applied sciences

MDPI

Article

Using Artificial Intelligence for Detecting Diabetic Foot Osteomyelitis: Validation of Deep Learning Model for Plain Radiograph Interpretation

Francisco Javier Álvaro-Afonso [1,2], Aroa Tardáguila-García [1,2,*], Mateo López-Moral [1,2], Irene Sanz-Corbalán [1,2], Esther García-Morales [1,2] and José Luis Lázaro-Martínez [1,2]

[1] Diabetic Foot Unit, University Podiatric Clinic, Faculty of Nursing, Physiotherapy and Podiatry, Complutense University of Madrid, 28040 Madrid, Spain; alvaro@ucm.es (F.J.Á.-A.); matlopez@ucm.es (M.L.-M.); irsanz01@ucm.es (I.S.-C.); eagarcia@ucm.es (E.G.-M.); diabetes@ucm.es (J.L.L.-M.)
[2] San Carlos Clinical Hospital Health Research Institute (IdISSC), 28040 Madrid, Spain
* Correspondence: aroa.tardaguila@ucm.es

Abstract

Objective: To develop and validate a ResNet-50-based deep learning model for automatic detection of osteomyelitis (DFO) in plain radiographs of patients with diabetic foot ulcers (DFUs). Research Design and Methods: This retrospective study included 168 patients with type one or type two diabetes and clinical suspicion of DFO confirmed via a surgical bone biopsy. An experienced clinician and a pretrained ResNet-50 model independently interpreted the radiographs. The model was developed using Python-based frameworks with ChatGPT assistance for coding. The diagnostic performance was assessed against the histopathological findings, calculating sensitivity, specificity, the positive predictive value (PPV), the negative predictive value (NPV), and the likelihood ratios. Agreement between the AI model and the clinician was evaluated using Cohen's kappa coefficient. Results: The AI model demonstrated high sensitivity (92.8%) and PPV (0.97), but low-level specificity (4.4%). The clinician showed 90.2% sensitivity and 37.8% specificity. The Cohen's kappa coefficient between the AI model and the clinician was −0.105 (p = 0.117), indicating weak agreement. Both the methods tended to classify many cases as DFO-positive, with 81.5% agreement in the positive cases. Conclusions: This study demonstrates the potential of IA to support the radiographic diagnosis of DFO using a ResNet-50-based deep learning model. AI-assisted radiographic interpretation could enhance early DFO detection, particularly in high-prevalence settings. However, further validation is necessary to improve its specificity and assess its utility in primary care.

Keywords: artificial intelligence; diabetic foot; diabetic foot osteomyelitis

Academic Editors: Ivan Julian-Rochina, Jose Piñeiro Ramos and Lorenzo Brognara

Received: 4 July 2025
Revised: 26 July 2025
Accepted: 29 July 2025
Published: 1 August 2025

Citation: Álvaro-Afonso, F.J.; Tardáguila-García, A.; López-Moral, M.; Sanz-Corbalán, I.; García-Morales, E.; Lázaro-Martínez, J.L. Using Artificial Intelligence for Detecting Diabetic Foot Osteomyelitis: Validation of Deep Learning Model for Plain Radiograph Interpretation. *Appl. Sci.* **2025**, *15*, 8583. https://doi.org/10.3390/app15158583

1. Introduction

Diabetic foot osteomyelitis (DFO) is one of the most severe complications of diabetic foot infections, affecting approximately 20–60% of moderate-to-severe cases [1–3]. This condition is associated with high morbidity, prolonged hospitalization, and an increased risk of lower limb amputation [3,4]. Early and accurate diagnosis of DFO is critical for timely intervention and limb preservation, preventing amputations and reducing the mortality rates [5]. Although a bone biopsy remains the gold standard for diagnosing

DFO, less-invasive imaging modalities have demonstrated comparable accuracy in recent years [1].

The conventional diagnostic approaches for DFO include clinical assessment, laboratory markers, and imaging studies [6]. Among the imaging modalities, plain radiography is the most widely used first-line tool due to its accessibility, low cost, and non-invasive nature [6,7]. However, its diagnostic sensitivity ranges from 43% to 75%, particularly in the early stages before bone destruction becomes radiographically apparent. Moreover, plain radiography interpretation is highly operator-dependent, demonstrating low-level interobserver agreement even among experienced clinicians when evaluated in isolation without additional clinical context [8,9]. Advanced imaging techniques, such as positron emission tomography–computed tomography (PET-CT) and magnetic resonance imaging (MRI), offer higher sensitivity, but are expensive, not always available in routine clinical settings, and may be contraindicated in certain patient populations [10]. Recent meta-analyses suggest that combining plain radiography with simple diagnostic tools, such as the probe-to-bone (PTB) test, can enhance the diagnostic accuracy [1].

Considering these diagnostic limitations, artificial intelligence (AI) and deep learning (DL) have emerged as promising tools for medical image analysis. They aim to enhance diagnostic accuracy and reduce interobserver variability [11–13]. Deep convolutional neural networks (CNNs) such as ResNet-50 have shown a strong performance in medical imaging tasks, including the detection of fractures, tumors, glaucoma, and vascular changes [14–23]. While AI has been widely studied in medical imaging, its application to diabetic foot complications remains limited. Most existing studies have focused on ulcer detection, segmentation, and risk prediction using clinical data and photographic images [24,25]. Only a few have addressed the diagnosis of DFO, and those that have typically rely on advanced imaging techniques such as MRI [26]. The use of AI for DFO detection based on plain radiographs is therefore still underexplored. Our study addresses this gap by validating a ResNet-50-based DL model using histopathological confirmation as the gold standard, offering a novel and accessible diagnostic tool, particularly valuable in resource-limited settings. AI models can analyze large volumes of imaging data, detect subtle radiographic patterns that may elude the human eye, and ensure consistent, objective evaluations.

This study aims to develop and validate a ResNet-50-based DL model for the automatic detection of DFO in plain radiographs of patients with DFU. This could potentially provide a robust and scalable tool for enhancing clinical decision making.

To our knowledge, this is the first study to validate a ResNet-50-based deep learning model for the detection of DFO using plain radiographs, with histopathological confirmation. This represents a novel approach within the context of diabetic foot diagnostics, where AI applications remain largely underexplored.

2. Materials and Methods

2.1. Study Design and Data Acquisition

We compiled a retrospective dataset of radiographic images from 2020 to 2024 at our center. This dataset included individuals with type 1 and type 2 diabetes and diabetic foot ulcers (DFUs) with a clinical suspicion of osteomyelitis. These individuals underwent surgery and had bone biopsies for histopathological confirmation. A clinician experienced in managing DFU and a pretrained ResNet-50 model interpreted the radiographs. Both the clinician and the AI model were informed about the location of the DFU before interpretation. Osteomyelitis was diagnosed based on the presence of at least one of the following radiographic findings: active periosteal reaction, cortical disruption (the loss of cortex with bony erosion), affected bone marrow (focal loss of

trabecular pattern or marrow radiolucency), and a sequestrum (devitalized bone with a radio-dense appearance that has become separated from the normal bone) [6]. The diagnostic performance of the clinician and the AI model was assessed against the histopathological findings from bone biopsy samples. Those samples were obtained in the operating theater and examined by a pathologist [27].

2.2. Model Selection, Implementation, Training and Optimization

A CNN architecture based on ResNet-50 was implemented [28]. ResNet-50 was chosen due to its proven effectiveness in medical image classification and its ability to extract hierarchical features from radiographic images [29]. The model was initialized with ImageNet weights and refined using transfer learning techniques to optimize the performance for DFO detection [30].

The training dataset consisted of manually labeled radiographs, categorized as either osteomyelitis-positive or osteomyelitis-negative cases (Figure 1a,b). The data were divided into training (80%) and validation (20%) subsets. To boost diagnostic precision, we used supervised learning with balanced classes to confidently differentiate between the DFO-positive and DFO-negative cases. Hyperparameters, such as the batch size and the learning rate, were optimized for convergence.

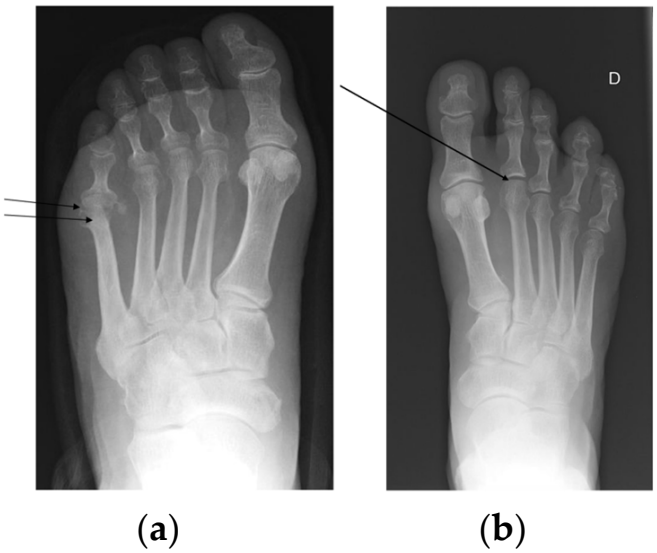


Figure 1. (**a**) A radiographic example of an osteomyelitis-positive case. The affected region is located at the fifth metatarsal. The observed radiographic signs include the loss of cortex with bony erosion and affected bone marrow (focal loss of trabecular pattern or marrow radiolucency). (**b**) A radiographic example of an osteomyelitis-negative case. The analyzed region is located beneath the second metatarsal head. No radiographic signs indicative of osteomyelitis, such as cortical disruption and medullary involvement, were observed.

A web-based interface was created using Flask, permitting users to upload radiographic images, receive real-time AI-based diagnostic predictions, and download structured reports (Figure 2). The system was established in a Python-based environment, integrating Torchvision and PIL for image preprocessing [31]. To streamline model deployment, ChatGPT (GPT-4.5, OpenAI) was used to support the integration of the AI model into the web-based diagnostic interface, focusing on code refinement and system functionality.

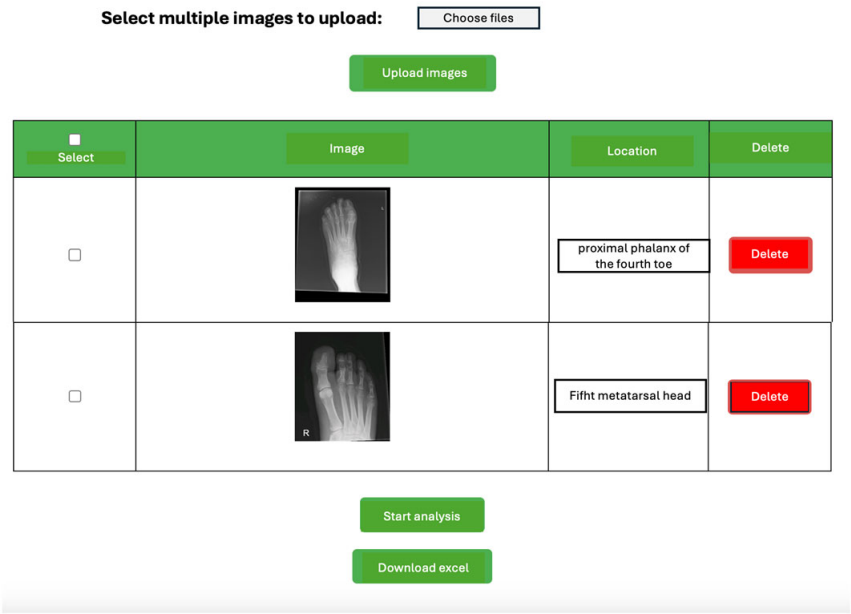


Figure 2. The web-based interface developed using Flask. The system allows users to upload radiographic images, specify lesion locations, and receive AI-based diagnostic predictions. The system allows users to initiate analysis and generate structured diagnostic reports for clinical interpretation.

2.3. Image Preprocessing and Augmentation

The radiographic images underwent uniform preprocessing to enhance model generalization and reduce overfitting [32].

Rescaling: All the images were resized to 224 × 224 pixels to meet the input requirements of the ResNet-50 architecture.

Normalization: Standard normalization was applied using the mean and standard deviation values from the ImageNet dataset (mean = [0.485, 0.456, 0.406]; std = [0.229, 0.224, 0.225]).

Augmentation: During training, data augmentation techniques—including horizontal flipping, contrast adjustments, and small rotations—were consistently applied to all training folds to increase model robustness.

Uniformity and acquisition protocol: All the images were acquired from the same radiological equipment and processed using a consistent acquisition protocol, thereby minimizing the risk of batch effects related to imaging source or variability between devices.

2.4. Inference and Decision Thresholds

Once trained, the model performed binary classification, outputting a confidence score between 0 and 1. To improve diagnostic accuracy, the initial classification threshold of 0.50 was adjusted to 0.60 according to the performance of the validation set to balance the rates of false positives and false negatives [33]. The final diagnostic output was displayed in the web application, with the optional probability score indicating the confidence level of each prediction [34].

2.5. Software and Code Availability

The AI pipeline was developed using Python (version 3.8), integrating several key libraries, such as PyTorch (used for constructing and training the CNN model); Flask (used for the development of the web-based interface for user interaction); Pandas (for tasks surrounding data manipulation and analysis); and OpenCV (used for various image processing operations).

2.6. Validation and Clinical Comparison

The diagnostic accuracy of both the seasoned diabetic foot clinician and the AI model was compared to a gold standard (histological bone analysis). Sensitivity, specificity, the positive predictive value (PPV), the negative predictive value (NPV), and likelihood ratios were determined for both the clinician and the model. In addition, the inter-rater agreement between the experienced clinician and the AI model was examined.

Statistical analysis was performed using the SPSS v27 statistical package. The Kolmogorov–Smirnov test was utilized to verify the assumption of normality for all continuous variables using IOs version 21.0 (SPSS, Inc. Chicago, IL, USA). To assess the inter-rater reliability between the AI model and the experienced clinician, Cohen's kappa coefficient was implemented, referencing the Landis and Koch criteria to explore the strength of association [35]. McNemar's test was additionally applied to contrast the performances of the AI and the clinician in differentiating between the osteomyelitis-positive and osteomyelitis-negative cases. To evaluate the various accuracy parameters (such as sensitivity and specificity), a diagnostic test calculator was employed (MedCalc Software Ltd. Diagnostic Test Evaluation Calculator. https://www.medcalc.org/calc/diagnostic_test.php (Version 23.1.7, accessed on 12 March 2025)).

2.7. Sample Size Justification

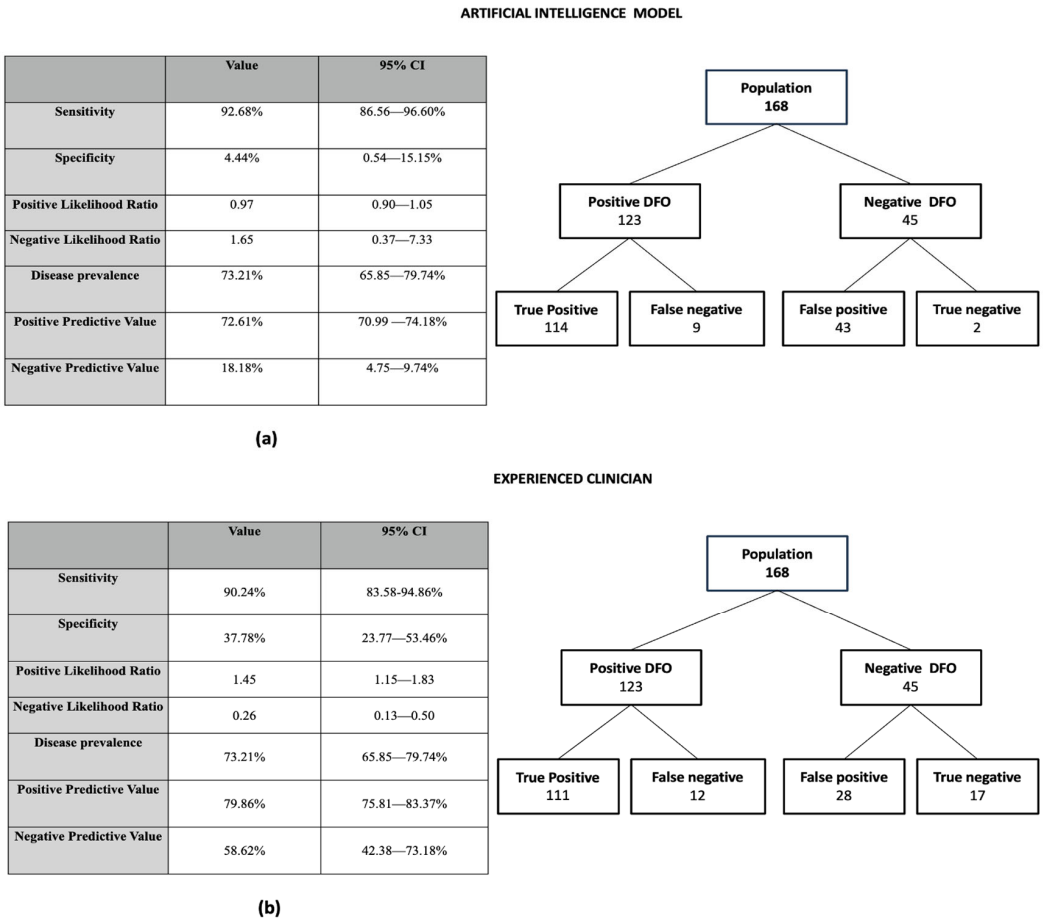
The sample size of our study is 168 patients, a number supported by previous validation studies evaluating the diagnostic accuracy of radiography for osteomyelitis using histopathology as the gold standard. A recent meta-analysis identified seven key studies, with sample sizes ranging from 19 to 356 patients. This includes the study by Aragón-Sánchez et al., which had the largest cohort of 356 patients [1,36–42]. Our sample size falls within the higher range of these previous studies, exceeding the median sample size among them. Since prior studies with fewer participants have been deemed methodologically valid for assessing radiographic diagnostic performance, our sample size of 168 patients is deemed sufficient to ensure statistical power and generalizability of the findings.

3. Results

This study included a total of 168 patients, with a mean age of 61.9 $\pm$ 11.6 years. The cohort was made up of 127 men (75.6%) and 41 women (24.4%). The median duration of having had diabetes was 14.0 years (IQR: 8.0–25.0), with 20 patients (11.9%) diagnosed with type one diabetes, and 148 (88.1%) with type two diabetes. The median glycated hemoglobin level was 55.2 mmol/mol (IQR: 46.4–69.4). The median wound duration was 7.5 weeks (IQR: 2.0–20.0).

The most common location for DFUs was the central metatarsals (32.1%, n = 54), followed by the lesser toes (23.8%, n = 40), the fifth metatarsal (21.4%, n = 36), the first metatarsal (14.3%, n = 24), the hallux (7.1%, n = 12), and lastly the rearfoot (1.2%, n = 2). In terms of ulcer types, 58.3% (n = 98) were type III B, 35.1% (n = 59) were type III D, 4.2% (n = 7) were of type IIB, and the remainder, 2.4% (n = 4), were type IIID.

Figure 3a,b outline the validation metrics for the AI model and the experienced clinician, respectively. The tables within these figures present crucial diagnostic performance parameters—these include sensitivity, specificity, predictive values, and likelihood ratios. Moreover, the confusion matrix diagrams display the distribution of classification outcomes within the study population, offering a comparative perspective on the diagnostic accuracy of both these methodologies.



	Value	95% CI
Sensitivity	92.68%	86.56—96.60%
Specificity	4.44%	0.54—15.15%
Positive Likelihood Ratio	0.97	0.90—1.05
Negative Likelihood Ratio	1.65	0.37—7.33
Disease prevalence	73.21%	65.85—79.74%
Positive Predictive Value	72.61%	70.99 —74.18%
Negative Predictive Value	18.18%	4.75—9.74%

	Value	95% CI
Sensitivity	90.24%	83.58-94.86%
Specificity	37.78%	23.77—53.46%
Positive Likelihood Ratio	1.45	1.15—1.83
Negative Likelihood Ratio	0.26	0.13—0.50
Disease prevalence	73.21%	65.85—79.74%
Positive Predictive Value	79.86%	75.81—83.37%
Negative Predictive Value	58.62%	42.38—73.18%

Figure 3. (**a**) The model validation metrics and a confusion matrix representation. This table summarizes the diagnostic performance of the AI model, while the diagram illustrates the distribution of classification outcomes within the study population. (**b**) The experienced clinician validation metrics and a confusion matrix representation. This table summarizes the diagnostic performance of the experienced clinician, while the diagram illustrates the distribution of classification outcomes within the study population.

Additionally, Table 1 illustrates the two-by-two agreement matrix between the AI model and the experienced clinician for the diagnosis of diabetic foot osteomyelitis (DFO). The number of cases classified as positive or negative by each method is presented. Observed agreement was assessed using Cohen's kappa coefficient, which yielded a value of $\kappa = -0.105$ ($p = 0.117$), indicating poor agreement between both the approaches and suggesting differences in their diagnostic criteria.

Table 1. Agreement between the AI model and the experienced clinician in diagnosing diabetic foot osteomyelitis. This table presents the cross-tabulation of classification outcomes, showing the number of cases classified as positive (yes) and negative (not) by both the methods.

AI Model	Experienced Clinician—Not	Experienced Clinician—Yes	Total
Not	0	11	11
Yes	29	128	157
Total	29	139	168

To complement this numerical summary, Figure 4 presents an annotated Venn diagram that visually represents the overlap in positive classifications. The AI model identified 157 cases as positive, and the clinician 139 cases, with 128 classified as positive by both.

The diagram clearly highlights the 29 cases labeled as positive only by the AI model and the 11 cases labeled as positive only by the clinician, offering a visual representation of the discordant classifications.

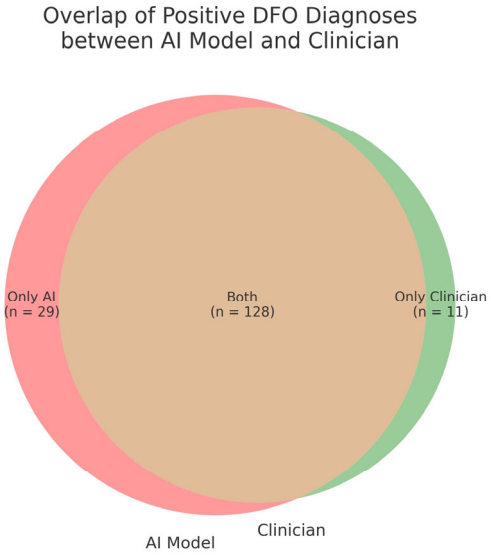


Figure 4. An annotated Venn diagram showing the overlap in positive diabetic foot osteomyelitis (DFO) classifications between the AI model and the experienced clinician. The left circle represents the cases classified as positive by the AI model (n = 157), and the right circle those classified as positive by the clinician (n = 139). The overlapping region (n = 128) corresponds to cases classified as positive by both these methods. The "Only AI" segment includes cases identified as positive exclusively by the AI model (n = 29), while the "Only Clinician" segment includes those identified only by the clinician (n = 11).

4. Discussion

This study shows that the AI model achieved high sensitivity (92.8%) for detecting DFO on plain radiographs, surpassing the performance of an experienced clinician (90.24%). The model also demonstrated a high positive predictive value (PPV) of 0.97. In comparison to the previous studies, which used histopathology as the gold standard, our AI model showed superior sensitivity to those of Morales-Lozano et al. (89%), Aragón-Sánchez et al. (82%), and Álvaro-Afonso et al. (86%) [36,37,42]. These outcomes highlight the potential of AI-assisted radiographic interpretation in specialized diabetic foot units.

Despite its strong sensitivity, the AI model demonstrated low-level specificity. This phenomenon is in alignment with previous findings, as indicated in a recent meta-analysis [1], which scrutinized seven validation studies that used histopathology as the gold standard. In these studies, the specificity value was consistently lower than the sensitivity value. This trend can be ascribed to the high prevalence of DFO in specialized diabetic foot centers, where validation studies are typically undertaken. For example, Morales-Lozano et al. [37] reported a prevalence of 79.5%, Aragón-Sánchez et al. [36] reported 72.5%, and Álvaro-Afonso et al. [42] reported 66.7%. In our study, the prevalence of DFO was 73.2%, which reinforces that our cohort represents a high-prevalence setting.

In addition to internal validation parameters, assessing reproducibility is crucial for evaluating a diagnostic tool. Our study found a very low level of agreement between the AI model and the experienced clinician, with a Cohen's kappa coefficient of K = −0.105 (p = 0.117). Similar findings have been reported in previous studies assessing the reproducibility of radiographic interpretation for DFO among clinicians with varying levels of expertise. Álvaro-Afonso et al. (2013) [8] found low-level inter-clinician agreement (K = 0.35, $p < 0.01$) among experts evaluating radiographs for DFO. A subsequent validation study in

2018 showed similar results (K = 0.40, $p < 0.001$) [42]. Likewise, Morales-Lozano et al. [37] reported a low level of inter-clinician agreement (K = 0.136, $p = 0.104$). One potential explanation for the low-level agreement observed in our study is that both the AI model and the experienced clinician interpreted radiographs solely based on the ulcer location, without access to additional clinical information. Prior studies have demonstrated that incorporating clinical assessments such as the PTB test improves diagnostic consistency. Álvaro-Afonso et al. [9] reported that when radiographic interpretation was combined with PTB, the degree of inter-clinician agreement significantly increased (K = 0.77, $p < 0.001$). Similarly, Aragón-Sánchez et al. [36] demonstrated that combining radiography with PTB increased the level of sensitivity from 82% (radiography alone) to 97% (radiography + PTB).

A significant observation is that the AI model ($n = 157$) and the proficient clinician ($n = 138$) both tended to categorize a considerable number of patients as positive for DFO. They concurred on 128 positive cases, equating to an 81.5% agreement rate for positive diagnoses, as shown in Table 1. However, there was a lack of concordance for the negative instances. More precisely, the clinician classified 29 cases as negative that were identified as positive by the AI model, while the AI model classified 11 cases as negative that the clinician tagged as positive. These results underline the necessity for additional refinement of the AI model, particularly to enhance specificity in identifying patients without DFO. This distribution of discordant cases is further illustrated in Figure 4, which provides a visual summary of the overlap between both the diagnostic approaches using an annotated Venn diagram. While the level of numerical agreement for positive classifications is high, the graphical representation emphasizes the asymmetry in non-matching positive classifications between the AI model and the clinician. This visualization highlights the need to better understand the factors contributing to diagnostic divergence, particularly in cases with subtle or borderline radiographic features.

The findings of this study suggest that AI-based diagnostic models for DFO might be especially beneficial in specialized diabetic foot units, where the disease prevalence exceeds 70%. The subsequent challenge involves optimizing these models to enhance specificity, thus ensuring their application in primary care environments. By bolstering specificity, AI-driven tools could aid non-specialist clinicians in spotting high-risk cases, thereby enabling timely referrals to specialized centers.

Importantly, the integration of AI tools into clinical workflows is not intended to replace expert judgment, but rather to provide valuable support in settings where access to specialized expertise is limited. In primary care or rural environments, where imaging interpretation is often suboptimal and access to MRI or nuclear medicine may be restricted, an AI system could serve as a triage mechanism, flagging high-risk radiographs for referral or further assessment. Additionally, in specialized units, such tools may function as second readers to reduce perceptual errors and mitigate the impact of cognitive fatigue in high-volume practices.

4.1. Study Limitations

Several limitations must be acknowledged. Firstly, the retrospective nature of this study could introduce selection bias. Secondly, the radiographic interpretations were conducted without knowledge of additional clinical characteristics such as the PTB results. This lack of information could have influenced the diagnostic accuracy. Importantly, neither the AI model nor the clinician had access to this clinical information.

Furthermore, the AI model (PLR = 0.97 [0.90–1.05]) and the clinician (PLR = 1.45 [1.11–1.83]) exhibited low positive likelihood ratios. These findings suggest that while the model performs well in high-prevalence settings, further validation is required to determine its applicability in primary care settings, where DFO prevalence is lower.

In addition, the low-level specificity observed in our model (4.4%) reflects its prioritization of sensitivity in order to minimize missed diagnoses in a high-risk population. While this trade-off may be acceptable in specialized diabetic foot units—where the goal is early detection and timely referral—it may limit the model's usefulness in settings with lower DFO prevalence. Future iterations should aim to improve specificity by integrating complementary clinical variables (e.g., probe-to-bone test, ulcer duration, lesion site, and inflammatory markers) and optimizing the classification thresholds, particularly to adapt the model to broader clinical environments.

Additionally, while the model was trained and validated specifically for diabetic foot ulcers (DFUs), its applicability to other types of chronic ulcers or osteomyelitis of non-diabetic origin remains unknown. Future research is required to evaluate the generalizability of the model across different ulcer etiologies and clinical contexts.

4.2. Clinical Implications and Future Directions

The proposed AI system is integrated into a web-based interface, enabling real-time radiographic analysis and structured diagnostic support. This technology has the potential to improve clinical workflows by standardizing the evaluation of plain radiographs, reducing interobserver variability, and expediting decision making. The ability to upload images, annotate lesion locations, and download automated reports contributes to its usability in both high-volume and resource-limited settings.

Future iterations of this platform could benefit from the integration of multimodal data—such as probe-to-bone test results, inflammatory biomarkers, and clinical photographs—through ensemble learning strategies. Such integration may enhance both the sensitivity and specificity of the system, broadening its utility at different levels of care.

Although artificial intelligence has been widely explored in fields such as oncology, musculoskeletal imaging, and cardiology [11–13,16,30,43,44], its application to diabetic foot infections, and particularly to diabetic foot osteomyelitis detection, remains limited. This study contributes to filling that gap by validating a deep learning model tailored specifically for radiographic assessment of diabetic foot osteomyelitis. As such, it provides a foundation for future prospective validations and implementation studies.

Further research is required to explore the performance of AI-based diagnostic systems across diverse healthcare settings, including primary care and community health environments. To enhance generalizability, future iterations of the model should incorporate a greater proportion of negative cases and be prospectively validated in lower-prevalence populations. Integrating multimodal clinical data—such as ulcer duration, anatomical location, laboratory markers, and probe-to-bone test results—may also improve specificity. Moreover, evaluating the integration of these tools into real-world clinical practice, considering usability, interpretability, and clinician trust, will be essential to ensure successful adoption. In the long term, AI-driven decision support systems may assist clinicians in achieving earlier and more accurate diagnoses of DFO, ultimately contributing to the prevention of major amputations and a reduction in diabetes-related morbidity and mortality.

5. Conclusions

This study demonstrates the potential of artificial intelligence to support the radiographic diagnosis of diabetic foot osteomyelitis (DFO) using a ResNet-50-based deep learning model. The AI system achieved excellent sensitivity and a high positive predictive value, exceeding the clinician's performance in sensitivity, although showing a substantially lower level of specificity. These findings support the usefulness of AI-assisted interpretation as a complementary diagnostic tool in specialized diabetic foot units, where early detection of bone infection is critical for preserving limb integrity.

While the AI model outperformed the clinician in terms of sensitivity, it showed a substantially lower level of specificity. This trade-off suggests that the model may be useful in high-prevalence or specialist settings as a diagnostic aid, but further refinement is required before broader clinical deployment. Integrating additional clinical variables and expanding the training dataset with more negative cases may help improve the diagnostic performance and improve generalizability.

The development of a web-based diagnostic interface further supports the clinical translation of this technology, enabling rapid, accessible, and standardized evaluation of radiographs. Future research should focus on prospective validation in real-world scenarios, exploration of multimodal models, and assessment of clinical impact on decision making and patient outcomes.

Overall, AI-based tools offer a promising avenue to enhance diagnostic accuracy, reduce interobserver variability, and facilitate timely management of diabetic foot infections. While they are not intended to replace expert clinical judgment, these systems may play a valuable role in supporting clinicians across diverse healthcare settings, ultimately contributing to better patient care and reduced complication rates.

Author Contributions: Conceptualization, F.J.Á.-A., A.T.-G. and J.L.L.-M.; methodology, F.J.Á.-A., A.T.-G. and I.S.-C.; software, F.J.Á.-A.; validation, F.J.Á.-A., A.T.-G. and J.L.L.-M.; formal analysis, F.J.Á.-A.; investigation, F.J.Á.-A., A.T.-G. and J.L.L.-M.; resources, F.J.Á.-A. and A.T.-G.; data curation, F.J.Á.-A.; writing—original draft preparation, F.J.Á.-A., A.T.-G. and J.L.L.-M.; writing—review and editing, F.J.Á.-A., A.T.-G., M.L.-M., I.S.-C., E.G.-M. and J.L.L.-M.; visualization, M.L.-M., I.S.-C. and E.G.-M.; supervision, J.L.L.-M.; project administration, F.J.Á.-A. and J.L.L.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki. The corresponding institutional review board granted ethical approval, which ensured compliance with data protection regulations (24/738-E, approval date: 19 November 2024).

Informed Consent Statement: Patient consent was waived due to the retrospective nature of this study.

Data Availability Statement: The data are not publicly available due to privacy or ethical restrictions.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-4.5, OpenAI) for assistance with Python code implementation, debugging, and model deployment within the web interface. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
CNNs	Deep convolutional neural networks
DFO	Diabetic foot osteomyelitis
DFU	Diabetic foot ulcer
DL	Deep learning
MRI	Magnetic resonance imaging
NPV	Negative predictive value
PET-CT	Positron emission tomography–computed tomography
PPV	Positive predictive value
PTB	Probe-to-bone

References

1. Calvo-Wright, M.D.M.; Álvaro-Afonso, F.J.; López-Moral, M.; García-Álvarez, Y.; García-Morales, E.; Lázaro-Martínez, J.L. Is the Combination of Plain X-ray and Probe-to-Bone Test Useful for Diagnosing Diabetic Foot Osteomyelitis? A Systematic Review and Meta-Analysis. *J. Clin. Med.* **2023**, *12*, 5369. [CrossRef] [PubMed] [PubMed Central]
2. Shahbazian, H.; Yazdanpanah, L.; Latifi, S.M. Risk assessment of patients with diabetes for foot ulcers according to risk classification consensus of International Working Group on Diabetic Foot (IWGDF). *Pak. J. Med. Sci.* **2013**, *29*, 730–734. [CrossRef] [PubMed] [PubMed Central]
3. Armstrong, D.G.; Boulton, A.J.M.; Bus, S.A. Diabetic Foot Ulcers and Their Recurrence. *N. Engl. J. Med.* **2017**, *376*, 2367–2375. [CrossRef] [PubMed]
4. Lázaro Martínez, J.L.; García Álvarez, Y.; Tardáguila-García, A.; García Morales, E. Optimal management of diabetic foot osteomyelitis: Challenges and solutions. *Diabetes Metab. Syndr. Obes.* **2019**, *12*, 947–959. [CrossRef] [PubMed] [PubMed Central]
5. Armstrong, D.G.; Swerdlow, M.A.; Armstrong, A.A.; Conte, M.S.; Padula, W.V.; Bus, S.A. Five year mortality and direct costs of care for people with diabetic foot complications are comparable to cancer. *J. Foot Ankle Res.* **2020**, *13*, 16. [CrossRef] [PubMed] [PubMed Central]
6. Senneville, É.; Albalawi, Z.; van Asten, S.A.; Abbas, Z.G.; Allison, G.; Aragón-Sánchez, J.; Embil, J.M.; Lavery, L.A.; Alhasan, M.; Oz, O.; et al. IWGDF/IDSA guidelines on the diagnosis and treatment of diabetes-related foot infections (IWGDF/IDSA 2023). *Diabetes Metab. Res. Rev.* **2024**, *40*, e3687. [CrossRef] [PubMed]
7. Senneville, É.; Albalawi, Z.; van Asten, S.A.; Abbas, Z.G.; Allison, G.; Aragón-Sánchez, J.; Embil, J.M.; Lavery, L.A.; Alhasan, M.; Oz, O.; et al. Diagnosis of infection in the foot of patients with diabetes: A systematic review. *Diabetes Metab. Res. Rev.* **2024**, *40*, e3723. [CrossRef] [PubMed]
8. Álvaro-Afonso, F.J.; Lázaro-Martínez, J.L.; Aragon-Sanchez, J.; García-Morales, E.; Cecilia-Matilla, A.; Beneit-Montesinos, J.V. Interobserver and intraobserver reproducibility of plain X-rays in the diagnosis of diabetic foot osteomyelitis. *Int. J. Low Extrem. Wounds* **2013**, *12*, 12–15. [CrossRef]
9. Álvaro-Afonso, F.J.; Lázaro-Martínez, J.L.; Aragon-Sanchez, J.; García-Morales, E.; Garcia-Alvarez, Y.; Molines-Barroso, R.J. Inter-observer reproducibility of diagnosis of diabetic foot osteomyelitis based on a combination of probe-to-bone test and simple radiography. *Diabetes Res. Clin. Pract.* **2014**, *105*, e3–e5. [CrossRef]
10. Lauri, C.; Tamminga, M.; Glaudemans, A.W.J.M.; Juárez Orozco, L.E.; Erba, P.A.; Jutte, P.C.; Lipsky, B.A.; IJzerman, M.J.; Signore, A.; Slart, R.H.J.A. Detection of Osteomyelitis in the Diabetic Foot by Imaging Techniques: A Systematic Review and Meta-analysis Comparing MRI, White Blood Cell Scintigraphy, and FDG-PET. *Diabetes Care* **2017**, *40*, 1111–1120. [CrossRef] [PubMed]
11. Quellec, G.; Charrière, K.; Boudi, Y.; Cochener, B.; Lamard, M. Deep image mining for diabetic retinopathy screening. *Med. Image Anal.* **2017**, *39*, 178–193. [CrossRef] [PubMed]
12. Paredes, D.; Saha, A.; Mazurowski, M.A. Deep learning for segmentation of brain tumors: Can we train with images from dif-ferent institutions? In *Medical Imaging 2017: Computer-Aided Diagnosis*; SPIE: Bellingham, WA, USA, 2017; Volume 10134.
13. Shen, D.G.; Wu, G.R.; Suk, H.I. Deep Learning in Medical Image Analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [CrossRef] [PubMed]
14. Liu, K.; Qin, S.; Ning, J.; Xin, P.; Wang, Q.; Chen, Y.; Zhao, W.; Zhang, E.; Lang, N. Prediction of Primary Tumor Sites in Spinal Metastases Using a ResNet-50 Convolutional Neural Network Based on MRI. *Cancers* **2023**, *15*, 2974. [CrossRef] [PubMed]
15. Foysal, M.; Hossain, A.; Yassine, A.; Hossain, M.S. Detection of COVID-19 Case from Chest CT Images Using Deformable Deep Convolutional Neural Network. *J. Healthc. Eng.* **2023**, *2023*, 4301745. [CrossRef] [PubMed]
16. Fulton, L.V.; Dolezel, D.; Harrop, J.; Yan, Y.; Fulton, C.P. Classification of Alzheimer's Disease with and without Imagery using Gradient Boosted Machines and ResNet-50. *Brain Sci.* **2019**, *9*, 212. [CrossRef]
17. Dos Santos, G.C.; Araujo, A.L.D.; de Amorim, H.A.; Giraldo-Roldan, D.; de Sousa-Neto, S.S.; Vargas, P.A.; Kowalski, L.P.; Santos-Silva, A.R.; Lopes, M.A.; Moraes, M.C. Feasibility study of ResNet-50 in the distinction of intraoral neural tumors using histopathological images. *J. Oral. Pathol. Med.* **2024**, *53*, 444–450. [CrossRef]
18. Aanjankumar, S.; Sathyamoorthy, M.; Dhanaraj, R.K.; Surjit Kumar, S.R.; Poonkuntran, S.; Khadidos, A.O.; Selvarajan, S. Prediction of malnutrition in kids by integrating ResNet-50-based deep learning technique using facial images. *Sci. Rep.* **2025**, *15*, 7871. [CrossRef]
19. Kumar, V.; Prabha, C.; Sharma, P.; Mittal, N.; Askar, S.S.; Abouhawwash, M. Unified deep learning models for enhanced lung cancer prediction with ResNet-50-101 and EfficientNet-B3 using DICOM images. *BMC Med. Imaging* **2024**, *24*, 63. [CrossRef]
20. Kovalyk-Borodyak, O.; Morales-Sánchez, J.; Verdu-Monedero, R.; Sancho-Gomez, J.L. Glaucoma detection: Binocular approach and clinical data in machine learning. *Artif. Intell. Med.* **2025**, *160*, 103050. [CrossRef]
21. Yang, S.; Gong, L.; Zhao, H.; Liu, J.; Chen, X.; Shen, S.; Zhu, X.; Luo, W. Classification of pain expression images in elderly with hip fractures based on improved ResNet50 network. *Front. Med.* **2024**, *11*, 1421800. [CrossRef]

22. Qi, Y.; Hu, Y.; Lin, C.; Song, G.; Shi, L.; Zhu, H. A preoperative predictive model based on multi-modal features to predict pathological complete response after neoadjuvant chemoimmunotherapy in esophageal cancer patients. *Front. Immunol.* **2025**, *16*, 1530279. [CrossRef] [PubMed]
23. Tang, M.; Sugiyama, T.; Takahari, R.; Sugimori, H.; Yoshimura, T.; Ogasawara, K.; Kudo, K.; Fujimura, M. Assessment of changes in vessel area during needle manipulation in microvascular anastomosis using a deep learning-based semantic segmentation algorithm: A pilot study. *Neurosurg. Rev.* **2024**, *47*, 200. [CrossRef] [PubMed]
24. Yogapriya, J.; Chandran, V.; Sumithra, M.G.; Elakkiya, B.; Shamila Ebenezer, A.; Suresh Gnana Dhas, C. Automated Detection of Infection in Diabetic Foot Ulcer Images Using Convolutional Neural Network. *J. Healthc. Eng.* **2022**, *2022*, 2349849. [CrossRef] [PubMed] [PubMed Central]
25. Yap, M.H.; Hachiuma, R.; Alavi, A.; Brüngel, R.; Cassidy, B.; Goyal, M.; Zhu, H.; Rückert, J.; Olshansky, M.; Huang, X.; et al. Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. *Comput. Biol. Med.* **2021**, *135*, 104596. [CrossRef] [PubMed]
26. Cakir, M.; Tulum, G.; Cuce, F.; Yilmaz, K.B.; Aralasmak, A.; Isik, M.İ.; Canbolat, H. Differential Diagnosis of Diabetic Foot Osteomyelitis and Charcot Neuropathic Osteoarthropathy with Deep Learning Methods. *J. Imaging Inform. Med.* **2024**, *37*, 2454–2465. [CrossRef] [PubMed] [PubMed Central]
27. Cecilia-Matilla, A.; Lázaro-Martínez, J.L.; Aragon-Sanchez, J.; Garcia-Morales, E.; Garcia-Alvarez, Y.; Beneit-Montesinos, J.V. Histopathologic characteristics of bone infection complicating foot ulcers in diabetic patients. *J. Am. Podiatr. Med. Assoc.* **2013**, *103*, 24–31.
28. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [CrossRef]
29. Kim, H.E.; Cosa-Linan, A.; Santhanam, N.; Jannesari, M.; Maros, M.E.; Ganslandt, T. Transfer learning for medical image classification: A literature review. *BMC Med. Imaging* **2022**, *22*, 69. [CrossRef]
30. Wu, Y.; Lu, X.; Hong, J.; Lin, W.; Chen, S.; Mou, S.; Feng, G.; Yan, R.; Cheng, Z. Detection of extremity chronic traumatic osteomyelitis by machine learning based on computed-tomography images: A retrospective study. *Medicine* **2020**, *99*, e19239. [CrossRef]
31. Egger, J.; Wild, D.; Weber, M.; Bedoya, C.A.R.; Karner, F.; Prutsch, A.; Schmied, M.; Dionysio, C.; Krobath, D.; Jin, Y.; et al. Studierfenster: An Open Science Cloud-Based Medical Imaging Analysis Platform. *J. Digit. Imaging* **2022**, *35*, 340–355. [CrossRef]
32. Mustapha, B.; Zhou, Y.; Shan, C.; Xiao, Z. Enhanced Pneumonia Detection in Chest X-Rays Using Hybrid Convolutional and Vision Transformer Networks. *Curr. Med. Imaging* **2025**, *21*, e15734056326685. [CrossRef]
33. Rajaraman, S.; Ganesan, P.; Antani, S. Deep learning model calibration for improving performance in class-imbalanced medical image classification tasks. *PLoS ONE* **2022**, *17*, e0262838. [CrossRef]
34. Esposito, C.; Landrum, G.A.; Schneider, N.; Stiefl, N.; Riniker, S. GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. *J. Chem. Inf. Model.* **2021**, *61*, 2623–2640. [CrossRef] [PubMed]
35. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [CrossRef] [PubMed]
36. Aragon-Sanchez, J.; Lipsky, B.A.; Lazaro-Martinez, J.L. Diagnosing diabetic foot osteomyelitis: Is the combination of probe-to-bone test and plain radiography sufficient for high-risk inpatients? *Diabet. Med.* **2011**, *28*, 191–194. [CrossRef] [PubMed]
37. Morales Lozano, R.; Gonzalez Fernandez, M.L.; Martinez Hernandez, D.; Beneit Montesinos, J.V.; Guisado Jimenez, S.; Gonzalez Jurado, M.A. Validating the probe-to-bone test and other tests for diagnosing chronic osteomyelitis in the diabetic foot. *Diabetes Care* **2010**, *33*, 2140–2145. [CrossRef]
38. Segall, G.M.; Nino-Murcia, M.; Jacobs, T.; Chang, K. The role of bone scan and radiography in the diagnostic evaluation of sus-pected pedal osteomyelitis. *Clin. Nucl. Med.* **1989**, *14*, 255–260. [CrossRef]
39. Weinstein, D.; Wang, A.; Chambers, R.; Stewart, C.A.; Motz, H.A. Evaluation of magnetic resonance imaging in the diagnosis of osteomyelitis in diabetic foot infections. *Foot Ankle.* **1993**, *14*, 18–22. [CrossRef]
40. Blume, P.A.; Dey, H.M.; Daley, L.J.; Arrighi, J.A.; Soufer, R.; Gorecki, G.A. Diagnosis of pedal osteomyelitis with Tc-99m HMPAO labeled leukocytes. *J. Foot Ankle Surg.* **1997**, *36*, 120–126. [CrossRef]
41. Enderle, M.D.; Coerper, S.; Schweizer, H.P.; Kopp, A.E.; Thelen, M.H.; Meisner, C.; Pressler, H.; Becker, H.D.; Claussen, C.; Häring, H.U.; et al. Correlation of imaging techniques to histo-pathology in patients with diabetic foot syndrome and clinical suspicion of chronic osteomyelitis. The role of high-resolution ultrasound. *Diabetes Care* **1999**, *22*, 294–299. [CrossRef]
42. Alvaro-Afonso, F.J.; Lazaro-Martinez, J.L.; Garcia-Morales, E.; Garcia-Alvarez, Y.; Sanz-Corbalan, I.; Molines-Barroso, R.J. Cortical disruption is the most reliable and accurate plain radiographic sign in the diagnosis of diabetic foot osteomyelitis. *Diabet. Med.* **2019**, *36*, 258–259. [CrossRef] [PubMed]

43. Ghassemi, N.; Shoeibi, A.; Rouhani, M. Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images. *Biomed. Signal Process. Control.* **2020**, *57*, 101678. [CrossRef]
44. Fu, J.; Yang, Y.; Singhrao, K.; Ruan, D.; Kishan, A.; King, C.; Low, D.; Lewis, J. Deep Learning Approaches for Male Pelvic Synthetic CT Generation Using 2D and 3D Convolutional Neural Networks. *Med. Phys.* **2018**, *45*, E701.