Cognitive Computation
https://doi.org/10.1007/s12559-024-10267-3

Diabetic Foot Ulcer Detection: Combining Deep Learning Models for Improved Localization

Rusab Sarmun[1] · Muhammad E. H. Chowdhury[2] · M. Murugappan[3,4,5] · Ahmed Aqel[6] · Maymouna Ezzuddin[2] · Syed Mahfuzur Rahman[7] · Amith Khandakar[2] · Sanzida Akter[8] · Rashad Alfkey[9] · Anwarul Hasan[3]

Received: 29 March 2023 / Accepted: 3 March 2024

Abstract
Diabetes mellitus (DM) can cause chronic foot issues and severe infections, including Diabetic Foot Ulcers (DFUs) that heal slowly due to insufficient blood flow. A recurrence of these ulcers can lead to 84% of lower limb amputations and even cause death. High-risk diabetes patients require expensive medications, regular check-ups, and proper personal hygiene to prevent DFUs, which affect 15–25% of diabetics. Accurate diagnosis, appropriate care, and prompt response can prevent amputations and fatalities through early and reliable DFU detection from image analysis. We propose a comprehensive deep learning-based system for detecting DFUs from patients' feet images by reliably localizing ulcer points. Our method utilizes innovative model ensemble techniques—non-maximum suppression (NMS), Soft-NMS, and weighted bounding box fusion (WBF)—to combine predictions from state-of-the-art object detection models. The performances of diverse cutting-edge model architectures used in this study complement each other, leading to more generalized and improved results when combined in an ensemble. Our WBF-based approach combining YOLOv8m and FRCNN-ResNet101 achieves a mean average precision (mAP) score of 86.4% at the IoU threshold of 0.5 on the DFUC2020 dataset, significantly outperforming the former benchmark by 12.4%. We also perform external validation on the IEEE DataPort Diabetic Foot dataset which has demonstrated robust and reliable model performance on the qualitative analysis. In conclusion, our study effectively developed an innovative diabetic foot ulcer (DFU) detection system using an ensemble model of deep neural networks (DNNs). This AI-driven tool serves as an initial screening aid for medical professionals, augmenting the diagnostic process by enhancing sensitivity to potential DFU cases. While recognizing the presence of false positives, our research contributes to improving patient care through the integration of human medical expertise with AI-based solutions in DFU management.

Keywords Diabetic foot ulcer (DFU) · Weighted bounding box fusion (WBF) · Machine learning · Deep learning · Diabetic Foot Ulcer Challenge 2020 (DFUC2020)

Introduction

Diabetes mellitus is a chronic disease characterized by an abnormally high blood sugar level, leading to serious and sometimes life-threatening complications such as lower limb amputations, cardiovascular issues, loss of vision, and renal impairment [1]. The International Diabetes Federation reported that about 9.3% of the world's total population is affected by diabetes and the number is predicted to increase up to 10.2% by the year 2030 [2]. A primary complication of diabetes is neuropathy, particularly in the feet, which can lead to incurable infections. Diabetic individuals often struggle with healing foot ulcers due to impaired blood circulation [3, 4], which can exacerbate infections, potentially necessitating amputation in the long run. The recurrence rate of diabetic foot ulcers (DFUs) is also extremely high at roughly 40% after the first year and 60% within three years after onset [5, 6]. In the United States alone, almost one million people with diabetes undergo amputations annually due to inadequate diagnosis and management of DFUs [7]. Such amputation wounds are again susceptible to complications in addition to having a negative influence on the quality of life [8]. The treatment and care of an advanced DFU patient are difficult and costly [9]. Effective treatment of DFU requires attentive screening and documentation. Consequently, it is essential to discover a reliable way for the early detection and regular screening of DFU so that they may be treated quickly and cost-effectively before progressing to the next stage.

Extended author information available on the last page of the article

Published online: 01 April 2024

In recent years, artificial intelligence (AI)-based computer-assisted diagnosis (CAD) has been gaining popularity for a wide range of diseases due to the development and effectiveness of artificial neural networks (ANNs) and deep learning (DL) frameworks. AI-based applications are a crucial tool in assisting overworked medical professionals to promote better practices. These automate repetitive procedures by offering decision help at the point of care with swift and definite detections of negative changes in the course of wound healing. There have been several attempts to diagnose DFU with artificial intelligence-based techniques since 2015 [10]. Both traditional machine learning (ML) and computer vision (CV) techniques were utilized at the same time to analyze DFU images [11, 12]. With recent advancements in DL techniques in the CV domain, DFU researchers also focused on the use of DL in DFU for instance Goyal et al. proposed automated segmentation in 2017 [13]. Since then, several investigations have tried to diagnose DFU from the planar thermograms by detecting hot zones that might be an indicator of tissue injury or inflammation [14–16]. In the meantime, some researchers have done substantial studies on classifying, detecting, and segmenting foot images to detect DFU. However, with the lack of large and properly annotated datasets, early detection of DFU cases remains a challenging problem. The researchers could only achieve up to 0.74 mean average precision (mAP) in detecting DFU cases on DFUC2020 which is one of the popular datasets in this domain [10].

Das et al. [17] have put forward a stacked parallel convolution layer-based custom model called DFU SPNet to classify normal and abnormal DFU skin from foot images. The DFU SPNet model is made up of three blocks of parallel convolution layers, each of which has a variety of kernel sizes to extract both local and global features. The study also focused on exploring multiple optimizers and learning rate combinations ultimately achieving an area under curve (AUC) score of 97.4% [17]. Alzubaidi et al. proposed a novel network for the automated classification of DFU images called DFU QUTNet [18]. This network was built to increase the network breadth while preserving a relatively good depth compared to other modern networks. This helps gradient propagation and avoids the complexity of adding extra layers to conventional CNN networks [18]. Yet, their research primarily addresses the issue as a classification problem, which limits its capacity to accurately pinpoint the exact location of ulcers. Another study devised a novel method of capturing DFU images consistently using a mirrored capture box. The DFU regions were identified using cascaded two-stage support vector classification, followed by performing segmentation and feature extraction using a two-stage super-pixel classification method [19]. Recently, four different types of super-resolution tools (super-resolution using a generative adversarial network (SRGAN), enhanced deep residual networks (EDSR), enhanced super-resolution generative adversarial networks (ESRGAN), and image super-resolution (ISR)) were used to enhance the resolution of the DFU images in the DFU2020 challenge dataset and to detect the DFU [56]. Thotad et al. (2022) implemented a system based on EfficientNet to classify normal and abnormal DFU skin. This method triumphed over several models like DFU-Net [20], VGG16 [21], and GoogleNet [22] in precision, recall, and $F1$ score. However, the total number of images in the dataset used here was only 855 which is not significant enough to develop a robust model to detect DFU [23].

Early studies focused on DFU detection using deep neural network (DNN), and the researchers primarily used the DFU Challenge dataset to develop AI models. There are only a few studies that describe the localization of DFUs using DNN along with DFU detection in the literature. Clinical experts must be able to identify the severity of DFU based on localization information to provide a proper course of treatment in clinical practice. In fact, it is also highly useful to develop a remote healthcare system or software prototype for DFU management. Goyal et al. (2019) conducted a thorough study on real-time DFU localization for mobile devices. The study used two-tier transfer learning on a multitude of deep learning models including SSD MobileNet [24], Faster RCNN Inception-v2, and RFCN-ResNet101 [25]. Faster RCNN-Inceptionv2 reached the highest mAP of 0.918. Then, the model was implemented for real-time detection via an Android app and an NVIDIA Jetson TX2 module. It was found that the models used in this study failed to accurately predict small ulcer points, and no further steps were taken to combine the predictions to improve accuracy [26]. In addition to this study, other studies have demonstrated that mobile devices can capture images of feet and identify DFU cases accurately. According to Yap et al., the FootSnap application was developed to monitor diabetic feet using an iPad. A high degree of inter- and intra-operator reliability was shown when both diabetic feet (30 images) and non-diabetic feet (30 images) were analyzed by two different operators on different days [11, 26]. Yap et al. (2021) conducted a comprehensive study of the DFUC2020 dataset with a range of state-of-the-art deep learning networks and also attempted to ensemble them for better results. A higher $F1$ score was obtained with one of the ensemble combinations; however, the mAP decreased. Yet, the deformable convolution [27] reported a maximum mAP, which is a variant of Faster RCNN [28].

While most previous methods in this field have concentrated on developing novel capture tools, they lack the necessary sensitivity for medical diagnosis, which is crucial because missing positive cases can have serious consequences. The other notable research gaps in this field include the following: (i) a very limited number of studies have addressed localization of DFU, (ii) ensemble classifiers

based on DNNs have not been explored in DFU detection and localization, (iii) the use of external datasets to validate the earlier work methodology is highly limited, and (iv) there is a high level of computational complexity associated with most of the earlier work methodologies. In this study, we utilize ensemble-based DNNs to achieve a high level of sensitivity in detecting DFU cases, a critical aspect in medical scenarios for the patient's well-being. Our adopted bounding box detection strategy can ease the burden of the clinical expert in accurately pinpointing DFU regions. The innovative AI algorithms used in our research can help clinical experts with early diagnosis, improved treatment planning, reduced complications, and enhanced patient care. It will also aid remote health monitoring of patients in a home environment. Active observation outside of the hospital can reduce healthcare systems' resources in addition to lowering patient risk [29, 30]. This point is of the utmost importance in the context of the COVID-19 pandemic, as COVID infection correlates with more severe outcomes for diabetic patients. Therefore, minimizing diabetic patients' exposure to clinical settings is vital for their health. To this end, our study aims to elevate DFU patient care by employing advanced ensemble-based detection frameworks for ensuring dependable and accurate diagnostic solutions. The primary aim of the project is to provide a primary screening solution to aid medical professionals in rapid diagnosis.

In this paper, the following major contributions have been made:

(i) We have employed various state-of-the-art object detection models to detect DFU from foot images, leveraging the unique feature extraction capabilities of each architecture to identify a wide range of DFU cases.
(ii) Our adopted ensemble methods further enhance prediction accuracy by strategically merging the detection outcomes in a weighted combination.
(iii) Our designed post-processing step reduces overlapping bounding boxes with an area-to-overlap ratio greater than 0.8 threshold. This mitigates the redundant detections generated by the ensemble methods to improve the overall performance.
(iv) The DFUC2020 dataset [13], comprising over 2000 images, is used to train and develop our models. We utilize the transfer learning approach to enhance the network training for enabling effective model development even with a smaller dataset.
(v) An independent test is conducted on a new collection of 506 DFU images from IEEE DataPort [51] to evaluate the model's generalization capability across distinct datasets. Our qualitative analysis shows the model's remarkable adaptability to a wide spectrum of patient data, emphasizing its effectiveness for real-world applicability.

In this paper, the content is divided into four sections. The cutting-edge object detection models used in this study are discussed in the second section. The experimental methodology of our research is discussed in detail in the third section. The fourth section presents the results of the study and analyzes the improvement of our proposed detection systems with baseline detection models. Limitations and future scopes of our study are discussed in the fifth section, and the sixth section concludes the paper.

State-of-the-Art Models

YOLOv5

YOLOv5 is a state-of-the-art object detection framework that is capable of real-time detection. It improves upon its predecessors by reducing parameters and FLOPS (floating-point operations per second), thus improving inference speed and performance as well as reducing the model size. This is achieved through implementing the CSPDarknet backbone by incorporating CSPNet (cross-stage partial network) [31] into Darknet.

YOLOv5 also incorporates PANet (path aggregation network) [32] which implements a novel FPN(feature pyramid network) with an improved bottom-up path boosting propagation of low-level features. Concurrently, adaptive feature pooling, which connects the feature grid and all feature levels, is employed to ensure that important information in each feature level propagates straight to the subsequent subnetwork. PANet is able to optimize the utilization of precise localization signals in lower layers, which improves the localization of the object. The head of the YOLO model generates 3 distinct sizes of feature maps [33] providing the model the ability to effectively handle objects of different sizes. It uses the Binary Cross-Entropy with Logistic Loss (BCELL) for the calculation of the class and object losses. It creates more than one prediction bounding box that is further eliminated via NMS (non-maximum suppression) to solve the overlapping issue [34].

YOLOv7

YOLOv7 is currently the latest edition of YOLO by the original author and is currently one of the best object detection models available in terms of both inference speed and performance [35]. This iteration of YOLO introduces a number of architectural modifications designed to improve detection speed and precision. In terms of the backbone, YOLOv7 departs from its predecessors; rather than employing the Darknet, an extended efficient layer aggregation network (E-ELAN) is deployed as the computing block for the backbone. The idea of E-ELAN is built on the

usage of expand, shuffle, and merge cardinality to constantly improve the network's learning ability while preserving the gradient path. YOLOv7 uses gradient flow propagation channels to identify the model segments (modules) that need re-parameterization. The head component of the design is based on the notion of multiple heads. Thus, the lead head is responsible for the final classification, while the auxiliary heads aid in the training of the intermediary layers [36]. The architecture for YOLOv7 is shown in Fig. 1.

YOLOv8

YOLOv8 is the latest YOLO model released by Ultralytics [38]. This is actually their third YOLO model after they released YOLOv3 and YOLOv5 previously. YOLOv8 can be used for object detection, image segmentation, and classification. It also boasts higher mAP score on the COCO dataset [39] outperforming the previous versions of YOLO. YOLOv8, with its improved design, is able to achieve greater performance with fewer parameters. It incorporates advanced loss functions such as CIoU and DFL for more precise bounding box calculations and employs binary cross-entropy for determining classification loss. This results in significantly better performance, particularly in identifying smaller objects [40]. In the past, the main component of the YOLO architecture's backbone relied solely on the output from the final bottleneck layer. However, in the improved C2f block, it now concatenates outputs from all bottleneck layers. This enhancement allows the network to tap into and leverage information from various stages, resulting in a more robust and detailed information flow. Additionally, anchors are absent in the YOLOv8 model. This suggests that the prediction is made based on an item's center rather than how far away from a known anchor box it is. Because they may only accurately represent the distribution of boxes in the desired benchmark and not the unique dataset, early YOLO models' anchor boxes were infamously hard to get right. Anchor-free detection reduces the number of box predictions, which speeds up the process, and non-maximum suppression (NMS), a difficult post-processing step, which filters through potential detections after inference. YOLOv8 also introduces new convolution blocks changing one of the core building blocks and replacing the 6×6 convolution with 3×3. YOLOv8 uses more augmentation techniques than the previous versions while training to make the models more robust. [41] A visualization of YOLOv8's architecture is shown in Fig. 2.

Faster RCNN-ResNet101

Faster RCNN [43] consists of two modules. The first module is a deep fully connected CNN that suggests regions, and the second module is the detector that employs the suggested regions. ResNet101 [44] has been implemented as the feature map extractor. A region proposal network (RPN) takes the feature map as input and generates a series of rectangular bounding boxes, each with its own objectness score as output. NMS (non-maximum suppression) is used to eliminate the extra bounding boxes based on score. The architecture for the model is provided in the Fig. 3.

EfficientDet

EfficientDet is a model that builds upon the principles of conventional single-stage detectors, similar to models like YOLO or SSD, which perform object detection in a single pass through the network [45]. It is based on the EfficientNet model. A distinguishing characteristic of the EfficientDet-D1 [46] model is the inclusion of an enhanced version of the feature pyramid network (FPN), known as a bi-directional feature pyramid network (BiFPN). Traditional FPNs in object detection models are used to process feature maps at different scales, allowing the models to detect objects of various sizes. The BiFPN takes this a step further by facilitating more efficient and effective integration of these multi-scale features. It does this by allowing information to flow in both directions (top-down and bottom-up) across the pyramid levels, which results in a more refined feature representation. In addition to the BiFPN, EfficientDet-D1 employs separate networks for class prediction and bounding box prediction. The class network focuses on determining the category of each detected object (like a person, car, and dog), while the box network is dedicated to predicting the precise location and size of each object's bounding box.

Experimental Methodology

This section discusses the Diabetic Foot Ulcer Challenge 2020 (DFUC2020) Dataset [47] and the IEEE DataPort Diabetic Foot Dataset [48]. We will then discuss the experimental steps involved in implementing our proposed detection system.

Dataset Description

(a) Diabetic Foot Ulcer Challenge 2020 (DFUC2020) Dataset

The training segment of this Challenge dataset consists of 2000 images with DFU, and the testing segment is not publicly accessible. The images were captured at a distance of around 30–40 cm with an aperture setting of f/2.8 in close-up mode. Images were captured with three different digital cameras: Kodak DX4530, Nikon D3300, and Nikon COOLPIX P100. Depending on the healing stage

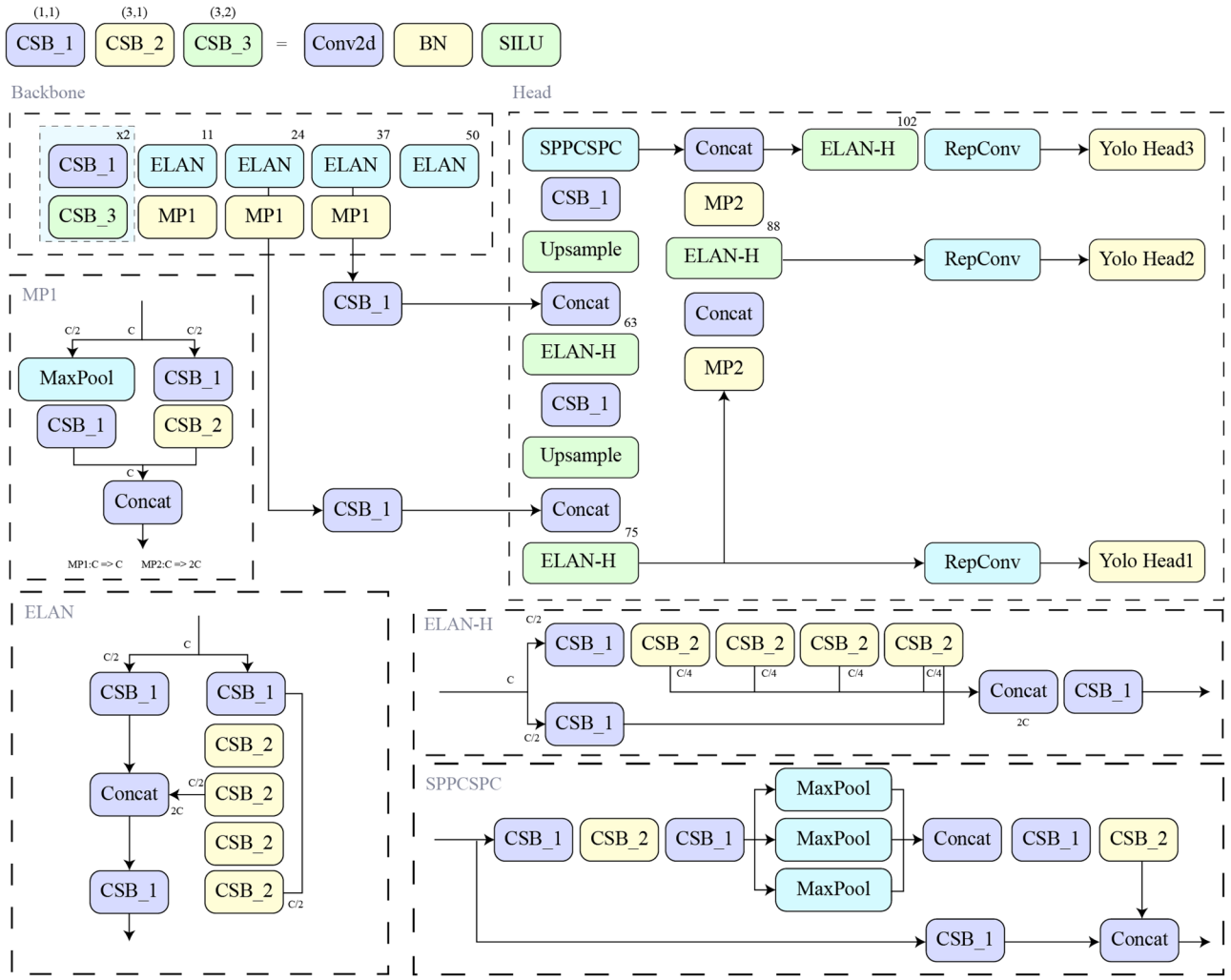

Fig. 1 YOLOv7 architecture [37]

of the ulcer, each image may contain multiple instances. Often, the same foot is photographed from different angles, backgrounds, and lighting conditions. There are a variety of ethnicities represented in the dataset but white is dominant with 1987 cases compared to only 13 non-white cases. The images were annotated by healthcare professionals with more than 5 years of experience treating and managing DFU.

(b) IEwEE DataPort Diabetic Foot Dataset

A total of 506 diabetic foot images are included in this dataset [51]. Most of the images in this dataset were taken with an L-shaped ruler measuring the wound size. Figure 4 shows sample images from both datasets.

Experimental Steps

DFUC2020 consisted of 2000 images split in the ratio of 80:10:10 among the training, validation, and test sets. As a result, the training set contained 1600 (80%) images, the validation set contained 200 (10%) images, and the test set contained 200 (10%) images. According to the three sets, there were 2010, 244, and 242 ulcer instances, respectively. Several deep learning-based object detection networks were then trained on this dataset to develop the models, including YOLOv5, YOLOv7 [35], YOLOv8 [38], EfficientDet [45], and Faster R-CNN [43]. The experimental setup on Google Colab utilized an NVIDIA Tesla T4 GPU with 15 GB of memory, complemented by a dual-core Intel Xeon CPU running at 2.00 GHz, and 26 GB of RAM. The software environment for these experiments included Python 3.9.16 and PyTorch 1.13.

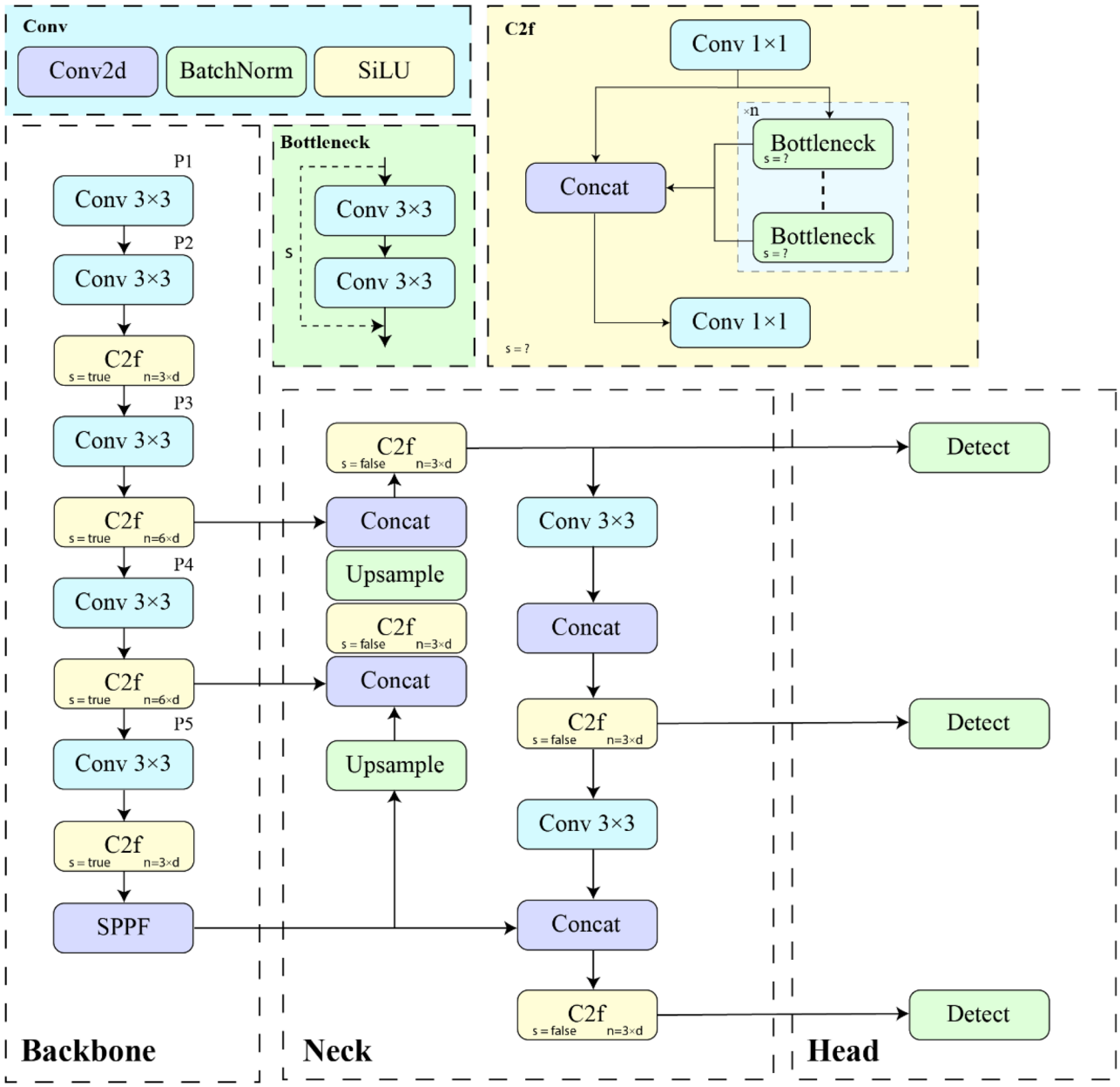

Fig. 2 YOLOv8 architecture

Evaluation Metrics

(a) Several performance metrics were used to evaluate the models, including precision, recall, $F1$ score, and mAP. As a rule of thumb, the Intersection over Union (IoU) for a predicted bounding box with the ground truth must be greater than or equal to 0.5 to be considered a true positive. $F1$ score is the harmonic mean of precision and recall, as it gives a more appropriate evaluation of the model's predictive performance in terms of both false negatives and false positives.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

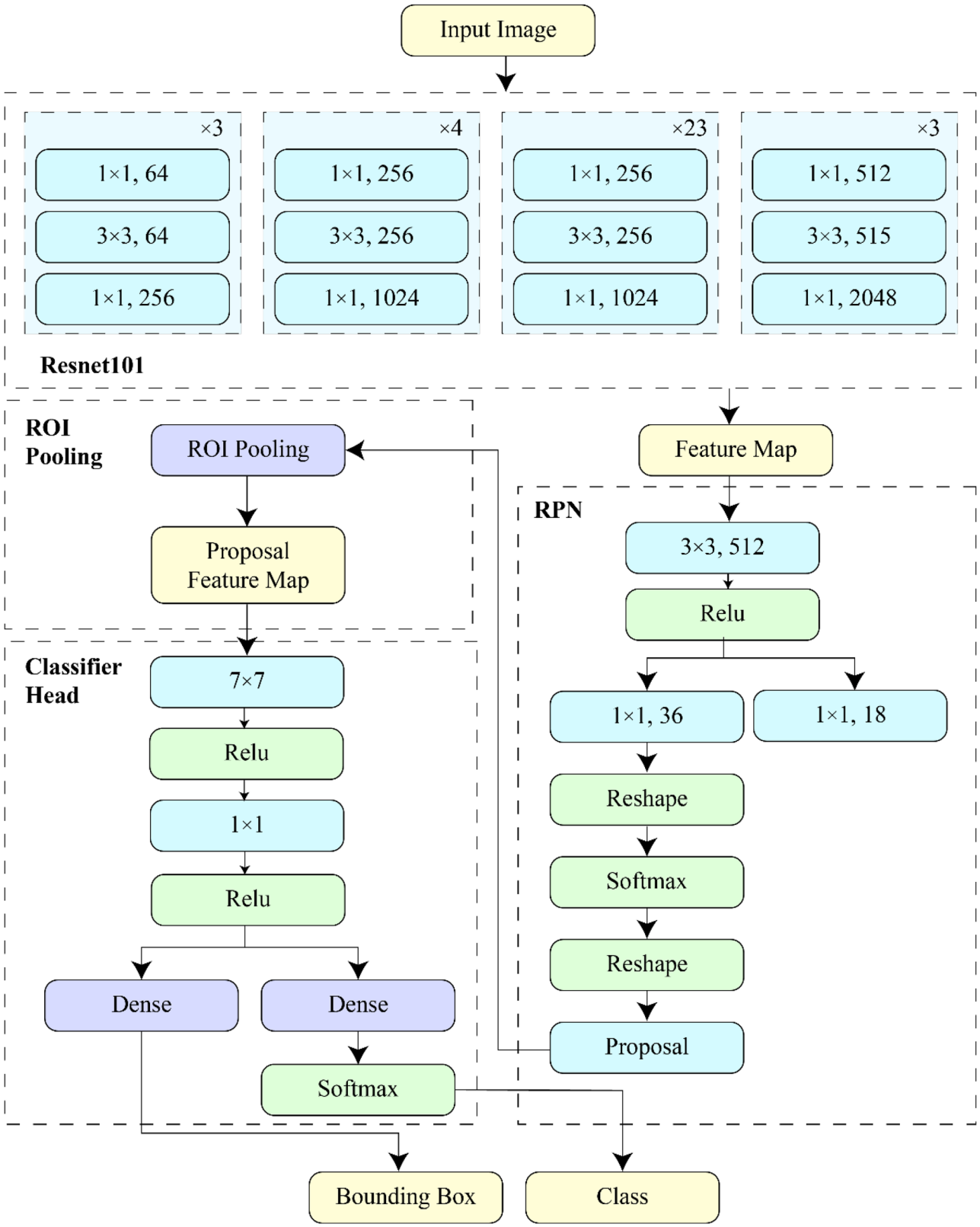

Fig. 3 Faster RCNN Resnet101 architecture [42]

Fig. 4 Sample images from the datasets: **a** Diabetic Foot Ulcer Challenge 2020 (DFUC2020) Dataset and **b** IEEE DataPort Diabetic Foot Dataset

$$F1\text{−Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

(b) mAP is another metric that is widely used for the evaluation of object detection tasks. Average precision is defined as the area under the precision-recall curve. One of the more popular techniques for calculating this area is the 11-point interpolation method. In this method, the shape of the precision-recall (PR) curve is approximated by computing the average of the maximum precision values across 11 equally spaced recall levels [49].

$$\text{AP}_{11} = \frac{1}{11} \sum_{R \in [0, 0.1, \ldots, 0.9, 1]} P_{\text{interp}}(R) \tag{4}$$

where

$$P_{\text{interp}}(R) = \max_{\widetilde{R}:\widetilde{R} \geq R} P\left(\widetilde{R}\right)$$

Instead of calculating the precision $P(R)$ at each recall level R, the AP is determined by considering the highest precision $P_{\text{interp}}(R)$ that has a recall value greater than R. mAP is simply the average of the AP for all the classes. If there are N classes:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{5}$$

For this study, we initialized each of these models with pre-trained weights from the COCO Dataset [39] and trained them for 100–150 epochs. The best model weights were determined by evaluating the mAP scores on the validation split. The performance of each model has been reported in the next section.

Proposed Method

On the dataset, an analysis of the qualitative results of each of the individual models shows that some models complement each other (shown in Supplementary Fig. 1), and ensembling their results has improved both qualitative and quantitative results. The combination of outputs from the three best-performing models was achieved using three different ensemble techniques. The purpose of these methods was to combine the predicted bounding boxes of the two models to produce a more robust prediction. This study's proposed methodology is illustrated in Fig. 5.

(a) NMS: The non-maximum suppression (NMS) algorithm is used to eliminate the overlapping bounding boxes over a certain threshold of IoU. A box with a relatively low objectness score is usually eliminated. This ensures that the most confident predictions from the two models are manifested in the final output. The process begins with a list of detection boxes and their scores. As soon as the highest-scoring detection is picked, it is removed from the initial set and added to the final detection set. Additionally, it eliminates any box that overlaps with the selected box in the initial set by more than a specified amount. The process is repeated for each of the remaining boxes. The rescoring function of NMS is as follows:

$$s_i = \begin{cases} s_i, & iou\left(m, b_i\right) < N_t \\ 0, & iou\left(m, b_i\right) \geq N_t \end{cases} \tag{6}$$

Here,

m denotes the selected highest scoring bounding box which is added to the final detection set,

b_i denotes a bounding box from the initial set,

s_i denotes the confidence score of the b_i bounding box,
N_t denotes the NMS threshold.

(b) Soft-NMS: Soft-NMS improves the system by reducing the objectness score of the overlapping bounding box instead of eliminating it. As a result, adjacent objects are less likely to be eliminated from predictions. Soft-NMS decays detection scores over a threshold as a linear function of the overlap with the bounding box. Therefore, detection boxes far away from the selected box are not impacted much, but those that are extremely close are penalized heavily [50]. The rescoring function of NMS is as follows:

$$s_i = \begin{cases} s_i, & IoU(m, b_i) < N_t \\ s_i(1 - \mathrm{IoU}(m, b_i)), & IoU(m, b_i) \geq N_t \end{cases} \tag{7}$$

Here,
m denotes the selected highest scoring bounding box which is added to the final detection set,
b_i denotes a bounding box from the initial set,
s_i denotes the confidence score of the b_i bounding box,
N_t denotes the NMS threshold.

However, this causes the penalty to incur suddenly as the IoU exceeds the threshold. This is tackled by updating the pruning step with a Gaussian penalty function applied in each iteration. The updated penalty function is as follows:

$$s_i = s_i e^{\frac{-\mathrm{iou}(m,b_i)}{\sigma}}, \forall b_i \notin D \tag{8}$$

Here,
m denotes the selected highest scoring bounding box which is added to the final detection set,
D denotes the final detection set,
b_i denotes a bounding box from the initial set,
s_i denotes the confidence score of the b_i bounding box,
σ denotes a constant that controls the intensity of the penalty.

(c) WBF: In weighted bounding box fusion, instead of eliminating or reducing some predictions, all bounding boxes and their scores are used to generate new average bounding boxes. This significantly improves the quality of the ensemble process [51]. The following formulas are used to calculate the weighted average bounding box and the new confidence score:

$$C = \frac{\sum_{i=1}^{T} C_i}{T} \tag{9}$$

$$X_{1,2} = \frac{\sum_{i=1}^{T} C_i * X_{1,2}}{\sum_{i=1}^{T} C_i} \tag{10}$$

$$Y_{1,2} = \frac{\sum_{i=1}^{T} C_i * Y_{1,2}}{\sum_{i=1}^{T} C_i} \tag{11}$$

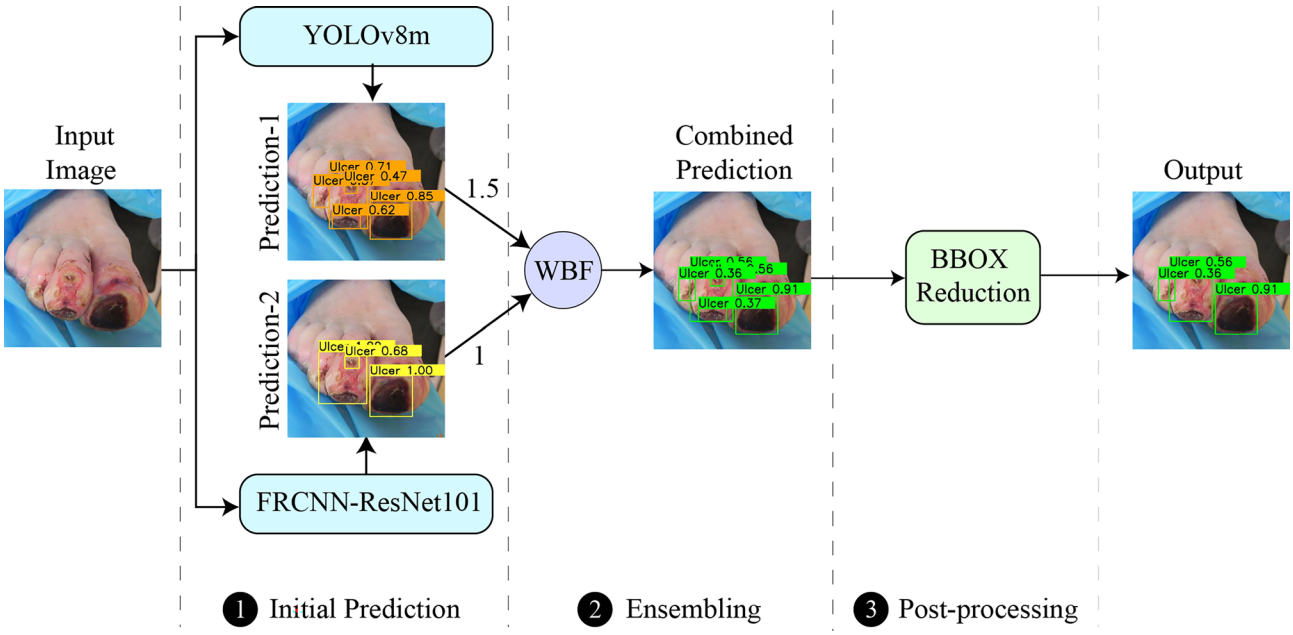

Fig. 5 Block diagram to illustrate the methodology of the study

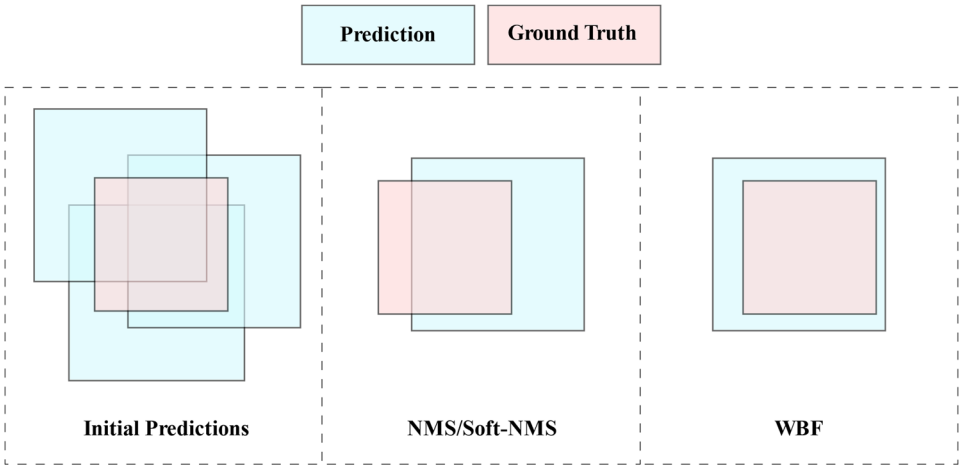

Fig. 6 Visual comparison between WBF and NMS/Soft-NMS

where.

T is the number of bounding boxes with scores greater than a certain threshold,

C_i is the confidence score of the ith bounding box,

$X_{1,2}$ and $Y_{1,2}$ are the x and y coordinates of the top left and bottom right points of the bounding box.

Figure 6 shows a visual comparison of the three ensemble methods. After ensemble, there were some overlapping bounding boxes in the resulting ensembled detections. A post-processing step was employed to mitigate overlaps by prioritizing the larger bounding boxes and removing all the smaller bounding boxes within a certain area-to-overlap ratio, to ensure that all ulcer points were accounted for while further filtering the detection. We empirically chose a threshold of 0.8 for the area-to-overlap ratio for this experiment in order to eliminate duplicate detections without hampering the detection of adjacent ulcer points.

Result and Analysis

This section presents the quantitative and qualitative evaluation results for each step of our approach in this study. We analyze the performance of the above-mentioned models and ensemble techniques on the DFUC2020 dataset. After that, we validate the results by inferencing on the IEEE DataPort Diabetic Foot Dataset. A discussion of the performance of each model is given in the first subsection, followed by an analysis of how they perform after ensembling. In the "Overlapping bounding box reduction" section, the proposed overlapping bounding box reduction technique is demonstrated to improve the results.

Individual Model Performance

We investigated different variations of YOLOv5, YOLOv7, and YOLOv8 models, and the Faster R-CNN ResNet101 and EfficientDet-D1. Pretrained weights from the MS COCO dataset were used to develop the models on the training set and the hyperparameters were tuned on the validation set. The loss and mAP curves are provided in the supplementary materials and the models' quantitative performances are presented in Table 1.

Table 1 shows that YOLOv8x, which is the extra-large version of YOLOv8, gave both the highest mAP@0.5 score of 0.856 and the highest $F1$ score of 0.811. The YOLOv8x model outperforms all other YOLO models, as well as FRCNN-ResNet101 and EfficientDet-D1. Among the non-YOLO models, FRCNN-ResNet101 performed better than EfficientDet-D1 in terms of both $F1$ score and mAP. YOLOv8m's optimal trade-off between inference time and mAP demonstrates the most practical applicability in terms of medical context, where timely and efficient diagnosis is pivotal for taking faster decision-making and enhancing healthcare. It also had a significantly lower total parameter count compared to the other similar performing models making it resource-efficient without compromising on robustness. As shown in Fig. 7, predictions and ground truth are provided for a sample test image so that qualitative results can be visualized. From the figure, it can be observed that while FRCNN-ResNet101 and other YOLO variants accurately detected the two regions identified in the ground truth, YOLOv5x identified an additional third region causing a false-positive prediction. Regarding the bounding box area, models such as YOLOv7x, YOLOv8m, and YOLOv8x demonstrated high precision, aligning most closely with the ground truth. However, FRCNN-Resnet101 predicted the top left ulcer point with a much larger bounding box. These types of predictions may be attributed to the model's lower mAP score, as the lower IoU overlap threshold of less than 0.5 leads to the exclusion of such predictions, despite the model's ability to accurately detect the affected area.

In the context of healthcare, especially when diagnosing conditions such as ulcers using deep learning models like

Table 1 Single model performance on test set

Model name	Precision	Recall	F1 score	mAP@0.5	Avg inference time (ms)	No. of parameters (mill)
YOLOv5l	0.848	0.715	0.776	0.815	21.8	46.1
YOLOv5x	0.868	0.704	0.777	0.819	40.7	86.1
YOLOv7	0.837	0.744	0.788	0.823	12.33	36.4
YOLOv7x	0.855	0.731	0.788	0.824	16.22	70.7
YOLOv8m	0.812	0.769	0.790	0.842	7.2	25.8
YOLOv8x	0.897	0.74	**0.811**	**0.856**	18.9	68.1
F-RCNN Resnet101	0.784	0.793	0.789	0.813	11.33	54.7
EfficientDet-D1	0.733	0.793	0.762	0.796	5.18	6.6

Values highlighted in bold denote the highest performance scores

YOLOv8m, explainability is crucial. Saliency maps generated for the model using Gradient-weighted Class Activation Mapping (GradCAM) are shown in Fig. 8. The reddish hues on the map indicate regions that positively contribute to the model's detection of the ulcer point in the image, while bluish tones suggest areas that are less influential to the detection outcome. The saliency maps in Fig. 8 reveal the model's highly centered attention in the ulcerated region, which is indicative of its ability to precisely pinpoint ulcer locations. Despite the presence of highlighted regions beyond the ulcer area, their effect on the model's results is minimal and does not detract from the overall accuracy of ulcer identification.

Model Ensemble Performance

Apart from the YOLO models, FRCNN-ResNet101 achieved higher accuracy in DFU detection compared to EfficientDet-D1. To develop an ensemble model for DFU prediction, we combined the FRCNN-ResNet101 model with the top three performing YOLO models (YOLOv7x, YOLOv8m, and YOLOv8x) to investigate the performance of DFU prediction. The predictions were combined using three different ensemble techniques (NMS, Soft-NMS, and WBF). The IoU threshold of 0.5 was chosen to determine detections for all the methods. Bounding boxes with less than

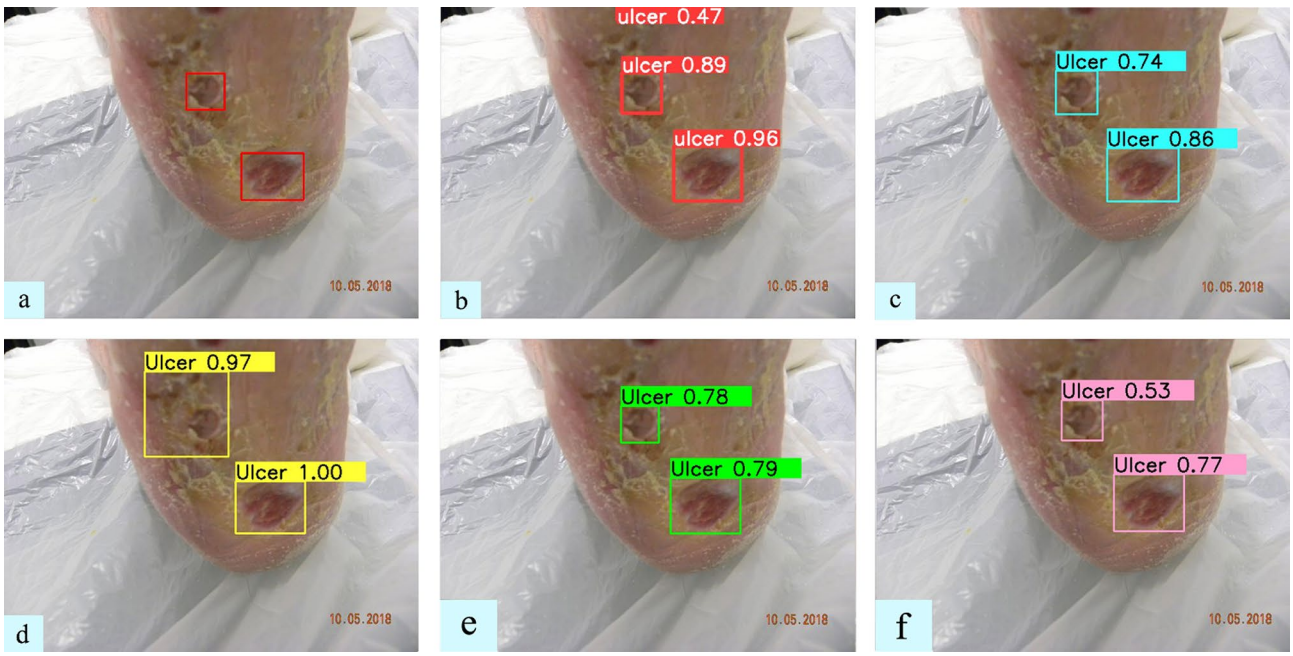

Fig. 7 Sample test image using best individual models. **a** Ground truth, **b** YOLOv5x, **c** YOLOv7x, **d** F-RCNN Resnet101, **e** YOLOv8m, and **f** YOLOv8x

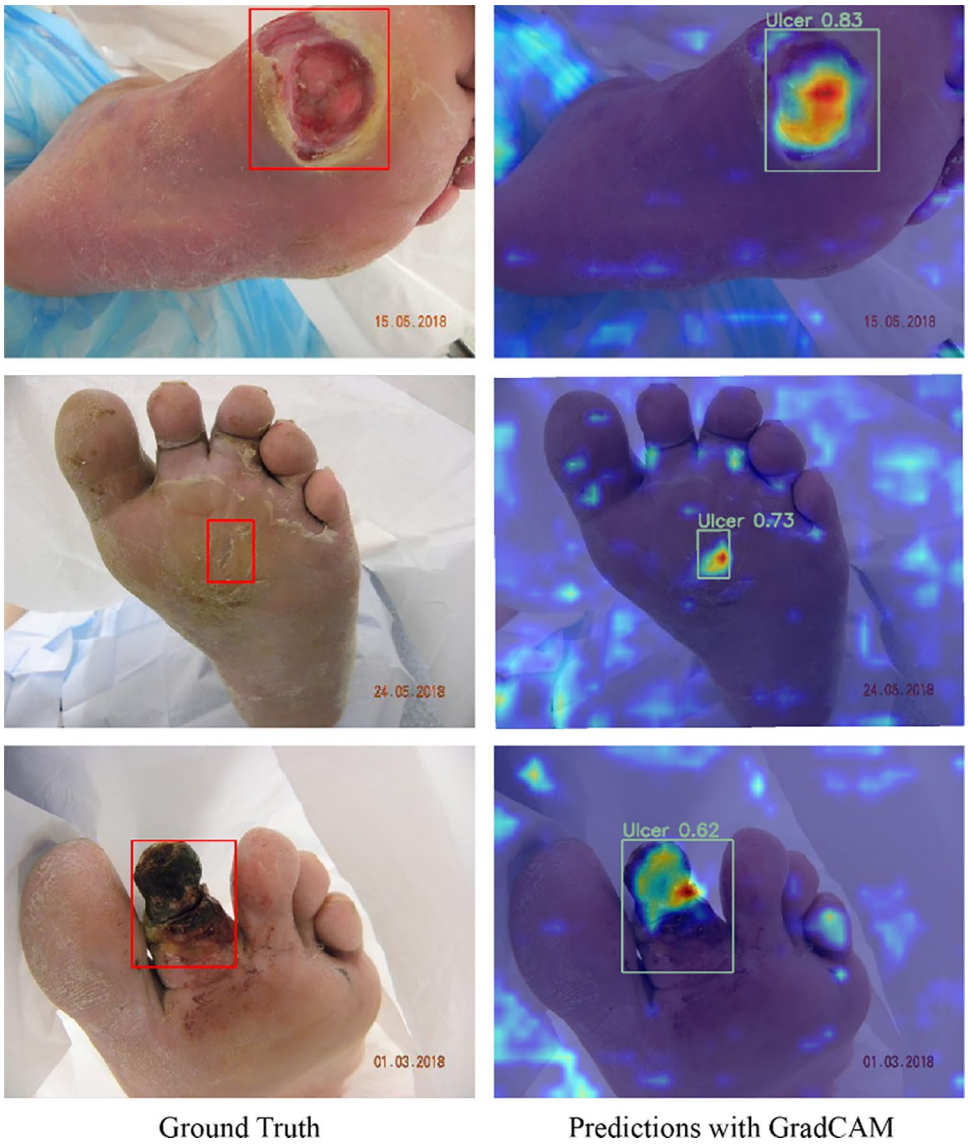

Fig. 8 YOLOv8 GradCAM saliency map visualization

0.001 confidence score were eliminated. A low Sigma value of 0.1 was chosen for Soft-NMS. According to Eq. 8, a lower sigma value highly suppresses the confidence scores of the overlapping bounding boxes without

Table 2 Different ensemble method performances on test set

Models	Method	Precision	Recall	F1 score	mAP@0.5
YOLOv7 + FRCNN-Resnet101	NMS	0.792	0.818	0.805	0.845
	Soft-NMS	0.687	0.826	0.796	0.826
	WBF	0.786	0.806	0.796	0.850
YOLOv7x + FRCNN-Resnet101	NMS	0.767	0.818	0.792	0.825
	Soft-NMS	0.681	0.839	0.752	0.829
	WBF	0.793	0.80	0.799	0.841
YOLOv8m + FRCNN-Resnet101	NMS	0.769	0.798	0.783	0.846
	Soft-NMS	0.701	0.806	0.750	0.827
	WBF	0.768	0.793	0.780	**0.864**
YOLOv8x + FRCNN-Resnet101	NMS	0.826	0.785	**0.810**	0.852
	Soft-NMS	0.713	0.789	0.749	0.823
	WBF	0.835	0.752	0.791	0.850

Values highlighted in bold denote the highest performance scores

Table 3 Individual model performance on the test set

Model type	Model name	F1 score	mAP@0.5
YOLO based	YOLOv5l	0.818	0.836
	YOLOv5x	0.792	0.832
	YOLOv7	0.792	0.85
	YOLOv7x	0.822	0.851
	YOLOv8m	0.790	**0.86**
	YOLOv8x	0.811	0.853
Other	F-RCNN Resnet101	0.777	0.8
	EfficientDet-D1	0.746	0.788

Values highlighted in bold denote the highest performance scores

completely eliminating them. Before applying Soft-NMS and WBF, weight values of 1.5 and 1 were used for the YOLO-based models and FRCNN-ResNet101, respectively. This bias was implemented as the YOLO-based models surpass FRCNN ResNet101 in individual performance in most cases, which is evident from Table 1 and Supplementary Fig. 1. The quantitative performance after ensemble is presented in Table 2.

As a result of using the NMS ensemble technique to combine predictions, in most cases, mAP is modestly higher than in the individual models. However, YOLO8x is slightly lower. Meanwhile, Soft-NMS performs poorly across all models, since it significantly lowers both mAP and $F1$ scores, with the exception of YOLOv7 and YOLOv7x, where it increases the mAP score from 0.824 to 0.829 and from 0.823 to 0.826, respectively. Other than the YOLOv8x model, where a slight decrease in mAP score can be observed, WBF provided the most excellent results, significantly improving mAP while minimally impacting the $F1$ score. Combining predictions from YOLOv8m and FRCNN-Resnet101, this technique achieves the highest mAP score of 0.864, which represents a significant improvement over both of their individual performances and surpasses the current leaderboard of the DFUC2020 challenge by 12.4% [52]. For the remaining experiments, only predictions based on the WBF approach have been considered for YOLOv8m and FRCNN-ResNet101 (Table 3).

Figure 9 shows a qualitative comparison between the ensemble outputs and the two fundamental models. Comparing the individual model performances to the results obtained through ensemble, it is evident that the ensembled results are significantly better. It can, for example, compensate for the fact that one of the models misses an ulcer point, as depicted in the figure. Even though all the ensembling techniques provide comparable qualitative results in detecting the DFU-affected areas in the ground truth almost perfectly, the WBF method is the most confident in detecting the regions.

Overlapping Bounding Box Reduction

The most reliable results obtained using the WBF ensemble on YOLOv8m and FRCNN ResNet101 are shown in Table 2. However, the ensemble method introduces multiple detections or overlapping detections for some images. To address this problem, we employed an overlapping bounding box reduction technique prioritizing the larger detection area.

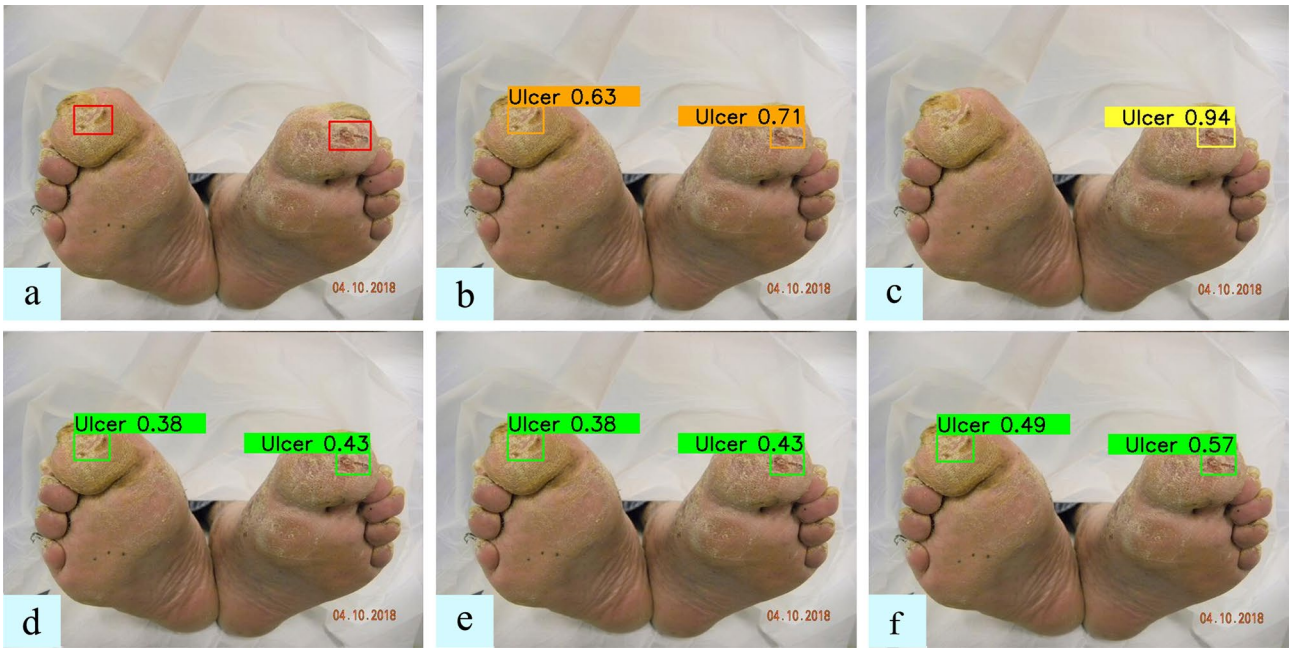

Fig. 9 Sample test image prediction with ensemble techniques. **a** Ground truth, **b** YOLOv8m, **c** F-RCNN Resnet101, **d** NMS, **e** Soft-NMS, and **f** WBF

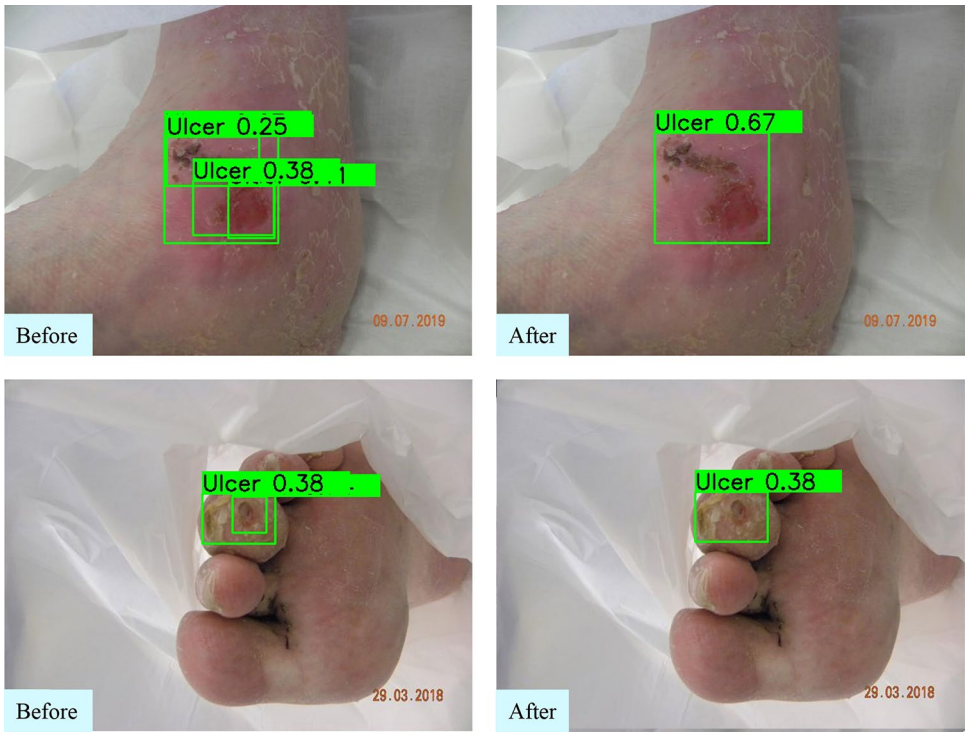

Fig. 10 Qualitative result improvement after overlap reduction

We removed any smaller bounding boxes that have an area-to-overlap with other bounding boxes with an intersection ratio greater than 0.8. The results are depicted side-by-side in Fig. 10. The figure shows how smaller bounding boxes can appear inside larger bounding boxes detecting the same DFU-affected area and how overlapping bounding box reduction can be applied to this problem.

External Validation

We validated our proposed DFU detection system using IEEE DataPort Diabetic Foot datasets to predict foot ulcers. The visual clarity of the DFUC2020 dataset surpassed that of the IEEE DataPort dataset, where many ulcer points were either out of focus or positioned at the edge of the foot. Sometimes the view of the foot is obstructed by the L-shaped scale. So, the model failed to predict some true positives from the suboptimal images within the IEEE DataPort's dataset. In addition, there were different background objects on the IEEE Dataport's dataset which led the model to make some false prediction. To tackle this issue, we cropped some overly prevalent background elements from the images of foot ulcers in the IEEE DataPort database to reduce the false positive

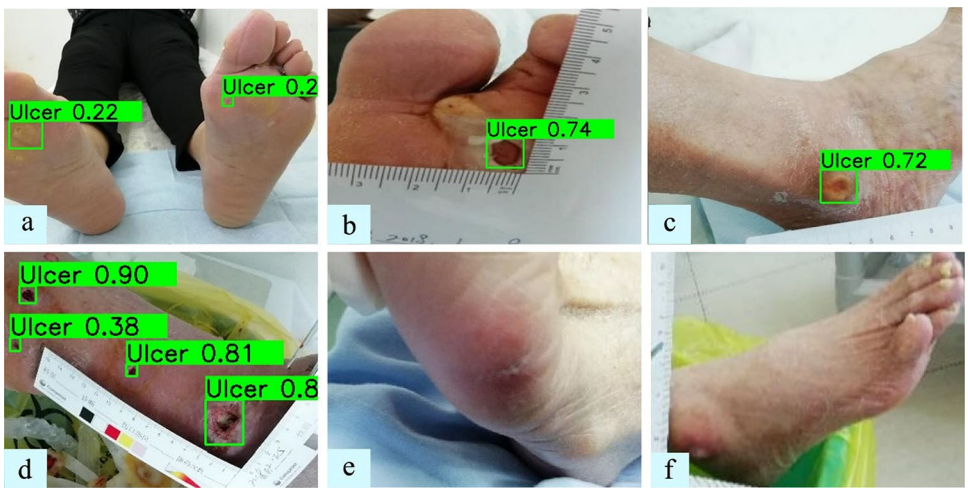

Fig. 11 Prediction on the external dataset

detections owing to the presence of irrelevant objects in the image. The above proposition was effective for most of the images in the validation dataset. The prediction results for the dataset are shown in Fig. 11 in which we can see that for the first four images, our method predicts ulcer areas almost accurately, but for the last two images (Fig. 11e, f), it failed to identify the ulcer area. In these last two images, the ulcer points are not clearly visible because they are out of focus. As the original dataset used in this study did not have such poor-quality images, where the DFU-affected areas were blurry, this kind of result is to be expected.

Ablation Study

This section presents the ablation study for evaluating the performance of various deep learning models on the validation set, with a focus on understanding the nuances of their architectures and the effectiveness of ensemble techniques in diabetic foot ulcer (DFU) detection.

Individual Network Performance

Our study categorized the evaluated models into two groups based on their network architectures: YOLO-based models and other models. Table 3 reveals that the newer architectures like YOLOv8 performed better than earlier versions like YOLOv5. One of the key reasons was that the output heads in YOLOv8, which serve as the last layers of the neural network, have been simplified in comparison to earlier iterations such as YOLOv5. YOLOv8 employs a solitary output head, in contrast to the three heads present in YOLOv5, and utilizes an anchor-free detection technique, unlike YOLOv5, which relies on an anchor-based strategy. This approach directly predicts the center of the object, reducing the number of bounding boxes and thereby increasing the efficiency of the post-processing stage. Additionally, YOLOv8 integrates Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) modules, aiding in producing multi-scale feature maps and combining features from different levels of the network, respectively. These modifications in YOLOv8's backbone architecture streamline information flow within the network and enhance the efficiency and effectiveness of object detection tasks [53].

Table 4 Ensemble of YOLO-based models performance

Model combinations	F1 score	mAP@0.5
YOLOv7 + YOLOv8m	0.779	0.845
YOLOv7 + YOLOv8x	0.751	0.84
YOLOv7x + YOLOv8m	0.805	0.861
YOLOv7x + YOLOv8x	0.786	0.845

YOLO-Based Model Ensemble

Since the YOLO-based models performed better than the other architectures, we have presented the combination of different YOLO models using the WBF module. Analysis with the other modules is discussed at a later section. Our results in Table 4 indicate that the ensemble performance did not significantly exceed the performance of individual models. This finding suggests that while YOLO models are individually robust, their similarities in architectural design and detection approach lead to a convergence in their detection capabilities. As a result, the ensemble models tend to reinforce the same strengths and weaknesses, rather than complementing and compensating for each other's limitations.

Ensemble of Different Architecture Models

Contrasting with the YOLO-based model ensemble, combining YOLO models with different architectures yielded more promising results. This approach leverages the complementary strengths of different detection algorithms, potentially addressing the limitations of a single-model approach. From Table 5, we see that the combination of YOLO models with FRCNN-ResNet101 mostly resulted in improved performance, with the combination of YOLOv8m and FRCNN-ResNet101 emerging as the most successful ensemble, yielding the best result in terms of mAP. This can be due to FRCNN's approach to object detection, which includes selective search and the use of a Region Proposal Network (RPN). This enhances its ability to detect objects more accurately compared to other models, which when combined, compliments YOLO's efficient architecture improving overall performance in different scenarios [54]. FRCNN-ResNet101's two-stage approach comprising of first feature extraction and then doing region proposal enhances the accuracy of detection on some complex scenarios compared to YOLO's single-stage approach [55]. However, the combination with EfficientDet-D1, which had the lowest performance among the individual models, did

Table 5 Ensemble of different architecture model performance

Model combinations	F1 score	mAP@0.5
YOLOv7 + EfficientDet-D1	0.747	0.83
YOLOv7 + FRCNN-Resnet101	0.791	0.86
YOLOv7x + EfficientDet-D1	0.787	0.834
YOLOv7x + FRCNN-Resnet101	0.806	0.853
YOLOv8m + EfficientDet-D1	0.779	0.851
YOLOv8m + FRCNN-Resnet101	0.793	**0.87**
YOLOv8x + EfficientDet-D1	0.743	0.825
YOLOv8x + FRCNN-Resnet101	0.795	0.848
EfficientDet-D1 + FRCNN-Resnet101	0.792	0.799

Values highlighted in bold denote the highest performance scores

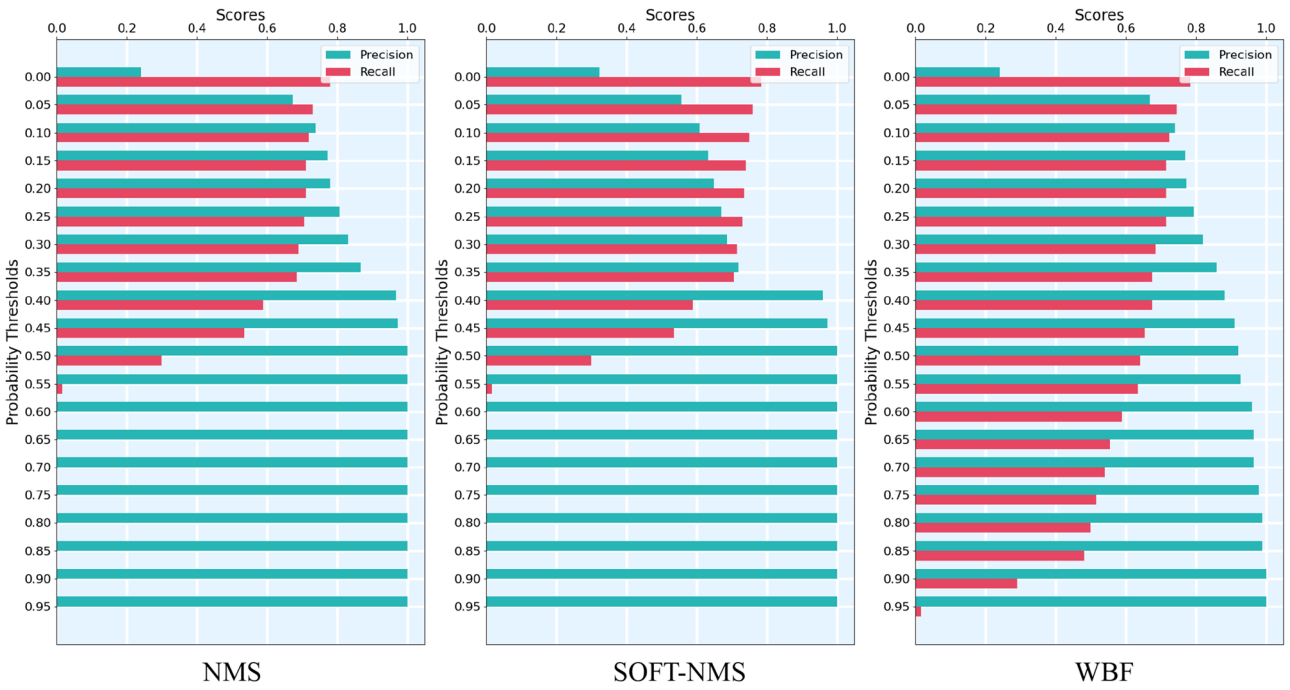

Fig. 12 Precision recall trade-off for different ensembling methods

not yield significant improvements. This could be due to EfficientDet-D1's limitations not being effectively addressed by the YOLO models' capabilities as both are single-stage detectors.

Precision Recall Tradeoff

The study also examined different confidence score thresholds to determine the optimal balance between precision and recall and select the best ensemble technique. As thresholds increased, the number of detections decreased, leading to a reduction in false positives but an increase in false negatives. Consequently, recall diminished as precision improved. Figure 12 illustrates this precision-recall tradeoff for various ensemble techniques at various confidence thresholds. The NMS and Soft-NMS showed a drastic drop in recall after the 0.55 threshold, while the WBF method demonstrated a more proportional trade-off. This is mostly because NMS removes the additional bounding boxes with lower confidence values that cross the IoU threshold for each detection. As a result, it showed higher precision in the first half compared to the Soft-NMS approach. Soft-NMS reduces the confidence scores of additional bounding boxes rather than fully eliminating them, resulting in increased recall in the first half. However, recall dropped drastically in the second half for both methods. WBF method on the other hand does not eliminate or reduce additional bounding boxes, but instead computes a weighted average based on the confidence scores. This resulted in a more gradual decline in recall maintaining a smoother trade-off across the whole confidence range, making WBF a robust option compared to NMS and Soft-NMS techniques. Based on the precision-recall trade-off graph, we chose 0.1 confidence threshold in this study as the optimal confidence score maximizing both precision and recall for the ensemble technique.

Overall, this ablation study reveals that the best YOLO models ensembled with FRCNN-ResNet101 using the WBF technique provide the best results for DFU detection. It also highlights the importance of considering architectural differences of combining models to enhance diagnostic accuracy.

Limitations and Future Scopes

In our research, we aimed to enhance the early detection of diabetic foot ulcers (DFUs), recognizing its crucial role in healthcare. Our approach incorporates a weighted bounding box fusion between two models (YOLOv8m and FRCNN-Resnet101) to create a robust system for DFU detection. While our approach has shown remarkable performance for DFU detection, it does have a few notable limitations:

- One of the key limitations of our work stemmed from the lack of diversity of non-DFU conditions in the available part of the DFUC2020 dataset. Other skin conditions such as keloids, onychomycosis and psoriasis may share visual similarities with DFUs, potentially leading to confusion. Sometimes, the model is

also confused by healed ulcer points, which is mainly due to the lack of severity categorization in the dataset.
- As detailed in the "Dataset Description" section, the DFUC2020 dataset is predominantly composed of images representing white individuals, with 1987 white cases compared to only 13 non-white cases. This significant ethnic disparity suggests potential limitations in the model's performance for people with non-white skin tones, due to underrepresentation in the dataset.
- In a few cases, the model predicted false positives (FP) outside of the foot as the model is distracted by other irrelevant objects.
- Our study was confined by the restricted access to the DFUC2020 dataset. Consequently, our model training and evaluation were based solely on the available DFUC2020 images. Although we recorded a 12.4% improvement in detection accuracy, we only evaluated on a limited 10% split of the available images from the DFUC2020 dataset.

Moving forward, the scope of our research will expand to mitigate these limitations and include more nuanced aspects of DFU detection. We aim to do the following:

- Conduct further investigations on a larger and ethnically diverse dataset containing non-DFU skin conditions and healthy foot images. This will help generalize the model to more diverse scenarios.
- Incorporate automatic foot area segmentation to narrow down the region of interest. This will allow the model to better narrow down on the ulcer regions without getting distracted by irrelevant objects.
- Extend our investigation to segment and classify ulcer points into several clinically relevant categories. This will potentially also help the model to better distinguish between healed and partially healed ulcer points.

Conclusion

In the conclusion of this study, we successfully developed an innovative diabetic foot ulcer (DFU) detection system using an ensemble of deep neural networks. Multiple state-of-the-art deep learning based state-of-the-art object detection models were evaluated in our study and their predictions were combined to enhance performance. While models like YOLOv5 and FRCNN-ResNet were effective in general but missed smaller ulcers, the more sensitive YOLOv7 and YOLOv8 models tended to generate more false positives in complex images. To balance these traits, ensembling via WBF was utilized to improve DFU prediction in this study. Although the results indicate that ensemble methods enhance the localization of DFUs, it is important to acknowledge the presence of false positives in our findings. However, this is within the intended scope of the project as the primary objective of our system is to serve as an initial screening tool for medical professionals and hospitals. It is designed to assist, and by no means replace, the critical human aspect of diagnosis in healthcare.

The role of this system is to augment the medical diagnostic process, particularly in the complex and varied field of DFU management. By providing doctors with an initial assessment, the system guides further detailed investigation. Ensuring higher sensitivity to potential cases of DFU is crucial for prompt and effective early screening. In summary, our research makes a significant contribution to medical diagnostics, offering a novel, AI-driven tool for the early detection of DFUs. While the system has its limitations, the wider implementation of our research in the area of diabetic foot ulcer detection, combined with ongoing improvements to address current shortcomings, is aimed to enhance patient care. The integration of human medical expertise with our AI-based solutions is set to offer a more all-encompassing, precise, and streamlined diagnostic approach.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12559-024-10267-3.

Acknowledgements We would like to thank Prof. Moi Hoon Yap for sharing the Diabetic Foot Ulcer Challenge 2020 (DFUC2020) Dataset under a data-sharing agreement which makes this work feasible.

Funding Open Access funding provided by the Qatar National Library. This work was made possible by Qatar National Research Fund (QNRF) NPRP12S-0227–190164 and International Research Collaboration Co-Fund (IRCC) grant: IRCC-2021–001. The statements made herein are solely the responsibility of the authors. The open access publication cost is covered by Qatar National Library.

Data Availability The dataset used in this study cannot be shared due to data-sharing agreement of the dataset provider.

Declarations

Ethical Approval This study is conducted on two publicly accessible datasets. The main training dataset is made available by Prof. Moi Hoon Yap as Diabetic Foot Ulcer Challenge 2020 (DFUC2020) Dataset while external validation set was available from IEEE DataPort. Therefore, ethical approval is not applicable for this study.

Informed Consent As this study uses two publicly available datasets, informed consent is not applicable for this study.

Conflict of Interest The authors declare no competing interests.

References

1. Higginson R, Burrows P, Jones B. Continuing professional development: diabetes and associated diabetic emergencies. Journal of Paramedic Practice. 2019;11(6):1–5.
2. Saeedi P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas. Diabetes Res Clin Pract. 2019;157: 107843.
3. Noor S, Zubair M, Ahmad J. Diabetic foot ulcer—a review on pathophysiology, classification and microbial etiology. Diabetes Metab Syndr. 2015;9(3):192–9.
4. Killeen AL, Brock KM, Dancho JF, Walters JL. Remote temperature monitoring in patients with visual impairment due to diabetes mellitus: a proposed improvement to current standard of care for prevention of diabetic foot ulcers. J Diabetes Sci Technol. 2020;14(1):37–45.
5. Reyzelman AM, et al. Continuous temperature-monitoring socks for home use in patients with diabetes: observational study. J Med Internet Res. 2018;20(12): e12460.
6. Armstrong DG, Boulton AJ, Bus SA. Diabetic foot ulcers and their recurrence. N Engl J Med. 2017;376(24):2367–75.
7. Ananian CE, et al. A multicenter, randomized, single-blind trial comparing the efficacy of viable cryopreserved placental membrane to human fibroblast-derived dermal substitute for the treatment of chronic diabetic foot ulcers. Wound Repair and Regeneration. 2018;26(3):274–83.
8. Boutoille D, Féraille A, Maulaz D, Krempf M. Quality of life with diabetes-associated foot complications: comparison between lower-limb amputation and chronic foot ulceration. Foot Ankle Int. 2008;29(11):1074–8.
9. Cavanagh P, Attinger C, Abbas Z, Bal A, Rojas N, Xu ZR. Cost of treating diabetic foot ulcers in five different countries. Diabetes Metab Res Rev. 2012;28:107–11.
10. Yap MH, Kendrick C, Reeves ND, Goyal M, Pappachan JM, Cassidy B. "Development of diabetic foot ulcer datasets: an overview," Diabetic Foot Ulcers Grand Challenge: Second Challenge, DFUC 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings. 2022:1–18.
11. Yap MH, et al. A new mobile application for standardizing diabetic foot images. J Diabetes Sci Technol. 2018;12(1):169–73.
12. Yap MH, et al. Computer vision algorithms in the detection of diabetic foot ulceration: a new paradigm for diabetic foot care? J Diabetes Sci Technol. 2016;10(2):612–3.
13. Goyal M, Yap MH, Reeves ND, Rajbhandari S, Spragg J. "Fully convolutional networks for diabetic foot ulcer segmentation," in 2017 IEEE int conf syst man cybern (SMC), 2017:618–623.
14. Hernandez-Contreras D, Peregrina-Barreto H, Rangel-Magdaleno J, Ramirez-Cortes J, Renero-Carrillo F. Automatic classification of thermal patterns in diabetic foot based on morphological pattern spectrum. Infrared Phys Technol. 2015;73:149–57.
15. Khandakar A, et al. Thermal change index-based diabetic foot thermogram image classification using machine learning techniques. Sensors. 2022;22(5):1793.
16. Khandakar A, et al. A machine learning model for early detection of diabetic foot using thermogram images. Comput Biol Med. 2021;137: 104838.
17. Das SK, Roy P, Mishra AK. DFU_SPNet: a stacked parallel convolution layers based CNN to improve Diabetic Foot Ulcer classification. ICT Express. 2022;8(2):271–5.
18. Alzubaidi L, Fadhel MA, Oleiwi SR, Al-Shamma O, Zhang J. DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. Multimedia Tools and Applications. 2020;79(21):15655–77.
19. Wang L, Pedersen PC, Agu E, Strong DM, Tulu B. Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. IEEE Trans Biomed Eng. 2016;64(9):2098–109.
20. Goyal M, Reeves ND, Davison AK, Rajbhandari S, Spragg J, Yap MH. Dfunet: convolutional neural networks for diabetic foot ulcer classification. IEEE Transactions on Emerging Topics in Computational Intelligence. 2018;4(5):728–39.
21. Simonyan K, Zisserman A, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint.* 2014. arXiv:1409.1556.
22. Szegedy C. et al., "Going deeper with convolutions," in *Proceedings of the* IEEE conf comp vis pattern recognition. 2015:1-9.
23. Thotad PN, Bharamagoudar GR, Anami BS, "Diabetic foot ulcer detection using deep learning approaches," Sens Int. 2022:100210.
24. Howard AG *et al.*, "Mobilenets: efficient convolutional neural networks for mobile vision applications" *arXiv preprint.* 2017. arXiv:1704.04861.
25. Dai J, Li Y, He K, Sun J. "R-fcn: object detection via region-based fully convolutional networks," Adv neural inf process systems. 2016;29.
26. Goyal M, Reeves ND, Rajbhandari S, Yap MH. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. IEEE J Biomed Health Inform. 2018;23(4):1730–41.
27. Zhu X, Hu H, Lin S, Dai J. "Deformable convnets v2: more deformable, better results," in Proceedings of the IEEE/CVF conf comp vis pattern recog, 2019, pp. 9308–9316.
28. Yap MH, et al. Deep learning in diabetic foot ulcers detection: a comprehensive evaluation. Comput Biol Med. 2021;135: 104596.
29. Rogers LC, Lavery LA, Joseph WS, Armstrong DG. "All feet on deck—the role of podiatry during the COVID-19 pandemic: preventing hospitalizations in an overburdened healthcare system, reducing amputation and death in people with diabetes," J Am Podiatr Med Assoc 2020;1(aop):0000–0000.
30. Rogers LC, et al. Wound center without walls: the new model of providing care during the COVID-19 pandemic. Wounds a compend clin res pract. 2020;32(7):178.
31. Wang C-Y, Liao H-YM, Wu Y-H, Chen P-Y, Hsieh J-W, Yeh I-H. "CSPNet: a new backbone that can enhance learning capability of CNN," in Proceedings of the IEEE/CVF conf comp vis pattern recognit workshops, 2020, pp. 390–391.
32. Wang K, Liew JH, Zou Y, Zhou D, Feng J. "Panet: Few-shot image semantic segmentation with prototype alignment," in Proceedings of the IEEE/CVF Int Conf Comp Vis 2019, pp. 9197–9206.
33. Redmon J, Farhadi A. "Yolov3: an incremental improvement," arXiv preprint 2018. arXiv:1804.02767
34. Rahman T, et al. HipXNet: deep learning approaches to detect aseptic loos-ening of hip implants using X-ray images. IEEE Access. 2022;10:53359–73.
35. Wang C-Y, Bochkovskiy A, Liao H-YM. "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proceedings of the IEEE/CVF Conf Comp Vis Pattern Recognit 2023, pp. 7464–7475.
36. Hussain M, Al-Aqrabi H, Munawar M, Hill R, Alsboui T. Domain feature mapping with YOLOv7 for automated edge-based pallet racking inspections. Sensors. 2022;22(18):6927.

37. Zhu W, et al. CPAM: cross patch attention module for complex texture tile block defect detection. Appl Sci. 2022;12(23):11959.
38. Glenn J, Ayush C, Jing Q. YOLO by Ultralytics (version 8.0.0). https://github.com/ultralytics/ultralytics. Accessed 14 Feb 2023.
39. Lin T-Y, et al. Microsoft coco: common objects in context. In: Euro conf comp vis. Springer; 2014. p. 740–55.
40. Terven J, Cordova-Esparza D. "A comprehensive review of YOLO: From YOLOv1 and beyond. arXiv 2023," *arXiv preprint* arXiv:2304.00501, 2023.
41. Jacob Solawetz F. "What is YOLOv8? The ultimate guide." https://blog.roboflow.com/whats-new-in-yolov8/. Accessed 14 Feb 2023.
42. Zhou Q-Q, et al. Automatic detection and classification of rib fractures on thoracic CT using convolutional neural network: accuracy and feasibility. Korean J Radiol. 2020;21(7):869.
43. Ren S, He K, Girshick R, Sun J. "Faster r-cnn: towards real-time object detection with region proposal networks," Adv neural inform process syst 2015;vol. 28.
44. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conf comp vis pattern recognit. 2016. p. 770–778.
45. Tan M, Pang R, Le QV. Efficientdet: scalable and efficient object detection. In: Proceedings of the IEEE/CVF con comp vis pattern recognit. 2020. p. 10781–10790.
46. Lin TY, Dollár P, Girshick R, He R, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE con comp vis pattern recognit. 2017. p. 2117–2125.
47. Cassidy B, et al. The DFUC 2020 dataset: analysis towards diabetic foot ulcer detection. touch REV Endocrinology. 2021;17(1):5.
48. Zhu H. Diabetic foot. In: Zhu HE, editor. IEEE DataPort: IEEE DataPort 2020. Accessed 14 Feb 2023.
49. Padilla R, Netto SL, Da Silva EA. A survey on performance metrics for object-detection algorithms. In: int con syst sig image process (IWSSIP),IEEE. 2020. p. 237–242.
50. Bodla N, Singh B, Chellappa R, Davis LS. "Soft-NMS--improving object detection with one line of code," in Proceedings of the IEEE int con comp vis 2017:pp. 5561–5569.
51. Solovyev R, Wang W, Gabruseva T. Weighted boxes fusion: ensembling boxes from different object detection models. Image Vis Comput. 2021;107: 104117.
52. Diabetic Foot Ulcer Challenge 2020. https://dfu2020.grand-challenge.org/evaluation/challenge/leaderboard/. Accessed 18 Feb 2023.
53. Reis D, Kupec J, Hong J, Daoudi A. "Real-time flying object detection with YOLOv8," 2023. *arXiv preprint* arXiv:2305.09972.
54. Srivastava S, Divekar AV, Anilkumar C, Naik I, Kulkarni V, Pattabiraman V. Comparative analysis of deep learning image detection algorithms. J Big data. 2021;8(1):1–27.
55. Kim J-A, Sung J-Y, Park S-H. "Comparison of Faster-RCNN, YOLO, and SSD for real-time vehicle type recognition," in 2020 IEEE int conf consum elec Asia (ICCE-Asia) 2020;IEEE:pp. 1–4.

Authors and Affiliations

Rusab Sarmun[1] · Muhammad E. H. Chowdhury[2] · M. Murugappan[3,4,5] · Ahmed Aqel[6] · Maymouna Ezzuddin[2] · Syed Mahfuzur Rahman[7] · Amith Khandakar[2] · Sanzida Akter[8] · Rashad Alfkey[9] · Anwarul Hasan[3]

✉ Muhammad E. H. Chowdhury
mchowdhury@qu.edu.qa

✉ Anwarul Hasan
hasan.anwarul.mit@gmail.com

Rusab Sarmun
rusabsarmun@gmail.com

M. Murugappan
m.murugappan@gmail.com

Ahmed Aqel
aa1205161@qu.edu.qa

Syed Mahfuzur Rahman
mahfuz3947@gmail.com

Sanzida Akter
sanzi575@gmail.com

Rashad Alfkey
rabdelmoaty@hamad.qa

[1] Department of Electrical and Electronic Engineering, University of Dhaka, Dhaka, Bangladesh

[2] Department of Electrical Engineering, Qatar University, 2713 Doha, Qatar

[3] Intelligent Signal Processing (ISP) Research Lab, Department of Electronics and Communication Engineering, Kuwait College of Science and Technology, Block 4, 13133 Doha, Kuwait

[4] Department of Electronics and Computer Engineering, Faculty of Engineering, Vels Institute of Sciences, Technology, and Advanced Studies, Chennai 600117, India

[5] Centre of Excellence for Unmanned Aerial Systems (CoEUAS), Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia

[6] Department of Industrial and Mechanical Engineering, Qatar University, 2713 Doha, Qatar

[7] Department of Biomedical Engineering, Military Institute of Science and Technology, Mirpur Cantonment, Dhaka 1216, Bangladesh

[8] Department of Neurology, BIRDEM General Hospital, Dhaka 1000, Bangladesh

[9] Acute Care Surgery and General Surgery, Hamad Medical Corporation, Doha, Qatar