Deep Learning Algorithms for the Automated Classification of Pre-Ulcerative Plantar Calluses and Corns

Krittidej Wittayavanichai
Phillips Exeter Academy
NH, United States
kwittayavanichai@exeter.edu

Nattamon Tanathornkirati
Registered Nurse,
Thammasat Diabetes Center
Pathum Thani, Thailand
tnattamo@tu.ac.th

Kanit Wittayavanichai
Department of Surgery Faculty of Medicine, Thammasat University
Bangkok, Thailand
witkanit@tu.ac.th

Poon Uthaisang
Harrow International School
Bangkok,Thailand
poon_ut@harrowschool.ac.th

Phannapha Yimpray
Registered Nurse,
Thammasat Diabetes Center
Pathum Thani, Thailand
get00707@tu.ac.th

Sarayut Nonsiri
School of Information Technology,
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
*Corresponding author:
sarayut@it.kmitl.ac.th

Taksaporn Wanichapol
Amnuay Silpa School
Bangkok, Thailand
79294@student.amnuaysilpa.ac.th

***Abstract*—Diabetic foot ulcers (DFUs) are a severe complication of diabetes mellitus and a leading cause of non-traumatic lower-extremity amputations. Pre-ulcerative lesions, notably plantar calluses and corns (helomata), serve as critical clinical precursors; however, standardized clinical grading systems and automated diagnostic tools for their early detection remain lacking. To address this gap, we propose a deep learning framework for the automated grading and severity classification of plantar calluses and corns from clinical images. Utilizing a custom dataset of 1,631 images compiled from only public sources, the lesions were categorized into 5 classes: Normal, Light callus, Hard callus, Corn, and Ulcer. To mitigate class imbalance and limited sample size, an augmentation pipeline using geometric transformations, photometric adjustments and non-linear distortion was applied during training. There are 8 pretrained models including ConvNeXtXLarge, NASNetLarge, EfficientNetV2L, ResNet152V2, InceptionResNetV2, VGG19, Xception, and DenseNet201 that were trained and evaluated. As a result, the ConvNeXtXLarge model achieved the highest accuracy of 79.80% and outperformed the other models across all evaluation metrics.**

Keywords—diabetic foot, callus, corn, pre-ulcerative lesion, deep learning, convolutional neural network, image classification

I. INTRODUCTION

Diabetes mellitus (DM) is a rapidly escalating global health challenge, with approximately 537 million documented cases worldwide as of 2021. This population is projected to reach 643 million by 2030 and 783 million by 2045 [1]. Regionally, Thailand faces a significant disease burden, where DM serves as a primary driver of mortality, disability, and heavy resource consumption within the healthcare system.

A critical symptom of DM is peripheral neuropathy, which severely degrades sensory feedback in the lower extremities. This reduced pain sensation and repetitive mechanical stress on the plantar surface leads directly to the formation of pre-ulcerative lesions, primarily calluses and corns. Biomechanical studies have quantified this relationship directly: in neuropathic diabetic feet, callus formation is governed by the ratio of shear to normal stress at prominent loading sites, and an established plantar callus markedly elevates local plantar pressure [2]. Hidden failure points beneath the callus, such as subcallosal hemorrhage and early tissue breakdown, are critical pre-ulcerous indicators. Without intervention, this chain of events may progress to Diabetic Foot Ulcer (DFU) if untreated. The progression to DFU is severe: diabetic patients face a 19–34% lifetime risk of developing them, they are the primary catalyst for non-traumatic lower extremity amputations, and they carry a 5-year mortality rate reaching up to 30% [3][4].

Addressing this pipeline at the pre-ulcer stage, before ulceration develops, is highly effective. Previous studies show that early prevention protocols—such as routine screening and pressure offloading—can reduce amputation rates significantly [5]. The International Working Group on the Diabetic Foot (IWGDF) 2023 prevention guideline explicitly recognizes plantar callus as a key pre-ulcerative lesion and recommends its identification and treatment within a risk-stratified screening framework [3]. Other methods such as education in foot care would also be greatly beneficial [6]. However, current diagnostic systems rely heavily on subjective manual examinations of foot ulcers, an approach explicitly recognized as imprecise and error prone [7]. This analog method introduces high inter-rater variability and creates a severe bottleneck in primary care settings lacking specialized resources.

While Artificial Intelligence (AI) and deep learning have transformed medical imaging, existing computational models for diabetic feet are overwhelmingly reactive. The majority of current deep learning research focuses on post-ulceration tasks, such as wound depth classification, ischemia evaluation, and infection detection, leaving the automated detection of pre-ulcerative lesions largely understudied [8]. The few studies that address earlier disease stages have relied predominantly on plantar thermography rather than standard clinical photographs [9] or have targeted structural deformities such as hallux valgus and hammer toes rather than the keratotic lesions that most directly precede ulceration [10]. To address this gap, this study proposes a deep learning classification framework designed for the automated severity grading of pre-ulcerative diabetic foot lesions using clinical photographs. By utilizing state-of-the-art Convolutional Neural Network (CNN) architectures and expert annotations, this framework has the potential to serve as a preliminary baseline for future screening models to assist clinicians in proactive risk screening.

II. RELATED WORKS

TABLE I. RELATED WORKS COMPARISON

No	Year, Ref.	Subject Matter	Techniques	Dataset	Results
1	2022, [9]	DM severity grading (6 classes) from plantar thermal images	Custom CNN vs. AlexNet	Plantar thermograms	Multi-class grading
2	2026, [10]	Diabetic foot deformities (callus, hallux valgus, hammer toes)	YOLOv5; Convolutional Autoencoder; DenseNet201	Woodlands Health Vascular Registry	Accuracy 88.57%
3	2024, [11]	Diabetic foot ulcers only (DFUs)	EfficientNet B3	Kaggle	Accuracy 98.2%
4	2023, [12]	Diabetic foot ulcers only (DFUs)	Faster R-CNN; Inception-ResNetV2	2 UK hospitals	Sens. 0.92 / Spec. 0.89 (0.93 post-proc.)
5	2024, [13]	Diabetic foot ulcer only (DFUs)	Xception; DenseNet121; ResNet50; InceptionV3; MobileNetV2; XAI	Kaggle	Accuracy 98.75%

Although the computational classification of Diabetic Foot Ulcers (DFUs) is well-established with high classification accuracy and various datasets that have been used to this point [11-13], pre-ulcerative lesions like calluses are rarely prioritized, with only a marginal number of related works including them within their diagnostic frameworks [9][10], as presented in Table I. Among these, Muralidhara et al. graded diabetic-foot severity into multiple classes from plantar thermal images but did not use standard RGB clinical photographs [9], whereas Teoh et al. detected structural deformities—callus, hallux valgus, and hammer toes—from foot photographs without grading callus severity [10]. Granular classification schemas for these pre-ulcerative states are even less common. However, a few notable exceptions have successfully stratified calluses into light, moderate, and heavy categories, providing the primary theoretical inspiration for the methodology [14]. To our knowledge, no prior work has graded plantar callus and corn severity across a Normal–Light callus–Hard callus–Corn–Ulcer continuum from clinical photographs using a head-to-head benchmark of modern CNN backbones, which is the gap this study addresses.

III. METHODOLOGY

The proposed research methodology is designed to evaluate the performance of pre-trained models for the automated classification of pre-ulcerative plantar calluses and corns. The experimental process comprises distinct phases: data collection, dataset partitioning, data augmentation, model training, and performance evaluation.

A. Dataset Collection

A dataset of 1,631 images was compiled from publicly available sources on the internet via web scraping, using the class names as search keywords. Images containing text, low-quality photos, and irrelevant results were removed. The remaining images were independently classified by three clinicians (one surgeon and two specialist nurses) into five categories: Normal, Light callus, Hard callus, Corn, and Ulcer (Table II). No patients were recruited and no identifiable private information was collected; all images were already in the public source.

TABLE II. DATASET CLASS DEFINITIONS

No	Name	Characteristics	Treatment	Image
0	Normal	Normal plantar skin	Preventive education, daily self-inspection, appropriate footwear, annual diabetic foot screening (once a year)	
1	Light callus	Superficial diffuse hyperkeratosis	Self-care, moisturizer, pressure reduction, monitoring every 6–12 months	
2	Hard callus	Thick, indurated keratotic plaque, fissure	Debridement, pressure reduction, offloading, custom footwear (nurse, podiatry-trained clinician, physiotherapist, physician)	
3	Corn	Nucleated keratin plug	Enucleation, focal offloading pain reduction, gait correction (physician or specialized diabetic foot nurse)	
4	Ulcer	Open wound	Urgent wound care and limb preservation (multidisciplinary diabetic foot team — critical care pathway)	

B. Data Setup and Augmentation

After labeling, the dataset was partitioned into training (80%), validation (10%), and testing (10%) subsets. To mitigate class imbalance, particularly in class 3, an augmentation strategy was implemented using the Python Augmentor package [15]. Minority classes were upsampled to match the majority class count, with image generation strictly capped at 2.5 times the original class size to prevent overfitting. To further enhance the training data, a stochastic transformation pipeline was applied to the original images. This pipeline incorporated a combination of geometric transformations—such as continuous and 90-degree random rotations, spatial flips, and zoom cropping—along with photometric adjustments for brightness and contrast. Additionally, random elastic grid distortions were utilized to introduce non-linear variations, simulating the diverse morphological presentations of keratotic lesions and ensuring generalized feature extraction by the model [16]. The augmentation was only performed on the training set after the partitioning to prevent data leakage.

TABLE III. DATASET PARTITIONING

No	Class	Internet	Training set (No augmentation)	Training set (Augmentation)
0	Normal	276	220	395
1	Light callus	328	262	395
2	Hard callus	494	395	395
3	Corn	149	119	297
4	Ulcer	384	307	395
		1,631	1,303	1,877

C. Model Training

This study evaluates the performance of eight deep learning models including ConvNeXtXLarge, NASNetLarge, EfficientNetV2L, ResNet152V2, InceptionResNetV2, VGG19, Xception, and DenseNet201 for the automated classification of pre-ulcerative plantar calluses and corns [17-25]. Although these models were pre-trained on the large ImageNet dataset [26], they still needed fine-tuning to recognize the specific features of this new dataset. Therefore, a transfer learning approach was applied to reuse the models' existing knowledge. The early layers were frozen to retain their foundational learned representations of low-level features, such as edges, corners, and textures. For the final layers, the designed fully connected layers were unfrozen and trained on this new dataset.

For the parameter settings, as presented in Table IV, the models were trained using the Adam optimizer with a constant learning rate of 0.0001. To prevent overfitting, an early stopping mechanism was set to terminate training if the validation loss did not improve for 10 consecutive epochs. Additionally, a model checkpoint callback was used to save only the network weights that achieved the lowest validation loss for use in the final testing phase.

TABLE IV. DATASET PARTITIONING

Parameters	Value
No. of Classes	5
Batch size	32
Optimizer	Adam
Learning rate	0.0001
Loss function	Sparse Categorical Cross-entropy
Epochs	100

D. Evaluation Metrics

To evaluate the classification performance of each pretrained model, standard metrics were derived from the empirical outputs of the testing set's confusion matrix.

Accuracy represents the overall proportion of correct classifications made by the model across all categories.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

Precision calculates the accuracy of the positive predictions, effectively measuring the model's ability to minimize false positives, which is critical when diagnostic resources are limited.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{2}$$

Recall also known as sensitivity, measures the system's ability to identify all actual positive cases, minimizing false negatives to ensure severe clinical conditions are not overlooked.

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{3}$$

Finally, the F1-Score is calculated as the harmonic mean of Precision and Recall. This metric provides a balanced and highly robust evaluation of model performance, which is particularly essential for this study given the inherent class imbalances between common conditions and less frequent lesions, such as corns.

$$\text{F1-Score} = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{4}$$

Due to the class imbalance in the dataset, performance metrics (precision, recall, and F1-score) are reported using weighted averages to accurately reflect the model's performance relative to the proportion of instances in each class.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The software environment was configured with Python 3.11, utilizing TensorFlow 2.21 and Keras as the primary deep learning frameworks.

As presented in Table V, the experimental results for the classification of pre-ulcerative plantar calluses and corns demonstrate that ConvNeXtXLarge achieved the highest values across all reported metrics under both the baseline and augmented training conditions when evaluated on the same held-out test set. This reflects other findings where ConvNeXt variants have equaled or beaten vision transformers [26, 27], an advantage usually credited to a convolutional backbone reworked with design choices borrowed from transformers. VGG19 and DenseNet201, actually lost several points once the training set was enlarged; only mid-tier networks such as ResNet152V2 and Xception improved [28].

TABLE V. MODEL PERFORMANCE COMPARISON

Model	Accuracy		Precision		Recall		F1-Score	
	Baseline	Augmented	Baseline	Augmented	Baseline	Augmented	Baseline	Augmented
EfficientNetV2L	0.702	0.702	0.705	0.699	0.702	0.702	0.696	0.692
ResNet152V2	0.726	0.750	0.735	0.773	0.726	0.750	0.720	0.744
InceptionResNetV2	0.720	0.726	0.714	0.714	0.720	0.726	0.711	0.711
ConvNeXtXLarge	**0.798**	**0.798**	**0.793**	**0.799**	**0.798**	**0.798**	**0.793**	**0.797**

TABLE V. MODEL PERFORMANCE COMPARISON

Model	Accuracy		Precision		Recall		F1-Score	
	Baseline	Augmented	Baseline	Augmented	Baseline	Augmented	Baseline	Augmented
NASNetLarge	0.708	0.714	0.700	0.713	0.708	0.714	0.701	0.707
VGG19	0.786	0.732	0.792	0.725	0.786	0.732	0.784	0.717
Xception	0.756	0.762	0.751	0.754	0.756	0.762	0.751	0.755
DenseNet201	0.780	0.744	0.781	0.739	0.780	0.744	0.778	0.738

Fig. 1 illustrates the confusion matrix of the best performing model overall. The classification involved five specific categories: Normal, Light Callus, Hard Callus, Corn, and Ulcer, achieving an overall accuracy of 79.8%.

Normal and Ulcer are well distinguished (⩾93% recall; Ulcer at 100% precision), while Corn achieves 81% recall. Most errors cluster at the light-callus/hard-callus boundary, where the two are mutually confused (11 and 10 cases), dropping light-callus recall to 58.8%. The model can identify callus-related lesions but struggles to grade their severity, suggesting that the distinction is graded rather than categorical.

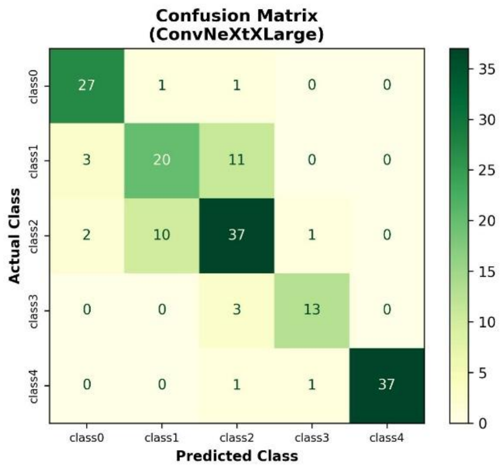

Fig. 1. Confusion matrix for the ConvNeXtXLarge trained on augmented data.

V. CONCLUSION

This study demonstrates the potential of deep learning for the automated classification of pre-ulcerative plantar calluses and corns. The comparative analysis revealed that ConvNeXtXLarge achieved the highest accuracy of 0.798. While promising, further performance improvement and clinical validation are required before use in clinical practice. Future work will focus on expanding the dataset and applying advanced data augmentation techniques to improve generalization. In addition, exploring the newer deep learning methods may further improve classification performance.

A primary contribution of this study is the introduction of a specialized dataset for the automated classification of diabetic foot lesions. This dataset comprises 1,631 images categorized into five clinical categories: Normal, Light Callus, Hard Callus, Corn, and Ulcer. Furthermore, a comprehensive comparative analysis of eight advanced deep learning architectures was conducted to evaluate their classification performance on this dataset. These findings provide a practical benchmark for future research on deep learning-based diabetic foot lesion classification.

REFERENCES

[1] International Diabetes Federation. IDF Diabetes Atlas [Internet]. 11th ed. Brussels: IDF; 2024 [cited 2025 Jan 1]. Available from: https://diabetesatlas.org/

[2] Amemiya A, Noguchi H, Oe M, Takehara K, Ohashi Y, Suzuki R, et al. Shear stress–normal stress (pressure) ratio decides forming callus in patients with diabetic neuropathy. J Diabetes Res. 2016;2016:3157123. https://doi.org/10.1155/2016/3157123

[3] Bus SA, Sacco ICN, Monteiro-Soares M, Raspovic A, Paton J, Rasmussen A, et al. Guidelines on the prevention of foot ulcers in persons with diabetes (IWGDF 2023 update). Diabetes Metab Res Rev. 2024;40(3):e3651. https://doi.org/10.1002/dmrr.3651

[4] Armstrong DG, Tan TW, Boulton AJM, Bus SA. Diabetic foot ulcers: a review. JAMA. 2023;330(1):62–75. https://doi.org/10.1001/jama.2023.10578

[5] Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. JAMA. 2005;293(2):217–28. https://doi.org/10.1001/jama.293.2.217

[6] Ebadi Fardazar F, Tahari F, Solhi M. Empowerment of type 2 diabetic patients visiting Fuladshahr diabetes clinics for prevention of diabetic foot. Diabetes Metab Syndr. 2018;12(6):853–8. https://doi.org/10.1016/j.dsx.2018.04.034

[7] Jishnu P, Sai Kumar BK, Jayaraman S. Automatic foot ulcer segmentation using conditional generative adversarial network (AFSegGAN): a wound management system. PLOS Digit Health. 2023;2(11):e0000344. https://doi.org/10.1371/journal.pdig.0000344

[8] Tulloch J, Zamani R, Akrami M. Machine learning in the prevention, diagnosis and management of diabetic foot ulcers: a systematic review. IEEE Access. 2020;8:198977–9000. https://doi.org/10.1109/ACCESS.2020.3035327

[9] Muralidhara S, Lucieri A, Dengel A, Ahmed S. Holistic multi-class classification and grading of diabetic foot ulcerations from plantar thermal images using deep learning. Health Inf Sci Syst. 2022;10(1):21. https://doi.org/10.1007/s13755-022-00194-8

[10] Teoh KL, Hari KS, Ng LL, Lo ZJ. Validation of a trained AI imaging model for detecting diabetic foot deformities. Int Wound J. 2026;23(2):e70821. https://doi.org/10.1111/iwj.70821

[11] Muthuraja M, Shanthi N, Aravindhraj N, Vilasini V, Saran Kumar A. Diabetic foot ulcer detection using deep learning approaches. In: 2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC); 2024. IEEE; 2024. p. 1–6. https://doi.org/10.1109/ICECCC61783.2024.10893620

[12] Cassidy B, Yap MH, Pappachan JM, Ahmad N, Haycocks S, O'Shea C, et al. Artificial intelligence for automated detection of diabetic foot ulcers: a real-world proof-of-concept clinical evaluation. Diabetes Res Clin Pract. 2023;205:110951. https://doi.org/10.1016/j.diabres.2023.110951

[13] Biswas S, Mostafiz R, Paul BK, Uddin KMM, Hadi MA, Khanom F. DFU_XAI: a deep learning-based approach to diabetic foot ulcer detection using feature explainability. Biomed Mater Devices. 2024;2(2):1225–45. https://doi.org/10.1007/s44174-023-00150-z

[14] Tollafield DR. Clinical photographic observation of plantar corns and callus associated with a nominal scale classification and inter-observer reliability study in a student population. J Foot Ankle Res. 2017;10:45. https://doi.org/10.1186/s13047-017-0225-2

[15] Bloice MD, Stocker C, Holzinger A. Augmentor: an image augmentation library for machine learning. J Open Source Softw. 2017;2(19):432. https://doi.org/10.21105/joss.00432

[16] [16] Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. In: Seventh International Conference on Document Analysis and

Recognition (ICDAR); 2003 Aug 3-6; Edinburgh, UK. IEEE; 2003. p. 958–63. https://doi.org/10.1109/ICDAR.2003.1227801

[17] Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18-24; New Orleans, LA. IEEE; 2022. p. 11976–86. https://doi.org/10.1109/CVPR52688.2022.01167

[18] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018 Jun 18-23; Salt Lake City, UT. IEEE; 2018. p. 8697–710. https://doi.org/10.1109/CVPR.2018.00907

[19] Tan M, Le QV. EfficientNetV2: smaller models and faster training. In: Proceedings of the 38th International Conference on Machine Learning (ICML); 2021 Jul 18-24; Virtual. PMLR; 2021. p. 10096–106. https://doi.org/10.48550/arXiv.2104.00298

[20] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer Vision – ECCV 2016. Lecture Notes in Computer Science, vol. 9908. Cham: Springer; 2016. p. 630–45. https://doi.org/10.1007/978-3-319-46493-0_38

[21] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the impact of residual connections on learning [Preprint]. arXiv. 2016 Feb 23 [cited 2025 Jan 1]. Available from: https://arxiv.org/abs/1602.07261

[22] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR); 2015 May 7-9; San Diego, CA. 2015. https://doi.org/10.48550/arXiv.1409.1556

[23] Chollet F. Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI. IEEE; 2017. p. 1800–7. https://doi.org/10.1109/CVPR.2017.195

[24] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21-26; Honolulu, HI. IEEE; 2017. p. 2261–9. https://doi.org/10.1109/CVPR.2017.243

[25] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2009 Jun 20-25; Miami, FL. IEEE; 2009. p. 248–55. https://doi.org/10.1109/CVPR.2009.5206848

[26] Vieira J, Mendonça F, Morgado-Dias F. Deep learning approaches for skin lesion detection. Electronics. 2025;14(14):2785. https://doi.org/10.3390/electronics14142785

[27] Baykal Kablan E, Ayas S. Skin lesion classification from dermoscopy images using ensemble learning of ConvNeXt models. Signal Image Video Process. 2024;18(8):6353–61. https://doi.org/10.1007/s11760-024-03321-y

[28] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6:60. https://doi.org/10.1186/s40537-019-0197-0