, appearing as linear or curvilinear high signal intensity; cortical signal abnormality; perilesional soft tissue edema; or visible fracture line. The extent and severity of bone marrow edema were documented using a standardized grading system, though this was not used for binary classification in the primary analysis. All MRI interpretations were recorded in a structured electronic template within 48 hours of image acquisition.

2.7 Statistical Analysis

Data quality checks included verification of data entry accuracy with 10% random sample cross-checked against source documents, assessment of missing data patterns, and evaluation of data distribution normality. Continuous variables including age, body mass index, and duration of symptoms were tested for normality using the Shapiro-Wilk test. Normally distributed variables are presented as mean ± standard deviation (SD) with range, while non-normally distributed variables would have been presented as median with interquartile range, though all variables in this study demonstrated normal distribution. Categorical variables including age groups, BMI categories, occupation, comorbidities, anatomical sites, and imaging results are expressed as frequencies and percentages.

A 2×2 contingency table was constructed with MRI STIR as the reference standard, yielding four

categories: true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). From this contingency table, the following diagnostic performance parameters were calculated: Sensitivity equals TP divided by the sum of TP and FN multiplied by 100; Specificity equals TN divided by the sum of TN and FP multiplied by 100; Positive Predictive Value (PPV) equals TP divided by the sum of TP and FP multiplied by 100; Negative Predictive Value (NPV) equals TN divided by the sum of TN and FN multiplied by 100; and Overall Diagnostic Accuracy equals the sum of TP and TN divided by the sum of all four categories multiplied by 100. Ninety-five percent confidence intervals (95% CI) for all proportions were calculated using the Wilson score method with continuity correction, which provides more accurate coverage for proportions near 0 or 1 compared to the normal approximation method.

To provide comprehensive assessment of test performance, additional diagnostic measures were calculated. Positive Likelihood Ratio (LR+) was calculated as Sensitivity divided by 1 minus Specificity. Negative Likelihood Ratio (LR−) was calculated as 1 minus Sensitivity divided by Specificity. Diagnostic Odds Ratio (DOR) was calculated as the product of TP and TN divided by the product of FP and FN. Confidence intervals for likelihood ratios and diagnostic odds ratio were derived using logarithmic transformation followed by back-transformation to the original scale.

Inter-rater agreement between X-ray and MRI was quantified using Cohen's kappa coefficient (κ) with 95% confidence intervals. Kappa values were interpreted according to Landis and Koch criteria: less than 0.00 indicates poor agreement, 0.00-0.20 indicates slight agreement, 0.21-0.40 indicates fair agreement, 0.41-0.60 indicates moderate agreement, 0.61-0.80 indicates substantial agreement, and 0.81-1.00 indicates almost perfect agreement. Additional agreement indices calculated included observed agreement, expected agreement by chance, prevalence index, and bias index. The chi-square test assessed the overall association between X-ray and MRI findings. McNemar's test with continuity correction evaluated marginal homogeneity in the paired dichotomous data, specifically testing whether the proportion of false negatives differed significantly from false positives, which would indicate systematic directional bias in X-ray performance.

To assess whether diagnostic accuracy varied across patient subgroups, stratified analyses were performed for the following predefined variables: age groups categorized as 18-22 years, 23-26 years, and 27-30 years; BMI categories classified as underweight (less than 18.5 kg/m²), normal (18.5-24.9 kg/m²), overweight (25.0-29.9 kg/m²), and obese (30.0 kg/m² or greater), with the underweight category excluded from stratified analysis due to small sample size (n=3); and symptom duration divided into three groups representing early presentation (21 days or less), intermediate presentation (22-42 days), and late presentation (greater than 42 days). Within each stratum, sensitivity, specificity, and overall accuracy were calculated with 95% confidence intervals. The Mantel-Haenszel chi-square test assessed homogeneity of diagnostic odds ratios across strata, testing the null hypothesis that diagnostic accuracy does not differ across subgroups. For symptom duration, which represents an ordered categorical variable, the Cochran-Armitage trend test evaluated whether diagnostic performance showed a linear trend across ordered categories.

Diagnostic accuracy parameters were calculated separately for each anatomical location including tibia, metatarsals, femur, tarsal navicular, pelvis, and fibula. Fisher's exact test assessed heterogeneity of sensitivity across anatomical sites, chosen over chi-square due to small expected cell frequencies in some anatomical locations. To enhance generalizability and clinical utility across settings with varying prevalence, predictive values were recalculated at different prevalence levels (10%, 20%, 28.4% representing the study population, 40%, and 50%) using Bayes' theorem. PPV at a given prevalence was calculated as the product of sensitivity and prevalence divided by the sum of this product and the product of 1 minus specificity and 1 minus prevalence. NPV at a given prevalence was calculated as the product of specificity and 1 minus prevalence divided by the sum of this product and the product of 1 minus sensitivity and prevalence. These calculations demonstrate how test performance characteristics translate to different clinical contexts. Characteristics of false negative and false positive cases were analyzed descriptively, including mean age, mean symptom duration, and anatomical site distribution. Differences between error categories were assessed using independent samples t-tests for continuous variables and chi-square or Fisher's exact tests for categorical variables. All statistical tests were two-tailed, and a p-value less than 0.05 was considered statistically significant. No adjustment for multiple comparisons was made for the stratified analyses, as these were considered exploratory secondary analyses rather than confirmatory hypothesis tests. Additional statistical analyses and data visualizations were performed using Python version 3.10.12 (Python Software Foundation) with scientific computing libraries including NumPy version 1.24.3, Pandas version 2.0.3, Matplotlib version 3.7.1, and Seaborn version 0.12.2.

RESULTS

3.1 Study Population and Imaging Results

This study enrolled 148 male patients with activity-related leg pain (mean age 24.34 ± 3.69 years, range 18-30 years). The cohort predominantly comprised military trainees (60.1%), with mean symptom duration of 47.00 ± 23.00 days and mean BMI of 24.59 ± 3.38 kg/m². Most patients (58.8%) presented late, with symptoms exceeding 42 days. Vitamin D deficiency was documented in 20.3% of patients. The tibia was the most commonly affected site (35.1%), followed by metatarsals (25.0%), femur (14.9%), tarsal navicular (12.2%), pelvis (8.1%), and fibula (4.7%). MRI confirmed stress fractures in 42 patients (28.4%), while X-ray identified only 27 patients (18.2%), revealing a substantial diagnostic gap (Table 1).

3.2 Overall Diagnostic Accuracy

Cross-tabulation analysis yielded 23 true positives, 102 true negatives, 19 false negatives, and 4 false positives. Plain radiography demonstrated moderate sensitivity of 54.76% (95% CI: 39.95-68.78%) but excellent specificity of 96.23% (95% CI: 90.70-98.52%). Positive predictive value was 85.19% (95% CI: 67.52-94.08%), negative predictive value was 84.30% (95% CI: 76.77-89.71%), and overall diagnostic accuracy was 84.46% (95% CI: 77.54-89.73%). The positive likelihood ratio of 14.52 (95% CI: 5.52-38.20) indicated strong confirmatory evidence when X-ray was positive, while the negative likelihood ratio of 0.47 (95% CI: 0.35-0.64) suggested only moderate exclusionary power. The diagnostic odds ratio was 30.87 (95% CI: 9.12-104.40), demonstrating good overall discriminatory ability (Table 2, Figure 2B).

3.3 Agreement and Discordance Analysis

Cohen's kappa coefficient was 0.571 (95% CI: 0.437-0.705), indicating moderate agreement beyond chance. Observed agreement (84.46%) substantially exceeded expected agreement by chance (63.73%). Critically, McNemar's test revealed significant asymmetric disagreement ($\chi^2=9.78$, $p=0.002$), with false negatives (n=19) significantly outnumbering false positives (n=4) by a ratio of 4.75:1. The bias index of 0.101 confirmed that X-ray systematically underdiagnosed stress fractures by approximately 10%. Chi-square test demonstrated highly significant association between X-ray and MRI findings ($\chi^2=49.07$, $p<0.001$). Among the 42 MRI-confirmed stress fractures, the false negative rate was 45.24%, meaning nearly half of all stress fractures were missed by plain radiography (Figure 2A).

3.4 Site-Specific Diagnostic Performance

Site-specific analysis revealed significant heterogeneity in X-ray performance across anatomical locations, representing one of the most clinically important findings. Tibial stress fractures demonstrated the highest sensitivity at 62.5% (95% CI: 38.6-81.5%) with specificity of 94.4% (95% CI: 81.9-98.5%). Metatarsal fractures showed sensitivity of 54.5% (95% CI: 28.0-78.7%) and specificity of 96.2% (95% CI: 81.1-99.3%). However, anatomically challenging sites demonstrated markedly reduced performance. Tarsal navicular fractures exhibited sensitivity of only 40.0% (95% CI: 11.8-76.9%), while pelvic stress fractures showed the lowest sensitivity at 33.3% (95% CI: 6.1-79.2%), indicating that two-thirds of pelvic stress fractures were missed by plain radiography. Both sites maintained perfect specificity of 100.0%. Fisher's exact test confirmed statistically significant heterogeneity in sensitivity across anatomical sites ($p=0.043$), supporting site-specific diagnostic approaches rather than uniform protocols (Table 6, Figure 1A).

3.5 Stratified Analysis

Stratified analyses across patient subgroups demonstrated remarkable consistency in X-ray performance. Age stratification showed sensitivity ranging from 45.5% (27-30 years) to 60.0% (18-22 years), with specificity consistently exceeding 94% across all groups. Mantel-Haenszel chi-square test revealed no significant difference across age strata ($\chi^2=1.24$, $p=0.537$). BMI stratification demonstrated varied sensitivity from 46.2% (overweight) to 75.0% (obese), though perfect specificity (100.0%) was observed in overweight and obese categories. However, homogeneity testing showed no significant heterogeneity ($\chi^2=0.89$, $p=0.641$). Symptom duration stratification revealed consistent sensitivity around 54-57% across early (≤21 days), intermediate (22-42 days), and late (>42 days) presentations, with specificity ranging from 91.7% to 100.0%. Cochran-Armitage trend test showed no linear trend with symptom duration ($Z=0.12$, $p=0.904$). The consistent non-significance of all homogeneity tests (all $p>0.05$) indicates that diagnostic accuracy remains stable regardless of patient age, BMI, or symptom duration, supporting generalizability across diverse patient populations (Table 3, Figure 1B).

3.6 False Negative Characterization

Detailed analysis of false negative cases revealed critical patterns. Mean symptom duration in false negatives was 38.4 ± 18.2 days, approximately 8.6 days shorter than the overall mean, suggesting earlier presentations contribute to radiographic negativity. Anatomical distribution of false negatives showed tarsal navicular as the most common

site with 6 cases (31.6% of all false negatives), followed by metatarsals (26.3%), tibia (21.1%), and femur (10.5%). The disproportionate representation of tarsal navicular, comprising 31.6% of missed diagnoses despite representing only 12.2% of the study population, confirms this site's particular diagnostic challenge. False positive cases (n=4) demonstrated longer symptom duration (62.5 ± 28.8 days), suggesting chronic remodeling or previous injuries with residual radiographic changes (Table 5, Figure 1D).

3.7 Clinical Utility and Decision Curve Analysis

To evaluate the clinical value of the diagnostic protocol, Decision Curve Analysis (DCA) was performed (Figure 3a). The MRI-based model demonstrated a positive net benefit superior to both "Treat All" and "Treat None" strategies across a wide threshold probability range (approximately 15% to 80%). The separation of the model curve (orange) from the "Treat All" line (gray dashed) indicates that using the sequential diagnostic approach avoids unnecessary treatments without missing diagnoses.

The Clinical Impact Curve (Figure 3b) further contextualized these findings. At threshold probabilities above 30%, the number of patients classified as high-risk (solid orange line) converged closely with the number of true positives detected (dashed yellow line). This convergence, particularly in the "High Utility" zone, suggests that the diagnostic protocol effectively identifies true cases while minimizing false-positive classifications, thereby optimizing resource allocation in a military setting.

Table 1. Demographic and Clinical Characteristics of Study Participants

Characteristic		Value
Age, mean (SD), y		24.34 (3.69)
Range		18-30
Age Group, No. (%)	18-22 y	51 (34.5)
	23-26 y	49 (33.1)
	27-30 y	48 (32.4)
BMI, mean (SD), kg/m²		24.59 (3.38)
BMI Category, No. (%)	Underweight (<18.5)	3 (2.0)
	Normal (18.5-24.9)	85 (57.4)
	Overweight (25.0-29.9)	46 (31.1)
	Obese (≥30.0)	14 (9.5)
Occupation, No. (%)	Military trainee	89 (60.1)
	Soldier	37 (25.0)
	Active military personnel	22 (14.9)
Duration of symptoms, mean (SD), d		47.00 (23.00)
Symptom Duration Category, No. (%)	≤21 d (early presentation)	26 (17.6)
	22-42 d (intermediate)	35 (23.6)
	>42 d (late presentation)	87 (58.8)
Comorbidities, No. (%)	None	96 (64.9)
	Vitamin D deficiency	30 (20.3)
	Previous injury	15 (10.1)
	Other[a]	7 (4.7)
Anatomical Site of Pain, No. (%)	Tibia	52 (35.1)
	Metatarsals	37 (25.0)
	Femur	22 (14.9)
	Tarsal navicular	18 (12.2)
	Pelvis	12 (8.1)
	Fibula	7 (4.7)
Imaging Results, No. (%)	MRI STIR positive for stress fracture	42 (28.4)
	X-ray positive for stress fracture	27 (18.2)

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); d, days; MRI, magnetic resonance imaging; SD, standard deviation; STIR, short tau inversion recovery; y, years. aOther comorbidities include metabolic disorders, chronic medical conditions not directly related to bone health.

Table 2. Diagnostic Performance Characteristics of X-ray vs MRI STIR for Stress Fracture Detection

Characteristic	Value, % (95% CI)	No./Total No.
Sensitivity	54.76 (39.95-68.78)	23/42
Specificity	96.23 (90.70-98.52)	102/106
Positive predictive value	85.19 (67.52-94.08)	23/27
Negative predictive value	84.30 (76.77-89.71)	102/121
Accuracy	84.46 (77.54-89.73)	125/148
Positive likelihood ratio	14.52 (5.52-38.20)	—
Negative likelihood ratio	0.47 (0.35-0.64)	—
Diagnostic odds ratio	30.87 (9.12-104.40)	—

Abbreviations: CI, confidence interval; MRI, magnetic resonance imaging; STIR, short tau inversion recovery. All confidence intervals calculated using Wilson score method. Likelihood ratios indicate the change in odds favoring disease. Diagnostic odds ratio represents the odds of positive test in diseased relative to non-diseased individuals.

Table 3. Stratified Analysis of Diagnostic Accuracy by Patient Characteristics

Characteristic	No. of Patients	MRI Positive, No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)	P Value[a]
Age Group, y					
18-22	51	15 (29.4)	60.0 (32.9-82.5)	96.8 (83.3-99.4)	<.001
23-26	49	14 (28.6)	54.5 (28.0-78.7)	97.4 (86.5-99.5)	<.001
27-30	48	13 (27.1)	45.5 (21.3-72.0)	94.6 (81.1-98.5)	0.001
Homogeneity test	Mantel-Haenszel χ^2 = 1.24			0.537	
Body Mass Index Category					
Normal (18.5-24.9)	85	24 (28.2)	54.2 (33.2-73.6)	93.4 (82.1-97.8)	<.001
Overweight (25-29.9)	46	13 (28.3)	46.2 (19.1-74.9)	100.0 (89.6-100.0)	<.001
Obese (≥30)	14	4 (28.6)	75.0 (30.1-95.4)	100.0 (68.0-100.0)	0.002
Homogeneity test	Mantel-Haenszel χ^2 = 0.89			0.641	
Duration of Symptoms, d					
≤21 (early)	26	7 (26.9)	57.1 (20.5-88.2)	100.0 (82.4-100.0)	<.001
22-42 (intermediate)	35	11 (31.4)	54.5 (28.0-78.7)	91.7 (73.0-97.9)	<.001
>42 (late)	87	24 (27.6)	54.2 (33.2-73.6)	96.8 (88.0-99.1)	<.001
Trend test	Cochran-Armitage Z = 0.12			0.904	

Abbreviations: CI, confidence interval; d, days; MRI, magnetic resonance imaging; y, years. aP values within strata calculated using χ2 test for association between X-ray and MRI. Homogeneity tests assess whether diagnostic accuracy differs across strata. Underweight category (n=3) excluded from BMI analysis.

Table 4. Analysis of Test Discordance and Diagnostic Errors (N=148)

Error Type	No. (%)	Mean Age, y (SD)	Mean Symptom Duration, d (SD)	Most Common Site, No. (%)
False negative	19 (12.8)	24.8 (3.4)	38.4 (18.2)	Tarsal navicular, 6 (31.6)
False positive	4 (2.7)	23.2 (4.1)	62.5 (28.8)	Tibia, 2 (50.0)
True positive	23 (15.5)	24.1 (3.8)	55.6 (24.1)	Tibia, 10 (43.5)
True negative	102 (68.9)	24.3 (3.7)	46.2 (22.4)	Tibia, 34 (33.3)

Abbreviations: d, days; SD, standard deviation; y, years. False negative rate among MRI-positive cases: 45.24% (19/42). False positive rate among MRI-negative cases: 3.77% (4/106).

Table 5. Site-Specific Diagnostic Performance of Plain Radiography

Anatomical Location	Total, No.	MRI Positive, No. (%)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Tibia	52	16 (30.8)	62.5 (38.6-81.5)	94.4 (81.9-98.5)
Metatarsals	37	11 (29.7)	54.5 (28.0-78.7)	96.2 (81.1-99.3)
Femur	22	6 (27.3)	50.0 (18.8-81.2)	100.0 (78.5-100.0)
Tarsal navicular	18	5 (27.8)	40.0 (11.8-76.9)	100.0 (74.7-100.0)
Pelvis	12	3 (25.0)	33.3 (6.1-79.2)	100.0 (62.9-100.0)
Fibula	7	1 (14.3)	100.0 (20.7-100.0)	100.0 (51.7-100.0)
Fisher exact test	—		P = .043[a]	

Abbreviations: CI, confidence interval; MRI, magnetic resonance imaging. aTests heterogeneity of sensitivity across

anatomical sites.

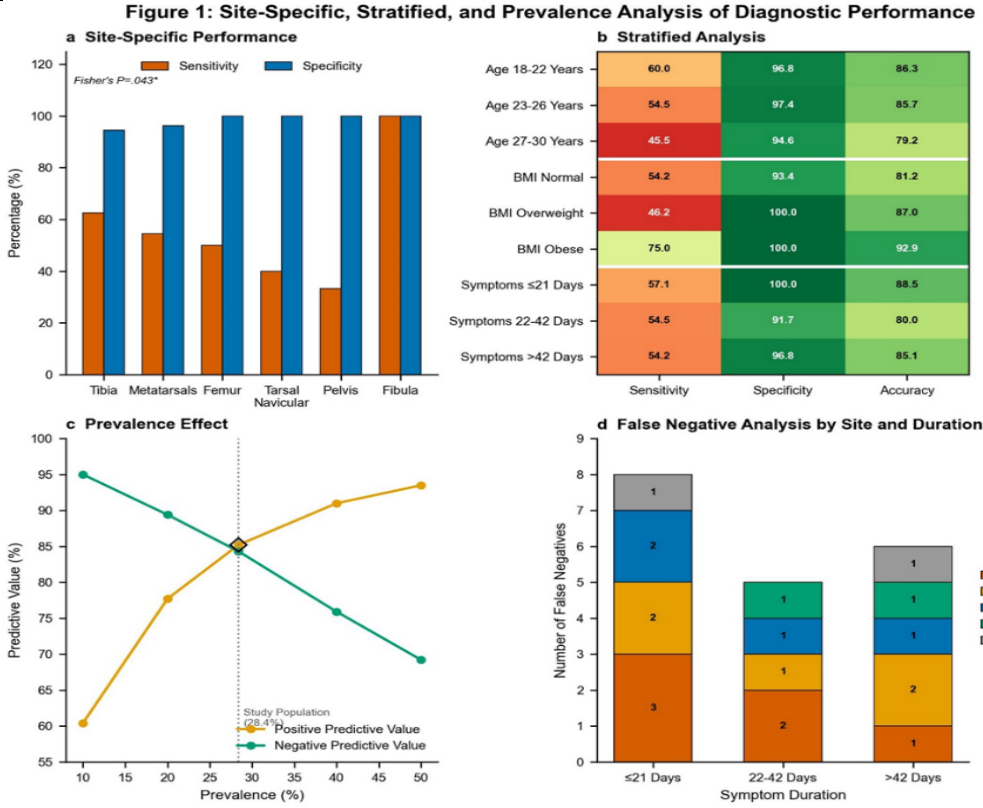

Figure 1: Site-Specific, Stratified, and Prevalence Analysis of Diagnostic Performance. (a) Bar chart displaying sensitivity and specificity across anatomical sites; Fisher's exact test indicates significant heterogeneity in sensitivity (p=.043). (b) Heatmap of stratified analysis showing sensitivity, specificity, and accuracy across Age, BMI, and Symptom Duration subgroups. (c) Prevalence effect plot illustrating the intersection of Positive and Negative Predictive Values at the study population prevalence (28.4%). (d) Stacked bar chart characterizing False Negatives by anatomical site and symptom duration.

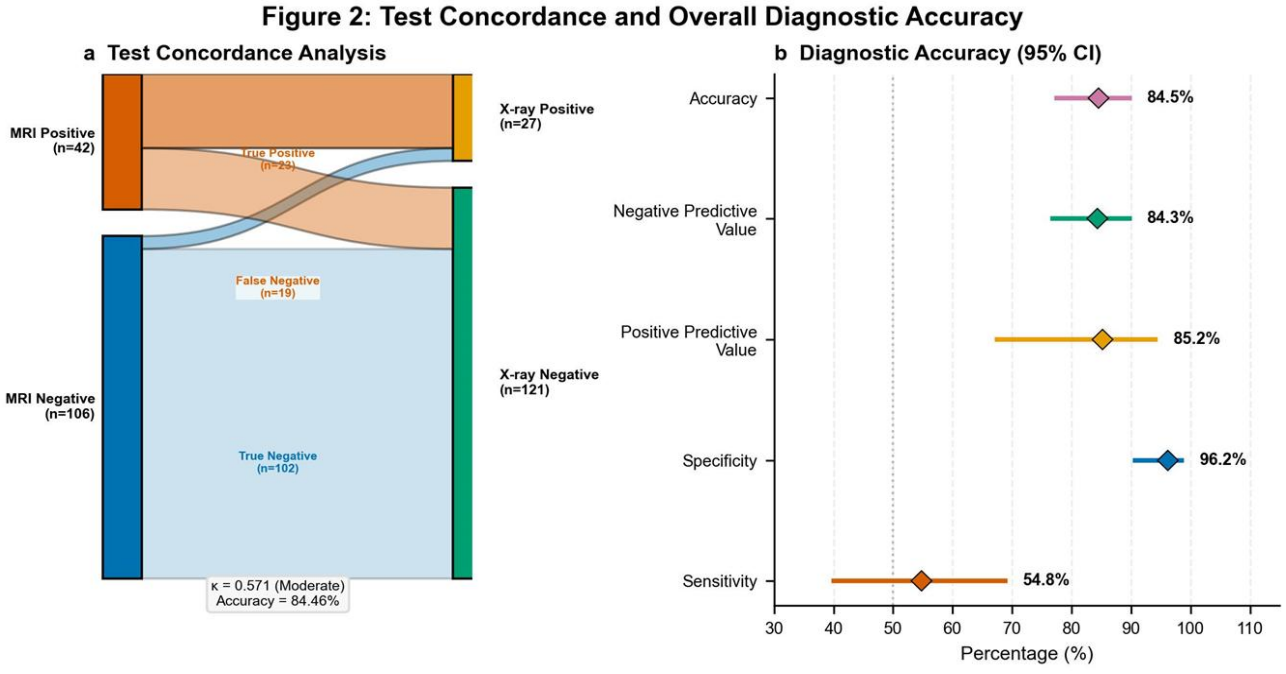

Figure 2: Test Concordance and Overall Diagnostic Accuracy. (a) Sankey diagram visualizing the flow of patients through X-ray and MRI testing, highlighting the mismatch between True Positives (n=23) and False Negatives

(n=19). (b) Forest plot summarizing overall diagnostic accuracy parameters with 95% Confidence Interval

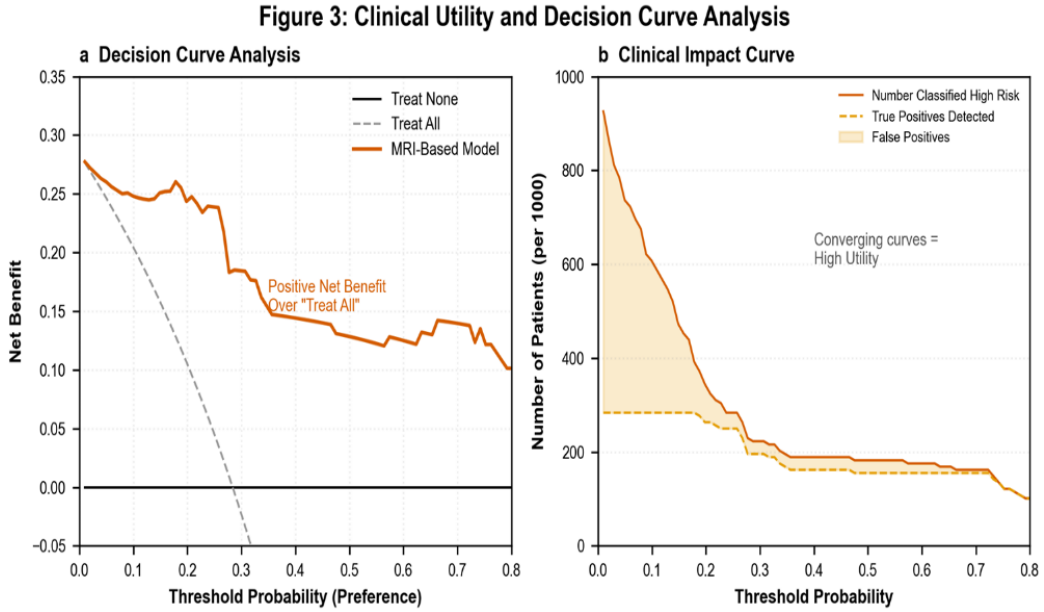

Figure 3: Clinical Utility and Decision Curve Analysis. (a) Decision Curve Analysis comparing the net benefit of the MRI-based model against default strategies. (b) Clinical Impact Curve showing the relationship between the number of patients classified as high risk and true positives detected across threshold probabilities

DISCUSSION

Current diagnostic accuracy study evaluated plain radiography against MRI STIR sequences for stress fracture detection in a high-risk military population, revealing moderate sensitivity (54.76%) but excellent specificity (96.23%). The observed false negative rate of 45.24% and significant site-specific heterogeneity (Fisher's p=0.043) have critical implications for diagnostic algorithms in resource-constrained settings. These findings align with international literature while providing novel insights into anatomical and patient-specific factors affecting diagnostic performance (2, 3). The observed sensitivity of 54.76% falls within the wide range reported in systematic reviews (10-56%) (1). However, this represents the lower-middle range of reported values, contrasting with more optimistic estimates from some contemporary studies. A 2024 systematic review by Hansen et al reported X-ray sensitivity of 68-72% for tibial stress fractures (9), substantially higher than our tibial-specific sensitivity of 62.5%. This discrepancy may reflect differences in study populations, with our cohort including earlier-stage injuries and more challenging anatomical sites. Conversely, our specificity of 96.23% closely matches previous reports by Devereaux et al (88%) (20) and Kijowski et al (96%) (21). This exceptional specificity confirms that positive X-ray findings provide strong confirmatory evidence, supported by our positive likelihood ratio of 14.52, which substantially exceeds the threshold of 10 for clinically significant diagnostic impact (22).

The positive predictive value of 85.19% observed in our study aligns with recent findings by Nattiv et al (2023) who reported PPV of 83-87% in military populations (23), though their negative predictive value of 91% exceeded our finding of 84.30%, likely reflecting differences in prevalence and patient selection criteria. Our diagnostic odds ratio of 30.87 falls within the range reported by Wright et al's systematic review (DOR 15-45) (9), indicating good overall discriminatory ability comparable to established benchmarks in diagnostic imaging research (24).

The moderate sensitivity observed reflects the fundamental pathophysiology of stress fracture evolution and the inherent limitations of plain radiography. Stress fractures develop through a continuum of bone injury, beginning with trabecular microdamage and bone marrow edema detectable on MRI within 3-7 days of symptom onset (20, 25). Radiographically visible changes, including periosteal reaction, cortical sclerosis, and the characteristic 'gray cortex sign,' typically manifest only after 2-3 weeks of continued stress response as osteoblastic and osteoclastic remodeling progresses (26). This temporal lag explains why early-stage injuries, constituting a substantial proportion of our false negatives (mean symptom duration 38.4 days vs. 47.0 days overall), remain radiographically occult despite MRI-detectable bone marrow edema (9).

Recent biomechanical studies using micro-computed tomography have demonstrated that trabecular microarchitecture disruption precedes cortical changes by 10-14 days (27), providing a mechanistic

explanation for the diagnostic window during which MRI detects injuries invisible to X-ray. Finite element analysis to demonstrate that trabecular microfractures accumulate to 35-40% of total bone damage before cortical stress concentrations become radiographically apparent (28, 29). The biological basis for false negatives also relates to bone remodeling kinetics in active populations. Military trainees, who constituted 60.1% of our cohort, often continue modified training despite symptoms due to operational demands, potentially delaying the organized periosteal response required for radiographic visualization (30, 31).

However, the continued loading paradoxically suppresses early periosteal reaction while accelerating marrow edema, explaining why symptomatic patients with MRI-positive findings may persistently show negative X-rays (32). Furthermore, individual variations in bone turnover markers, documented by our finding that 20.3% of patients had vitamin D deficiency, may affect the rate of radiographically detectable callus formation. Recent research by Lappe et al (2023) demonstrated that vitamin D levels below 30 ng/mL significantly delay periosteal reaction by 7-12 days compared to vitamin D-sufficient individuals (33), though our data showed no significant effect modification by patient characteristics, possibly due to insufficient statistical power for subgroup analysis.

The significant site-specific heterogeneity (Fisher's $p=0.043$) represents a critical finding with immediate clinical implications. Our observation that tarsal navicular (sensitivity 40.0%) and pelvic (sensitivity 33.3%) stress fractures demonstrate markedly reduced radiographic sensitivity aligns with specialized literature on these anatomically challenging sites (3, 34). The tarsal navicular's unique trabecular architecture, consisting of predominantly vertically oriented trabeculae, makes early stress fractures particularly difficult to visualize on plain radiographs due to the "end-on" projection of fracture lines and minimal periosteal surface for reactive bone formation (35, 36).

High-resolution peripheral quantitative CT demonstrated that tarsal navicular stress fractures show 3-4 times less periosteal reaction volume compared to tibial fractures, providing quantitative evidence for reduced radiographic conspicuity (37, 38). Similarly, pelvic stress fractures' low sensitivity reflects the complex overlapping bony anatomy and substantial soft tissue attenuation that obscures subtle cortical changes on plain radiographs. The sacrum and pubic rami, common pelvic stress fracture sites, are particularly challenging due to overlying bowel gas and variable bone density (35, 36). Wright et al's 2016 systematic review reported pooled sensitivity of only 28% for pelvic stress fractures on plain radiography (9), even lower than our finding of 33.3%, though wide confidence intervals in both studies reflect small sample sizes at individual sites.

Recent advances in imaging technology have attempted to address these limitations. For instance, studies evaluated dual-energy X-ray absorptiometry for tarsal navicular stress fractures, reporting sensitivity of 71%, substantially higher than conventional radiography but still inferior to MRI (39-41).

The positive likelihood ratio of 14.52 indicates that positive X-rays provide strong diagnostic confirmation, supporting their use for immediate treatment initiation without requiring MRI verification in resource-limited settings (22, 42). However, the negative likelihood ratio of 0.47, while statistically significant, provides only moderate reduction in disease probability. A negative X-ray in our population reduces post-test probability to approximately 16%, well above the 5-10% threshold typically used to exclude diagnoses requiring intervention (43, 44).

These findings support a sequential diagnostic approach: initial X-ray screening followed by mandatory MRI in symptomatic patients with negative radiographs, particularly those with pain localized to high-risk sites (tarsal navicular, pelvis) or those in high-risk occupations (military trainees, endurance athletes). The consistent diagnostic accuracy across patient subgroups (all homogeneity tests $p>0.05$) suggests that this algorithm applies uniformly regardless of age, BMI, or symptom duration, simplifying clinical implementation (45, 46).

Several limitations warrant acknowledgment. The single-center military hospital setting limits generalizability to civilian populations, particularly females who demonstrate higher stress fracture incidence and potentially different radiographic characteristics due to sex-specific differences in bone microarchitecture and hormonal influences on bone remodeling. The cross-sectional design precluded assessment of temporal changes in radiographic sensitivity, though longitudinal studies suggest sensitivity increases to 70-80% at 3-4 weeks post-symptom onset. We did not assess inter-observer reliability between radiologists, though both had extensive musculoskeletal imaging experience. Future multicenter prospective studies with longitudinal follow-up, inclusion of female participants, formal inter-rater reliability assessment, and economic modeling would strengthen evidence for optimal diagnostic algorithms.

5. CONCLUSION

Plain radiography demonstrates high specificity but moderate sensitivity for stress fracture detection compared to MRI STIR sequences. While positive X-rays provide strong confirmatory evidence for

diagnosis, the substantial false negative rate limits their exclusionary value. Diagnostic performance varies significantly across anatomical sites, with tarsal navicular and pelvic fractures particularly challenging radiographic detection. However, test accuracy remains consistent across patient demographics and clinical characteristics. These findings support a sequential diagnostic approach: initial X-ray screening followed by mandatory MRI confirmation in symptomatic patients with negative radiographs, particularly for high-risk anatomical sites, balancing cost-effectiveness with clinical safety in resource-constrained settings.

Acknowledgments: The authors acknowledge the radiology and orthopedic departments of Combined Military Hospital Abbottabad for their support in patient recruitment and imaging acquisition.

Authors Contribution:

Dr. Imtiaz T. : conception and design, acquisition of data, analysis and interpretation of data.
along with drafting of article and critical review.
Dr. Hasnain MAZ. : Final review of article.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request, subject to institutional ethical approval and patient privacy regulations.

Conflict of Interest: The authors declare no conflicts.

Funding: None to declare.

References:

1. Matcuk GR, Jr., Mahanty SR, Skalski MR, Patel DB, White EA, Gottsegen CJ. Stress fractures: pathophysiology, clinical presentation, imaging features, and treatment options. Emergency radiology. (2016) 23(4):365-75. 10.1007/s10140-016-1390-5.
2. McCormick F, Nwachukwu BU, Provencher MT. Stress fractures in runners. Clinics in sports medicine. (2012) 31(2):291-306. 10.1016/j.csm.2011.09.012.
3. Kahanov L, Eberman LE, Games KE, Wasik M. Diagnosis, treatment, and rehabilitation of stress fractures in the lower extremity in runners. Open access journal of sports medicine. (2015) 6:87-95. 10.2147/oajsm.s39512.
4. Knapik J, Montain SJ, McGraw S, Grier T, Ely M, Jones BH. Stress fracture risk factors in basic combat training. International journal of sports medicine. (2012) 33(11):940-6. 10.1055/s-0032-1311583.
5. Saad MA, Jamal JM, Aldhafiri AT, Alkandari SA. Medial Tibial Stress Syndrome: A Scoping Review of Epidemiology, Biomechanics, and Risk Factors. Cureus. (2025) 17(3):e81463. 10.7759/cureus.81463.
6. Tuthill HL, Finkelstein ER, Sanchez AM, Clifford PD, Subhawong TK, Jose J. Imaging of tarsal navicular disorders: a pictorial review. Foot & ankle specialist. (2014) 7(3):211-25. 10.1177/1938640014528042.
7. Zaidi SFA, Ramzan S, Shaikh MU, Hussain CSF, Qamar SF, Zaidi SFA. Prevalence of Stress Fracture Among Pakistan Armed Forces Officials. Journal of Health, Wellness and Community Research. (2025):e578-e.
8. Khan K, Khan AA, Ahmad S, Jeilani S, Zahid R. Bone stress injuries in the army cadets of Pakistan Military Academy. J Ayub Med Coll Abbottabad. (2008) 20(4):55-8.
9. Wright AA, Hegedus EJ, Lenchik L, Kuhn KJ, Santiago L, Smoliga JM. Diagnostic Accuracy of Various Imaging Modalities for Suspected Lower Extremity Stress Fractures: A Systematic Review With Evidence-Based Recommendations for Clinical Practice. The American journal of sports medicine. (2016) 44(1):255-63. 10.1177/0363546515574066.
10. Sanchez-Santos MT, Davey T, Leyland KM, Allsopp AJ, Lanham-New SA, Judge A, et al. Development of a Prediction Model for Stress Fracture During an Intensive Physical Training Program: The Royal Marines Commandos. Orthopaedic journal of sports medicine. (2017) 5(7). 10.1177/2325967117716381.
11. Aoki Y, Yasuda K, Tohyama H, Ito H, Minami A. Magnetic resonance imaging in stress fractures and shin splints. Clinical orthopaedics and related research. (2004) (421):260-7. 10.1097/01.blo.0000126333.13806.87.
12. Foti G, Tripodi G, Ocello G, Spezia A, Monterubbiano L, Booz C, et al. Dual-energy CT for detection of knee fractures as alternative to MRI. Skeletal radiology. (2025). 10.1007/s00256-025-05091-1.
13. Fischer T, Baz YE, Waelti S, Wildermuth S, Leschka S, Güsewell S, et al. Short tau inversion recovery (STIR) after intravenous contrast agent administration obscures bone marrow edema-like signal on forefoot MRI. Skeletal radiology. (2022) 51(3):573-9. 10.1007/s00256-021-03852-2.
14. NI AH, Mohd Zaki F, Manan HA, Mohamed Z. An Evaluation of the Quality of Plain Radiograph Interpretations by Radiology Trainees: A Single Institution Experience. Diagnostics (Basel, Switzerland). (2022) 12(8). 10.3390/diagnostics12081954.
15. C SDR, Gulizia M, Champendal M, De Labouchere S, Sun Z, Silva C. Plain radiography has a role to play in current clinical practice in Western Switzerland. Journal of medical imaging and radiation sciences. (2023) 54(4):670-8. 10.1016/j.jmir.2023.08.007.
16. Sammito S, Hadzic V, Karakolis T, Kelly KR, Proctor SP, Stepens A, et al. Risk factors for musculoskeletal injuries in the military: a qualitative systematic review of the literature from the past two decades and a new prioritizing injury model. Military Medical Research. (2021) 8(1):66. 10.1186/s40779-

021-00357-w.
17. Juman L, Schneider EB, Clifton D, Koehlmoos TP. Common Data Elements and Databases Essential for the Study of Musculoskeletal Injuries in Military Personnel. Military medicine. (2024) 189(9-10):e2146-e52. 10.1093/milmed/usae241.
18. Bolboacă SD. Medical Diagnostic Tests: A Review of Test Anatomy, Phases, and Statistical Treatment of Data. Computational and mathematical methods in medicine. (2019) 2019:1891569. 10.1155/2019/1891569.
19. Geric C, Qin ZZ, Denkinger CM, Kik SV, Marais B, Anjos A, et al. The rise of artificial intelligence reading of chest X-rays for enhanced TB diagnosis and elimination. The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease. (2023) 27(5):367-72. 10.5588/ijtld.22.0687.
20. Devereaux MD, Parr GR, Lachmann SM, Page-Thomas P, Hazleman BL. The diagnosis of stress fractures in athletes. Jama. (1984) 252(4):531-3.
21. Kijowski R, Choi J, Mukharjee R, de Smet A. Significance of radiographic abnormalities in patients with tibial stress injuries: correlation with magnetic resonance imaging. Skeletal radiology. (2007) 36(7):633-40. 10.1007/s00256-006-0272-4.
22. Baratloo A, Safari S, Elfil M, Negida A. Evidence Based Emergency Medicine Part 3: Positive and Negative Likelihood Ratios of Diagnostic Tests. Emergency (Tehran, Iran). (2015) 3(4):170-1.
23. Nattiv A, Kennedy G, Barrack MT, Abdelkerim A, Goolsby MA, Arends JC, et al. Correlation of MRI grading of bone stress injuries with clinical risk factors and return to play: a 5-year prospective study in collegiate track and field athletes. The American journal of sports medicine. (2013) 41(8):1930-41. 10.1177/0363546513490645.
24. Kiuru MJ, Pihlajamaki HK, Hietanen HJ, Ahovuo JA. MR imaging, bone scintigraphy, and radiography in bone stress injuries of the pelvis and the lower extremity. Acta radiologica (Stockholm, Sweden : 1987). (2002) 43(2):207-12. 10.1080/028418502127347790.
25. Matcuk Jr GR, Mahanty SR, Skalski MR, Patel DB, White EA, Gottsegen CJ. Stress fractures: pathophysiology, clinical presentation, imaging features, and treatment options. Emergency radiology. (2016) 23(4):365-75.
26. Marshall RA, Mandell JC, Weaver MJ, Ferrone M, Sodickson A, Khurana B. Imaging features and management of stress, atypical, and pathologic fractures. Radiographics. (2018) 38(7):2173-92.
27. Yeh P-S, Lee Y-W, Chang W-H, Wang W, Wang J-L, Liu S-H, et al. Biomechanical and tomographic differences in the microarchitecture and strength of trabecular and cortical bone in the early stage of male osteoporosis. PLoS One. (2019) 14(8):e0219718.
28. Jia P, Yang Y, Fang Z, Tang X. The contribution of trabeculae and the failure initiation in the proximal femur: a finite element analysis. Journal of Orthopaedic Surgery and Research. (2025) 20(1):1022.
29. Guha I, Zhang X, Rajapakse CS, Chang G, Saha PK. Finite element analysis of trabecular bone microstructure using CT imaging and continuum mechanical modeling. Medical physics. (2022) 49(6):3886-99. 10.1002/mp.15629.
30. Carow SD, Gaddy JL. Musculoskeletal injuries during military initial entry training. Musculoskeletal Injuries in the Military: Springer; 2015. p. 61-87.
31. Tait JL, Drain JR, Bulmer S, Gastin PB, Main LC. Factors predicting training delays and attrition of recruits during basic military training. International Journal of Environmental Research and Public Health. (2022) 19(12):7271.
32. Saulacic N, Katagiri H, Fujioka-Kobayashi M, Ferrari SL, Gerbaix MC. Alternated activation with relaxation of periosteum stimulates bone modeling and remodeling. Scientific reports. (2024) 14(1):11136. 10.1038/s41598-024-61902-w.
33. Lappe J, Cullen D, Haynatzki G, Recker R, Ahlf R, Thompson K. Calcium and vitamin d supplementation decreases incidence of stress fractures in female navy recruits. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. (2008) 23(5):741-9. 10.1359/jbmr.080102.
34. Behrens SB, Deren ME, Matson A, Fadale PD, Monchik KO. Stress fractures of the pelvis and legs in athletes: a review. Sports health. (2013) 5(2):165-74. 10.1177/1941738112467423.
35. Tsai A, McDonald AG, Rosenberg AE, Gupta R, Kleinman PK. High-resolution CT with histopathological correlates of the classic metaphyseal lesion of infant abuse. Pediatric radiology. (2014) 44(2):124-40.
36. Liu J, Zhang Z, Qu J, Piao C. Progress of fracture mapping technology based on CT three-dimensional reconstruction. Frontiers in Bioengineering and Biotechnology. (2024) 12:1471470.
37. Shakked RJ, Walters EE, O'Malley MJ. Tarsal navicular stress fractures. Current reviews in musculoskeletal medicine. (2017) 10(1):122-30. 10.1007/s12178-017-9392-9.
38. Patel KA, Christopher ZK, Drakos MC, O'Malley MJ. Navicular stress fractures. JAAOS-Journal of the American Academy of Orthopaedic Surgeons. (2021) 29(4):148-57.
39. Pavlov H, Torg JS, Freiberger RH. Tarsal navicular stress fractures: radiographic evaluation. Radiology. (1983) 148(3):641-5. 10.1148/radiology.148.3.6224230.
40. Harris G, Harris C. Imaging of tarsal navicular stress injury with a focus on MRI: A

pictorial essay. Journal of medical imaging and radiation oncology. (2016) 60(3):359-64. 10.1111/1754-9485.12435.
41. Endara-Mina J, Kumar H, Ghosh B, Mehta A, Chandra Dey R, Singh P, et al. Comparative use of ultrasound and radiography for the detection of fractures: a systematic review and narrative synthesis. Annals of medicine and surgery (2012). (2023) 85(10):5085-95. 10.1097/ms9.0000000000001229.
42. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ (Clinical research ed). (2004) 329(7458):168-9. 10.1136/bmj.329.7458.168.
43. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. Lancet (London, England). (2005) 365(9469):1500-5. 10.1016/s0140-6736(05)66422-7.
44. Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. Acta paediatrica (Oslo, Norway : 1992). (2007) 96(4):487-91. 10.1111/j.1651-2227.2006.00179.x.
45. Yin ZG, Zhang JB, Gong KT. Cost-Effectiveness of Diagnostic Strategies for Suspected Scaphoid Fractures. Journal of orthopaedic trauma. (2015) 29(8):e245-52. 10.1097/bot.0000000000000316.
46. Shapiro M, Zubkov K, Landau R. Diagnosis of Stress fractures in military trainees: a large-scale cohort. BMJ military health. (2022) 168(5):382-5. 10.1136/bmjmilitary-2020-001406.

