IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 June 2023, accepted 17 June 2023, date of publication 27 June 2023, date of current version 3 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289839

RESEARCH ARTICLE

Pressure Ulcer Categorization and Reporting in Domiciliary Settings Using Deep Learning and Mobile Devices: A Clinical Trial to Evaluate End-to-End Performance

PAUL FERGUS[1], CARL CHALMERS[2], WILLIAM HENDERSON[2], DANNY ROBERTS[2], AND ATIF WARAICH[2]

[1]Liverpool John Moores University, L3 3AF Liverpool, U.K.
[2]Mersey Care NHS Foundation Trust, L34 1PJ Prescot, U.K.

Corresponding author: Paul Fergus (p.fergus@ljmu.ac.uk)

This work was supported by the Department for Digital, Culture, Media & Sport (DCMS) under the Liverpool 5G Create Project.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Health Research Authority (HRA) under Approval No. IRAS: 253949.

ABSTRACT Pressure ulcers are a challenge for patients and healthcare professionals. In the UK, pressure ulcers affect 700,000 people each year. Treating them costs the National Health Service €3.8 million every day. Their etiology is complex and multifactorial. However, evidence has shown a strong link between old age, disease-related sedentary lifestyles, and unhealthy eating habits. Direct skin contact with a bed or chair without frequent position changes can cause pressure ulcers. Urinary and faecal incontinence, diabetes, and injuries that restrict body position and nutrition are also known risk factors. Guidelines and treatments exist but their implementation and success vary across different healthcare settings. This is primarily because healthcare practitioners have a) minimal experience in dealing with pressure ulcers, and b) a general lack of understanding of pressure ulcer treatments. Poorly managed, pressure ulcers can lead to severe pain, a poor quality of life, and significant healthcare costs. In this paper, we report the findings of a clinical trial conducted by Mersey Care NHS Foundation Trust that evaluated the performance of a faster region-based convolutional neural network and mobile platform that categorised and documented pressure ulcers automatically. The neural network classifies category I, II, III, and IV pressure ulcers, deep tissue injuries, and pressure ulcers that are unstageable. District nurses used their mobile phones to take pictures of pressure ulcers and transmit them over 4/5G communications to an inferencing server for classification. The approach uses existing deep learning technologies to provide a novel end-to-end pipeline for pressure ulcer categorisation that works in ad hoc domiciliary settings. The strength of the approach resides within MLOPS, model deployment at scale, and the platforms in-situ operation. While solutions exist in the NHS for analysing pressure ulcers none of them automatically classify and report pressure ulcers from a service users' residential home automatically. We acknowledge that there is a great deal of work to do, but the approach offers a convincing solution to standardise pressure ulcer categorisation and reporting. The results from the study are encouraging and show that using 216 images, collected over an eight-month trial, it was possible to generate a mean average Precision=0.6796, Recall=0.6997, F1-Score=0.6786 with 45 false positives using an @.75 confidence score threshold.

INDEX TERMS Pressure ulcers, MLOPS, faster region-based convolutional neural networks, classification, deep learning, machine learning, clinical practice, patient care, in-situ operation.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

I. INTRODUCTION

In the UK, pressure ulcers affect 700,000 people each year [1]. According to National Health Service (NHS)

Improvement, pressure ulcers cost the NHS more than €3.8 million every day to manage and treat [2]. In England, 24,674 patients developed a new pressure ulcer between April 2015 and March 2016 [2]. UK-wide, the number of new pressure ulcers in 2017/2018 was 200,000. The cost to the NHS for treating a category I pressure ulcer is €1,124 while a category IV is €14,108. [3], [4]. A House of Lords strategy discussion group in the UK in November 2017 reported that the NHS spent €5 billion on wound care every year - a similar financial cost to the NHS for managing obesity [5]. Malpractice claims against UK trusts relating to pressure ulcers increased by forty three percent in the three years leading up to 2017-18. The number of litigation cases increased from 279 in 2014-15 to 399 in 2017-18 with the bill to the NHS increasing fifty three percent from more than €13.6m to €20.8m. In total, pressure ulcer claims cost €72.4m over that period. While most cases are settled out of court for €20-30,000, some have cost the NHS as much as €3m [6].

Unrelieved pressure over bony parts of the body cause pressure ulcers [7]. Skin shearing, friction, moisture, and faecal soiling increase the risk of pressure ulcers significantly. These conditions are common in patients that are elderly, sick, debilitated or paralysed [8]. Poorly managed, pressure ulcers can lead to severe pain, reduced quality of life and significant economic costs to the NHS [9]. Pressure ulcers can be either a Category I, II, III, IV pressure ulcer, a Deep Tissue Injury (DTI) or Unstageable (Figure 1).

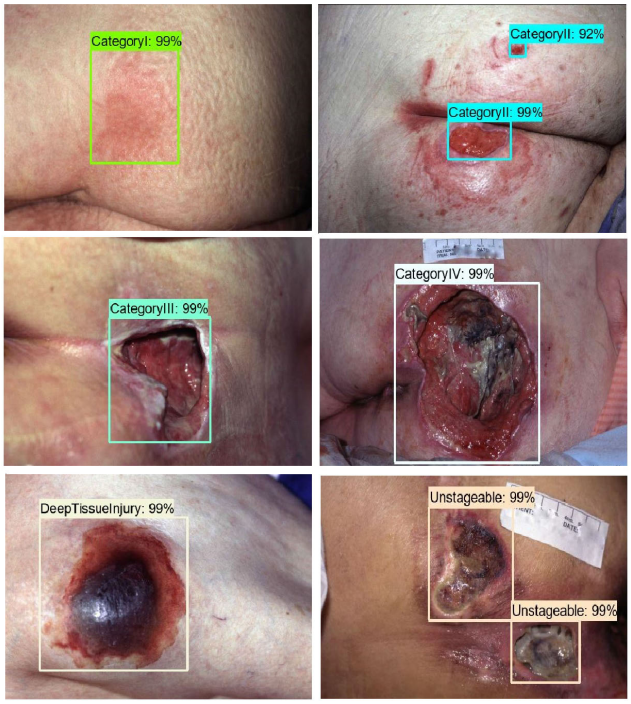

FIGURE 1. From top left to botton right you can see the classifications made during the trial for Category I, Category II, Category III, Category IV, Deep Tissue Injury and Unstaegable.

Pressure ulcers often occur on a) the ischial region (buttocks) typical for chair-bound patients, b) the back of the heal - in the supine position, c) the sacrum - in the supine position and d) the trochanteric region - in the lateral position [10]. When the surface of the skin is intact but reacts to injury by becoming red and hyperaemic, this is classed as a Category I. Category II ulcers occur in the epidermis and dermis layers where they can become necrotic and cause skin cover deficiency. Category III ulcers involve subcutaneous tissue and Category IVs have lesions that penetrate underlying muscle or bone. Category III and IV ulcers often have substantial amounts of necrotic tissue deep within the wound cavity. DTIs appear underneath intact skin and present themselves as deep bruises, which can deteriorate into a deep pressure ulcer. Unstageable wounds have an undetermined level of tissue damage covered with slough or eschar/necrotic tissue. Once an Unstageable pressure ulcer has been debrided, it can be categorised [11].

The National Institute for Care Excellence (NICE) provide guidelines for pressure ulcer risk assessment and prevention [12]. Clinicians use the NHS Safety Thermometer incident reporting system and the Strategic Executive Information System to document pressure ulcer incidents in the UK [13]. However, there is significant variation in their implementation and use [14], [15]. This is primarily because healthcare practitioners have a) varied experience in dealing with pressure ulcers, and b) a general lack of understanding of pressure ulcers and the treatment thereof [16]. The challenge is to provide a decision-support tool for healthcare practitioners that standardises pressure ulcer categorisation and reporting and makes pressure ulcer management more accessible to a wider group of healthcare professionals.

To address this challenge, we present a pressure ulcer management system that uses a Faster Region-based Convolutional Neural Network (Faster R-CNN) [17] and a mobile platform to automatically categorise and report pressure ulcers. We train a Faster R-CNN with a custom dataset of images to detect Category I, II, III, and IV pressure ulcers, DTIs, and Unstageable pressure ulcers that cannot be categorised in real-time. The proposed system does not replace human assessments, but enhances clinical practice, prevents diagnostic errors, and standardises how clinicians analyse and report pressure ulcers. The approach implements existing deep learning technologies to provide a novel end-to-end pipeline for pressure ulcer categorisation that works in ad hoc domiciliary settings. The strength of the approach resides within MLOPS [18], model deployment at scale, and its in-situ operation. The current pressure ulcer systems that exist do not automatically classify and report pressure ulcers from a service users' residential home in real-time. To the best of our knowledge, this is the first time clinicians have evaluated a deep learning end-to-end solution for automatic pressure ulcer categorisation and reporting in a service users' domiciliary setting.

The structure of the remainder of the paper is as follows. Section II discuses traditional automated image analysis. Section III discusses the relevant works undertaken in deep learning image analysis before Section IV introduces

the methodology for the proposed solution in this paper. Section V presents the results and Section VI discusses the findings, before Section VII concludes the paper and presents future work.

II. TRADITIONAL AUTOMATED IMAGE ANALYSIS

Automated medical image analysis has been an active area of research since computers digitised and processed scans. Between 1970 and 1990, clinicians used edge and line detector filters to analyse images. For example, snakes active contour models (ACM) where often implemented to perform segmentation in [19]. Later, clinicians used the approach in leg ulcer studies with piecewise B-spline arcs to adaptively initialise the ACMs [20].

Region-based approaches, also known as similarity-based segmentation, appeared in the late 1990s. Both [21] and [22] used this approach to build colour histogram models, and with Bayesian inference, were able to compute the posterior membership probability of pixels belonging to segments in a pressure ulcer image. By assigning pixels to different segments, the authors deconstructed ulcers to measure the wound and its constituent tissue.

Other image processing approaches include a) spectral clustering [23] which finds segments in images using morphological operators [24], b) relationship modelling between density and pixel intensity using synthetic frequencies extracted with contrast changes and energy density models [25], and c) toroidal geometry, where images over multiple contrast levels and varying synthetic frequencies are segmented with the method described in [26].

The focus moved from 2D to 3D image processing in the late 1990s with the introduction of the Measurement of Area and Volume Instrument System (MAVIS) [27]. MAVIS constructs three-dimensional mappings of pressure ulcers by projecting parallel stripes of alternating colours onto the region of interest. Clinicians then compute the volume of the ulcer using cubic spline interpolation. Similar 3D image processing approaches in [28] and [29] constructed 3D models of wounds by matching calibrated images captured from different angles and Stereoscopic 3D reconstruction.

While these approaches have proved to be useful in controlled environments, there use in domiciliary settings, where most pressure ulcers develop, less so due to the need for costly and complex lighting, specialised devices, and qualified staff trained to use the systems.

III. DEEP LEARNING IMAGE ANALYSIS

In the 1990s, scientists developed machine learning algorithms to perform semantic segmentation, data fitting, and statistical classification using image-specific features [30], [31]. Applications were primarily in the medical domain were clinicians manually performed feature extraction [32]. Today, Convolutional Neural Networks (CNNs) extract features from images automatically [33], [34]. In fact, since AlexNet (a CNN architecture) [35], DL has replaced most traditional image processing approaches given their ability to solve complex image processing problems.

In pressure ulcer studies there are some notable works. For example, [36] proposed a system that classifies tissue types and performs segmentation using CNNs. However, like the studies in [37] and [38], ML developers train models with low-quality images which have limited utility in complex wound analysis where clinicians often require high-resolution imagery. The challenge is getting high-resolution images which is fundamentally important for successfully training deep learning models [39]. At the time of writing, the Medetec dataset [40] is the most comprehensive open-source pressure ulcer dataset which contains 175 low resolution images of pressure ulcers - an insufficient number for training CNN models.

A common way to deal with this issue is to use the checkpoints of models trained on a large corpus of images and fine-tune them with images contained in smaller datasets (a technique known as transfer learning) [41]. This is an accepted method given that smaller organisations do not have the data or the compute to generate large-scale models - GPT-3 was pre-trained on 45 TB of text data with supercomputers (285,000 CPUs and 10,000 GPUs) [42]. Studies that do not have access to large datasets or compute use transfer learning in this way. For example, [43] fined-tuned a pre-trained model with a small dataset of pressure ulcer images to segment wounds and detect infection. Scientists discuss similar transfer learning approaches in [44], [45], and [46].

Obviously, the quality of the data and the type or problem clinicians want to solve informs DL practitioners on what type of DL architecture to use. Faster RCNNs are heavyweight detectors trained by scientists with a large corpus of high-quality data. Other less intensive models (in terms of data and compute requirements) for pressure ulcer analysis have been proposed. For example, scientists in [47], trained a single shot detector (SSD) based on the Mobilenet V2 Object Detection Model with low-resolution infrared thermography images [48]. While the results reported in the paper for training (confidence level=96-100%) are encouraging, the paper does not show any evaluation results for inference in a real-world setting. This is a major weakness of the paper which detracts from the fact results drop by thirty percent when clinicians use models in real-world environments due to the ad hoc nature of environmental conditions (devices, operational use, space, temperature, and ambient light). This is academic as the major limitation with this approach is the fact the model only has a "PU" class (hence the good training results) which severely limits its clinical utility when clinicians require automatic report generation and progress monitoring of distinct categories [49]. We found similar problems in other studies reported in the literature [50], [51], [52], [53]. When clinicians need complex medical image analysis DL practitioners use more advanced models, such as the Faster RCNN [54], [55], [56].

There is also a large body of work that uses deep learning to segment pressure ulcers in images [57], [58], [59], [60]. Segmentation models use object detection to first identify objects of interest and identify the class before mask algorithms segments objects into constituent parts. During the tagging process candidate objects are tagged using binding boxes, followed by pixel-level masking [61]. This is an important aspect of pressure ulcer analysis that allows clinicians to analyse and measure the constituent components of a pressure ulcer (i.e., granulating tissue, eschar, and slough). This paper does not consider segmentation, but it is something we will consider in future work once a sufficient pressure ulcer detector is developed. The scope of this paper is to automatically categorise pressure ulcers for the purpose of assessment and report standardisation.

A common aspect missing in the studies reviewed in this paper is an evaluation protocol to assess the model's usefulness in a clinical trial setting. Studies rarely report results beyond training and validation (i.e., mAP and IOU at .50 and .75 and precision/recall for small, medium, and large objects) [62], [63], [64], [65]. Clinicians need to evaluate the utility of the trained model in the settings they work in. What you tend to find is that the data used to train the model is significantly different to the data produced in clinical practice. As such the performance of the model significantly decreases. Running a clinical trial with the NHS is challenging, particularly when the technological solution is still in the initial stages of development. However, trials are important as they allow you to fully understand the strengths and weakness of the system through independent evaluation in a real-world clinical setting and continually re-train models to improve accuracy and clinical utility over time – the studies reviewed fail to report this critical aspect of on-going model training and evaluation.

IV. METHODOLOGY

This section describes the data collection strategy used in the study. The article discusses how we use a custom dataset, image augmentation, and transfer learning to train a Faster R-CNN model for categorising pressure ulcers. The paper also delves into the integration of the model into a mobile platform, used by clinicians in the clinical trial. The section concludes by presenting a set of evaluation metrics for assessing the model's performance during the training and inference stages of the clinical trial.

A. DATA COLLECTION AND PRE-PROCESSING

The Medetec pressure ulcer dataset provides a baseline image set in this study which contains 174 images of pressure ulcers (classes included are Category I, II, III, and IV, DTI and Unstageable) [40]. We added an additional 675 images (across the same classes) acquired from Google Images to the Medetec dataset. Images were used based on the following inclusion criteria: a) they have a minimum width and height of 600 pixels by 400 pixels to align picture quality with the quality of the images contained in the Medetec dataset (note we do not consider these to be high-quality images - but these were the only open access images we could obtain); b) they complement the images in the Medetec dataset where specific categories do not exist or are poorly represented; and c) they were not a duplicate of any existing image already included in the dataset. The Python Augmentor tool generates additional images by flipping, scaling, tilting, and rotating the 858 images. Each image is resized with a fixed ratio of 1024 by 1024 to match the input resolution of the Faster RCNN network. A district nurse with expertise in pressure ulcer categorisation tagged each pressure ulcer in the dataset as one of the six pressure ulcer classes - a total of 5084 objects in 4290 images: 685 tags for Category I, 1401 tags for Category II, 432 tags for Category III, 740 tags for Category IV, 899 for DTI and 927 for Unstageable. Figure 2 shows the class distribution.

It is clear to see from Figure 2 that there is a class imbalance problem in the study. Due to the small number of images collected from Medetec and Google images it was extremely difficult to appropriately deal with class balance in a sensible way. We could not under sample because we already had little data to train with. Therefore, the only option was to oversample using augmentation but there is only so much we could do with the small number of base images we had. This is because after a certain point augmentation stops introducing any additional variance. Nonetheless, despite the significant difficulties we had and the effort we put into this study, the model still performs reasonably well. The district nurse used Labelme to place binding boxes around objects to identify regions of interest. We export the tagged regions in each image as Extensible Mark-up Language (XML) in TensorFlow Pascal VOC format [66] which we later convert to Comma Separated Values (CSV) using Pandas and XML. Following a train and validation dataset split on the tagged classes, we use the Tensorflow Object Detection API and Pillow to convert the XML and associated images into TFRecords for training.

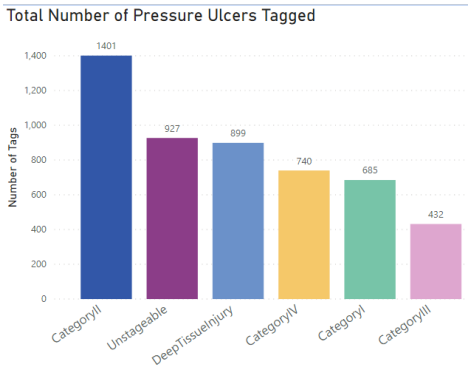

FIGURE 2. Class Distributions for the Tagged Dataset.

B. FASTER REGION-BASED CONVOLUTIONAL NEURAL NETWORK

The platform uses the Faster R-CNN architecture for object detection and classification on images containing pressure

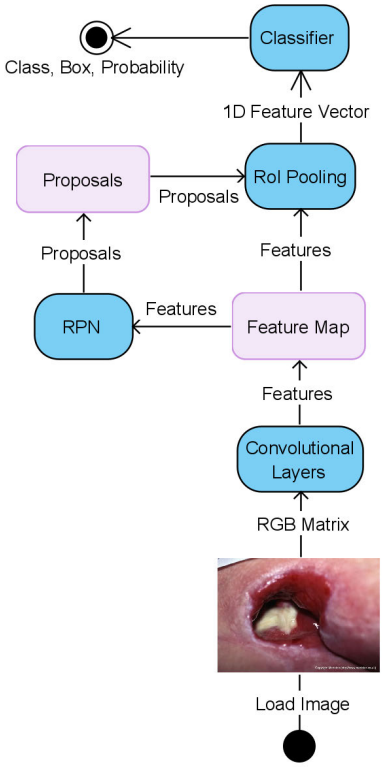

FIGURE 3. Faster R-CNN.

ulcers [67]. It has three parts: a) a CNN for classification and feature map generation, b) a region proposal network (RPN) for generating Regions of Interest (RoI), and c) a regressor, which finds the locations of each object and its classifications. Figure 3 provides an overview of the network architecture. The RPN identifies candidate pressure ulcer categories in photographs using previously learnt features in the base network (ResNet101 in this instance). The RPN replaces the selective search approach used in early R-CNN networks where the model generated region proposals at the pixel level rather than the feature map level. The RPN finds bounding boxes in the image using nine size and aspect ratios as shown in figure 4.

The size and aspect ratio configurations describe anchors (fixed bounding boxes) placed throughout the image. The RPN references the anchors to predict object locations. The RPN is a CNN, which uses the feature map provided in the base network to find a set of anchors of interest in an image. Note that the feature map dimensions are the same as those in the original image.

The RPN generates two outputs for each anchor bounding box a) a probability objectness score and b) a set of bounding box coordinates. The first output is a binary classification, the second a bounding box regression adjustment. During the training process, all the classified anchors are placed into one of two categories a) foreground: anchors that overlap the ground-truth object with an Intersection over Union (IoU) bigger than 0.5, or b) background: anchors that do not overlap any ground truth object or have less than a 0.1 IoU with ground-truth objects. The IoU is defined as:

$$IoU = \frac{Anchor\ box \cap Ground\ Truth\ box}{Anchor\ box \cup Ground\ Truth\ box} \tag{1}$$

Anchors are randomly sampled to create mini-batches with 256 balanced foreground and background anchors. Each batch is used to calculate the classification loss using binary cross-entropy. Anchors marked as foreground in the mini-batch are used to calculate the regression loss and the correct Δ to transform the anchor into the object. If no foreground anchors are found foreground anchors are selected that have the greatest IoU with overlapping ground truth objects. This ensures that foreground samples and targets are provided for the network to learn from rather than having no anchors at all.

Anchors will overlap; therefore, proposals will also overlap on the same object. Non-Maximum Suppression (NMS) is performed to delete intersecting anchor boxes with lower IoU values. IoU values greater than 0.7 describe positive object detection and values less than 0.3 describe background objects. Caution is required when setting the IoU threshold as setting it to low will result in proposals for objects being missed; too high and there will be too many proposals for the same object. It is typical to use 0.6 for the IoU threshold. The top N proposals, sorted by score, are selected after applying NMS. The loss functions for both the classifier and bounding box calculation are defined as:

$$L_{cls}(p_i, p_i^*) = -(p_i^* log(p_i) + (1 - p_i^*) log(1 - p_i)) \tag{2}$$

$$L_{reg}(t_i, t_i^*) = \Sigma_{i \in \{x,y,w,h\}} smooth_{L1}(t_i - t_i^*) \tag{3}$$

where

$$smooth_{L1}(t_i - t_i^*) = \begin{cases} 0.5x^2 & if\ |t_i - t_i^*| < 1 \\ |x| - 0.5 & 0ther \end{cases} \tag{4}$$

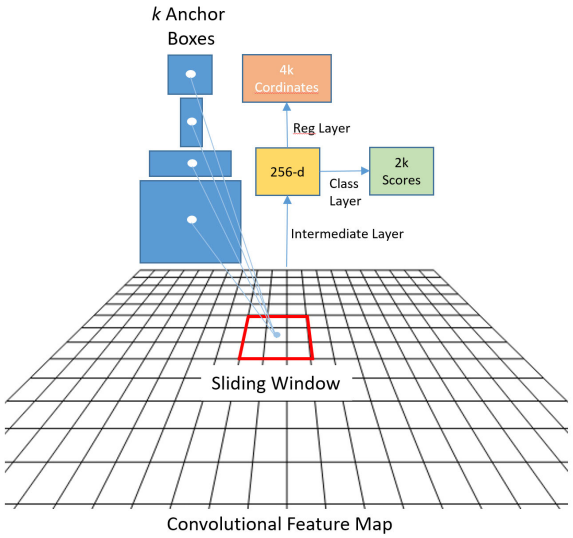

FIGURE 4. Region Proposal Network.

pi the object possibility, *ti* the 4k anchor coordinate, *pi** the ground truth label, *t** the ground truth coordinate, L_{cls} the classification loss (log loss), and L_{reg} the regression loss (smooth L1 loss)

Once the RPN step has completed there will be a set of object proposals. At this stage, the proposals do not have a class assigned to them. Each bounding box must be classified and assigned a category. In the Faster R-CNN implementation, the convolutional feature map is cropped using each proposal. Each crop is then resized to 14 * 14 * convdepth using interpolation. After cropping, max pooling with a 2 × 2 kernel is used to get a final 7 * 7 * 512 feature map for each proposal (via RoI Pooling). These dimensions are default parameters set by the Fast R-CNN; however, they are customizable depending on second stage use.

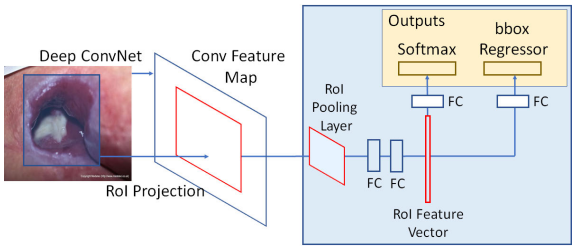

FIGURE 5. Fast R-CNN.

The Fast R-CNN takes the 7 ∗ 7 ∗ 512 feature map for each proposal, flattens it into a one-dimensional vector and connects it to two fully-connected layers of size 4096 with Rectifier Linear Unit (ReLU) activation. An additional fully-connected layer to identify object classes is implemented where *N* describes the total number of classes and *+1* the background. In parallel, a second fully-connected layer with *4N* units is implemented for bounding box regression prediction. The 4 parameters correspond to Δ_{center_x}, Δ_{center_y}, Δ_{width}, Δ_{height} for each of the *N* possible classes. Figure 5 describes the Fast R-CNN architecture. Targets in a Fast R-CNN are calculated in a similar way to the RPN targets but with different possible classes taken into account. Proposals and ground-truth boxes are used to calculate the IoU between them. Proposals with an IoU greater than 0.5 when compared with any ground truth box get assigned to that ground truth. Proposals with an IoU between 0.1 and 0.5 are assigned to the background. Proposals with no intersection are ignored. Targets for bounding box regression can then be calculated by determining the offset between the proposal and its corresponding ground-truth box. Note this only happens for proposals that have been assigned a class based on the IoU threshold. The Fast R-CNN is trained using backpropagation and Stochastic Gradient Descent. Calculating the loss function in the Fast R-CNN is defined as:

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda . [u \geq 1] L_{reg}(t^u, v) \quad (5)$$

where *p* describes the object possibility, *u* the classification class, *t* the ground truth label, *v* the ground truth coordinates for class *u*, L_{cls} the Loss function for classification, L_{reg} the Loss function for the bounding box regressor, and θ the balancing parameter. The L_{cls} is defined as:

$$L_{cls}(p, u) = -\log\left(\frac{e^{p_u}}{\sum_{j=1}^{K} e^{p_j}}\right) \quad (6)$$

where *p* is the object possibility, *u* the classification class, L_{cls} the Loss function for classification and *K* the number of classes. L_{reg} can be calculated using the equation described in 5 with t^u and *v* as input.

Following object classification, bounding box adjustments are performed. This is achieved by taking into account the class with the highest probability for that proposal. Proposals that have a background class assignment are ignored. Using the final set of objects class-based NMS is applied and, to minimise the final set of objects returned, a probability threshold is set.

Putting the complete model together there are two losses for the RPN and two for the R-CNN. The four losses are combined using a weighted sum to give classification losses more weight relative to regression losses, or give R-CNN losses more power over the RPNs'.

C. TRANSFER LEARNING

Transfer learning is adopted to fine-tune a pre-trained model using the six pressure ulcer classes in our dataset. This is an important technique as training CNNs on small datasets (which we have in this study) leads to extreme overfitting due to low variance. The base model is the residual neural networks-101 (Resnet101) model [68]. It has been pre-trained using the COCO dataset which contains 330 thousand images and 1.5 million object instances. Residual neural networks are deep neural networks based on a highway networks architecture [69]. They accelerate training in very deep neural networks and using skip connectors, avoid vanishing and exploding gradients. We do not claim any novelty in either the Faster RCNN or the transfer learning aspects but rather use them as a component in a novel end-to-end platform for automatically categorising and reporting pressure ulcers in domiciliary settings.

D. MODEL TRAINING

Model training is performed on an HP ProLiant ML 350 Gen 9 Server with x2 Intel Xeon E5-2640 v4 series processors, 768GB of RAM and four NVidia Quadro RTX8000 graphics cards with a combined 192GB of GPU memory. TensorFlow 2.2, TensorFlow Object Detection API, CUDA 10.2 and CuDNN version 7.6 are used in the training pipeline. In the TensorFlow pipeline.config file the following hyper parameters are set:

- To maintain aspect ratio resizer minimum and maximum coefficients are set to 1024 × 1024 pixels respectively. This minimises the scaling effect on the acquired data.
- The default setting for the feature extractor coefficient is retrained to provide a standard 16-pixel stride length to maintain a high-resolution aspect ratio and improve training time.

- The batch size coefficient is set to thirty-two to maintain GPU memory limits.
- The learning rate is set to 0.0004 to prevent large variations in response to the error.

In order to improve generalisation and to account for variance in the camera trap images the following augmentation settings were used:

- Random_adjust_hue which adjusts the hue of an image using a random factor.
- Random_adjust_contrast which adjusts the contrast of an image by a random factor.
- Random_adjust_saturation which adjusts the saturation of an image by a random factor.
- Random_square_crop_by_scale which was set with a scale_min of 0.6 and a scale_max of 1.3.

The Adam optimizer is implemented in Resnet 101 to minimise the loss function [70]. Unlike optimisers that maintain a single learning rate (alpha) throughout the entire training session (stochastic gradient descent), Adam calculates the moving average of the gradient m_t/squared gradients v_t and the parameters beta1/beta2 to dynamically adjust the learning rate. Adam is defined as:

$$\begin{aligned} m_t &= \beta_1 m_t - 1 + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_t - 1 + (1 - \beta_2) g_t^2 \end{aligned} \tag{7}$$

where m_t and v_t are estimates of the first and second moment of the gradients. Both m_t and v_t are initialised with 0's. Biases are corrected by computing the first and second moment estimates:

$$\begin{aligned} \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \end{aligned} \tag{8}$$

Parameters are updated using the Adam update rule:

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t. \tag{9}$$

The ReLU activation function is adopted to overcome the saturation changes around the mid-point of their input which is a common problem with sigmoid or hyperbolic tangent (tanh) activations [71]. ReLU is defined as:

$$g(x) = max(0, x) \tag{10}$$

E. CLINICAL TRIAL PROTOCOL

TensorFlow serving hosts the trained pressure ulcer model [72]. District nurses in the study use iOS and Android mobile devices over 4/5G communications to transmit photographs of pressure ulcers through a WordPress web interface hosted on Apache. A Rest-API submits photographs received server-side to TensorFlow Serving [73] for classification. A MySQL database on the server stores the URLs to classified images on disk. Clinicians can view classification results in the WordPress gallery 2-3 seconds after the photograph is taken. A custom-built server containing an Intel Xeon E5-1630v3 CPU, 64GB of RAM and an NVidia Tesla T4 GPU. TensorFlow 2.2, CUDA 10.2 and CuDNN 7.6 is used to inference the model. The taking of photographs did not impact service users or district nurses beyond normal clinical practice. The study ran between the 15th of March 2021 and the 21st of December 2021. Throughout the trial, specialist nurses reviewed the classifications made and either confirmed the category(s) was correct or reported what the correct category should be. Poor quality images, images with patient or nurse identifiable information, and images that did not contain pressure ulcers were removed from the study.

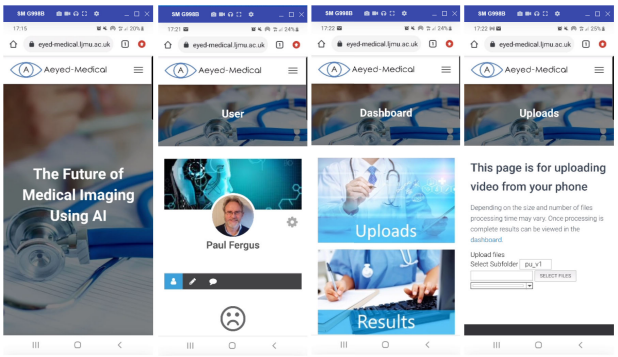

FIGURE 6. PUMS Mobile Phone Web Services Interface.

F. EVALUATION METRICS

The model's performance during training is evaluated using RPNLoss/objectiveness, RPNLoss/localisation, BoxClassifierLoss/classification, BoxClassifierLoss/localisation and TotalLoss. These metrics are collected from Tensorboard 2.6. The RPNLoss/objectiveness measures how well the model can generate suitable bounding boxes and categorise them as either a background or foreground object. RPNLoss/localisation measures how well the RPN is at generating bounding box regressor coordinates for foreground objects. In other words, how far each anchor target is from the closest bounding box. BoxClassifierLoss/classification measures the output layer/final classifier loss and describes the computed error for prediction. BoxClassifierLoss/localisation measures the performance of the bounding box regressor. All these measures are combined to produce a total loss metric.

The validation set during training is measured using mAP (mean average precision), which is a standard metric for evaluating the performance of an object detection model. mAP is defined as:

$$mAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q} \tag{11}$$

where Q is the number of queries in the set and $AveP(q)$ is the average precision (AP) for a given query a.

The mAP is calculated on the binding box locations for the final two checkpoints. IoU thresholds @.50 and @.75 are used to assess the overall performance of the model. This is achieved by measuring the percentage ratio of the overlap

between the predicted bounding box and the ground truth bounding box and is defined as:

$$IoU = \frac{AreaofOverlap}{AreaofUnion} \tag{12}$$

A threshold of @.50 measures the overall detection accuracy while the upper threshold of @.75 measures localisation accuracy.

Using the final trained model, inference is measured using photographs taken during the clinical trial to evaluate the performance of the model in a real-world situation. Inference is evaluated using Precision, Recall, F1-Score and Support. Precision is defined as:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

Recall is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

F1 Score is defined as:

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{15}$$

Support is used to describe the number of samples of the true response that reside within specific classes in the test set (the number of pressure ulcers in images obtained from the clinical trial).

The ground truths for images taken by nurses during the trial are provided by clinical staff at Mersey Care NHS Foundation Trust and used to calculate the detections generated by the in-trial model. Precision, Recall, F1-Score and Support are calculated with IoU@.50 for all experiments and confidence scores (CS) @.30, @.50, @.75, @.90. These metrics are used to provide an overall assessment for each class in the model during clinical trial inference. The experiments also report all false positives that reside outside of the IoU@.50 threshold at each of the four CS thresholds. The precision-recall receiver operator curve (ROC) is used to visually represent the cutoffs and the area under the curve (AUC).

V. EVALUATION

The results obtained during the training of the Faster RCNN model are presented first. This is followed by two additional evaluations to determine how well the trained model performs in a clinical setting. The first evaluates the model's ability to classify pressure ulcers in the photographs taken by district nurses. The second evaluates the same photographs cropped to only include the pressure ulcer (to remove noise and unnecessary information and to increase the size of the pressure ulcer).

A. TRAINING RESULTS FOR MODEL TRAINED ON THE MEDETEC AND GOOGLE DATASET

In the first experiment, the training set (Medetec and Google scrapped images - 4291 in total) are used to fit the model. Note this is a pre-trained Faster RCNN model fined tuned using a dataset containing the images from the Medetec pressure ulcers dataset and pressure ulcer images scrapped from Google images. The dataset is randomly split into training (90%), and validation (10%). The model is trained over 25000 steps (781 epochs) using a batch size of 32.

1) RESULTS FOR TRAINING DATASET

The results in Table 1 indicate that the model is generally good at producing candidate regions of interest (0.0593). The results also show that the RPN can effectively perform localisation on the objects identified (0.0598). The classification loss is higher (0.2015) than all other losses indicating the model is much less accurate at classifying identified objects of interest. This will correlate with the results presented for inference later in this section. In terms of box classifier localisation (0.0564), this is much more in line with the results produced by the RPN and shows that placing binding boxes around objects is not a real issue for the model. Table 1 shows the total loss (0.3770) for both the RPN and Box Classifier which is considered a good loss in object detection.

TABLE 1. Tensorboard Results for Training.

Metric	Smoothed	Value
RPNLoss/objectness	0.0593	0.0521
RPNLoss/localisation	0.0598	0.0103
BoxClassifierLoss/classification	0.2015	0.0622
BoxClassifierLoss/localisation	0.0564	0.0240
Total Loss	0.3770	0.1486

2) RESULTS FOR VALIDATION DATASET

Table 2 provides the detection boxes' mAP metrics across several configurations. mAP provides the mean average precision over all classes averaged over IoU thresholds ranging between .5 and .95 with .05 increments. Precision (0.7743) is relatively good indicating a reduced number of false positives. The three metrics for large, medium and small objects indicate the model is better at detecting large and medium objects in images rather than smaller ones. mAP @.50IoU is the mean average at 50% IoU and mAP @.75IoU is precision at 75% IoU. The results in Table 2 show that the best precision values are mAP (Large)=0.8045 and mAP@.50=0.9732. Utilising large objects and the mAP@.50 IoU threshold will minimise the number of false positives returned. In other words, the validation results suggest that the model will perform reasonably well with large/medium objects and less so with smaller objects. This means that photographs of pressure ulcers will need to be taken close to the actual wound. Table 3 provides the detection boxes AR metrics across the same configurations used to calculate Precision. AR@1 provides the average recall with 1 detection, AR@10 is the average recall with 10 detections and AR@100 is the average recall with 100 detections. Recall in this instance represents the number of ground truths detected divided by the total number of ground truths that exist. A significant jump is seen between 1 detection and 10 detections but little change between 10 and 100. Again, the results are reasonably

good when 10 or more detections are returned (0.8221 and 0.8249). In this instance, the results suggest that most of the ground truths presented were detected by the trained model. The recall values for AR@100(small, medium and large) show the average recall with 100 detections across small, medium and large objects in images. Again, the best results are obtained when large and medium objects in images are present (0.8496-0.7819) and less so for small objects (0.4212).

TABLE 2. Tensor Board Results for Eval - Precision.

Metric	Smoothed
DetectionBoxes/Precision/mAP	0.7743
DetectionBoxes/Precision/mAP (Large)	0.8045
DetectionBoxes/Precision/mAP (Medium)	0.7380
DetectionBoxes/Precision/mAP (Small)	0.1620
DetectionBoxes/Precision/mAP@.50IOU	0.9732
DetectionBoxes/Precision/mAP@.75IOU	0.9119

B. CLINICAL TRIAL RESULTS USING TRAINED MODEL

The trained model was deployed and used in the clinical trial to analyse pressure ulcer photographs taken by district nurses during routine patient visits. During the trial, 1016 images were collected. Following quality checking, this number was reduced to 624 by removing blurry images, images that contained identifiable patient or staff information, and images that did not contain pressure ulcers. A second review was performed to remove images that were similar (the same pressure ulcer taken repeatedly during the trial with little variance). The final test set contained 216 images (5 Category I images, 93 Category II, 11 Category III, 0 Category IV (none were seen during the trial), 30 DTI, and 77 unstageable) as shown in Figure 7.

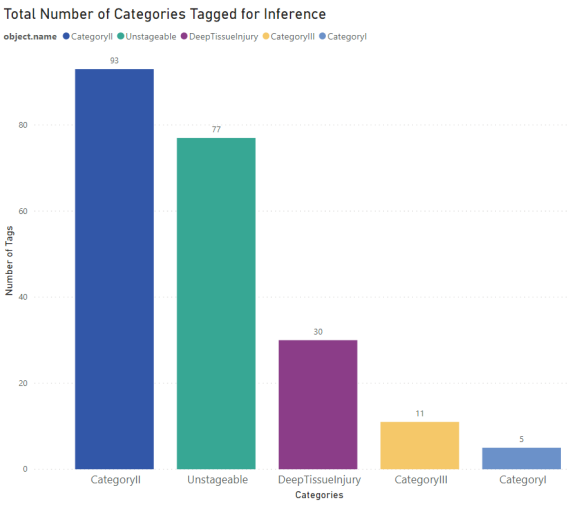

FIGURE 7. Inference Class Distribution.

1) INFERENCE USING UNCROPPED IMAGES

The 216 photographs are evaluated class-by-class using Precision and Recall, with IoU@.50 and CS @.30, @.50, @.75 and @.90. The F1-score is used to calculate the harmonic mean between the Precision and Recall values. The Support for each class is Category I=5, Category II=93, Category III=11, Category IV=0, Unstageable=77, and DTI=30. The number of false positives that reside outside of IoU@.50 are also provided. Table 4 shows the performance metrics using a CS @0.30. Note the mean average is divided by 5 as no Category IV pressure ulcers were seen during the trial.

In this evaluation, 109 false positives were reported. The results overall were poor across all classes. The best performing class was unstageable - the worse was category III which is reasonable considering support was only 11 and this category had the smallest number of tags for training (432).

The mAP for all classes was 0.4516 and 0.3900 for mAR. The F1-Score was reported as 0.3889. Increasing the CS threshold to @.50 does improve the results slightly however most Precision-Recall values are below 0.50 as shown in Table 5. The @.50 threshold does however significantly reduce the number of false positives from 109 to 46. Increasing the CS further to @.75 fails to balance the Precision-Recall values and suggests that a CS @.50 is the most optimal configuration for this model as shown in Table 6. Setting the CS to @.75 reduces the false positives to 16 in line with the higher precision values but decreases the model's ability to recall a sufficient number of ground truths.

TABLE 3. Tensorboard Results for Eval - Recall.

Metric	Smoothed
DetectionBoxes/Recall/AR@1	0.7308
DetectionBoxes/Recall/AR@10	0.8221
DetectionBoxes/Recall/AR@100	0.8249
DetectionBoxes/Recall/AR@100 (Large)	0.8496
DetectionBoxes/Recall/AR@100 (Medium)	0.7818
DetectionBoxes/Recall/AR@100 (Small)	0.4212

TABLE 4. Faster R-CNN Inference Results Using Uncropped Images with IOU@.50 CS@.30.

Class	Precision	Recall	F1-Score
CategoryI	0.2222	0.4444	0.2962
CategoryII	0.3555	0.3609	0.3581
CategoryIII	0.2400	0.3750	0.2926
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.6785	0.4222	0.5205
DTI	0.7619	0.3478	0.4775
Mean Average	**0.4516**	**0.3900**	**0.3889**

TABLE 5. Faster R-CNN Inference Results Using Uncropped Images with IOU@.50 CS@.50.

Class	Precision	Recall	F1-Score
CategoryI	0.3333	0.4444	0.3809
CategoryII	0.4105	0.2932	0.3420
CategoryIII	0.3571	0.3125	0.3333
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.7200	0.4000	0.5142
DTI	0.8500	0.3695	0.5150
Mean Average	**0.5341**	**0.3639**	**0.4170**

Setting the CS to @.90 further decreases the number of false positives to 4 but the overall Recall in many classes is again reduced as shown in Table 7.

TABLE 6. Faster R-CNN Inference Results Using Uncropped Images with IOU@.50 CS@.75.

Class	Precision	Recall	F1 Score
CategoryI	0.4285	0.3333	0.3749
CategoryII	0.4626	0.2330	0.3099
CategoryIII	0.3846	0.3125	0.3448
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.7500	0.3666	0.4924
DTI	0.8750	0.3043	0.4515
Mean Average	**0.5801**	**0.3099**	**0.3947**

TABLE 7. Faster R-CNN Inference Results Using Uncropped Images with IOU@.50 CS@.90.

Class	Precision	Recall	F1 Score
CategoryI	0.7500	0.3333	0.4615
CategoryII	0.4545	0.1503	0.2258
CategoryIII	0.5555	0.3125	0.3999
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.8571	0.3333	0.4799
DTI	0.9285	0.2826	0.4333
Mean Average	**0.7091**	**0.2824**	**0.4000**

A) Precision-Recall Curve for Original Images @IOU.50 and @CS.50

The precision-recall ROC curve in Fig. 8 shows the overall model performance. The AUC values for Category I, II, III, IV, DTI and Unstageable was 0.4660, 0.6296, 0.1979, 0.0000, 0.6691 and 0.4914 respectively. It is clear from these results the model's performance is poor. This is partly due to the ad-hoc way in which photographs of pressure ulcers were taken which did not reflect the images used for training. While best practice advice was given to district nurses, photograph quality varied due to poor lighting, pressure ulcer site access, and the distance mobile phones were held from the wound site when photographs were taken.

2) INFERENCE USING CROPPED IMAGES

To mitigate the issues raised in the previous evaluation the 216 images were standardised by cropping them with a 1024 by 1024 aspect ratio. Figure 9 shows an example of the original images on the left and the classified cropped pressure ulcer images on the right (note these images where from the Medetec and Google dataset - they were not images of patients who participated in the trial).

The assumption is that removing noise will improve the overall localisation and classification results. In other words, zoom into the image and make the pressure ulcers appear bigger. In this evaluation, the performance of the model was measured using the same metrics. Table 8 provides the results for IoU@.50 and a CS @.30. Adopting this strategy improved the mean average for Precision, Recall and F1-Score but nothing significant beyond the previous set of results. The mean averages remain below 0.50 for both Precision and the

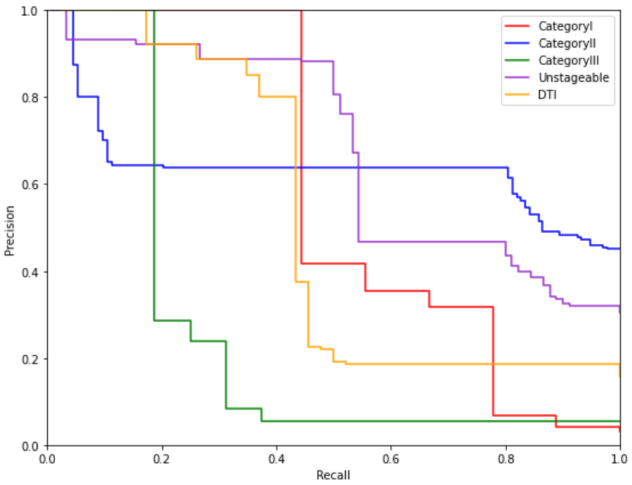

FIGURE 8. Model Trained on the Medetec and Google dataset with Uncropped Images.

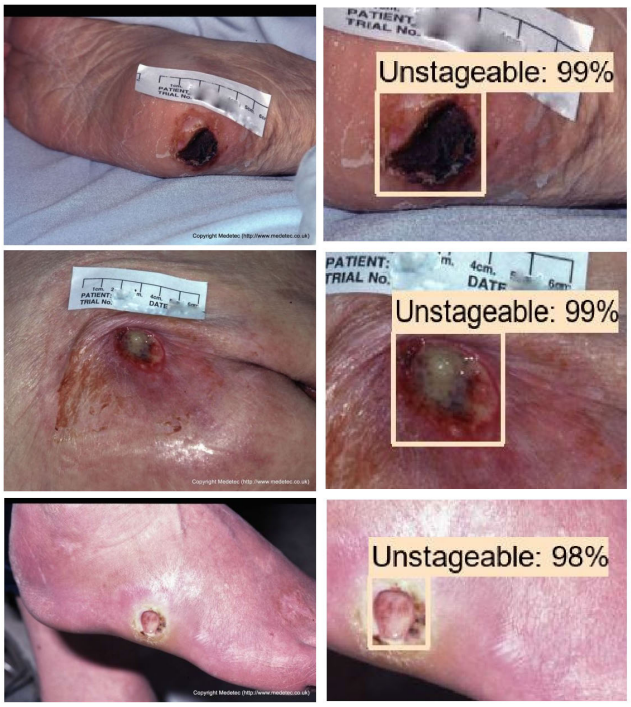

FIGURE 9. Example images represent original images on the Left and images on the right where the pressure ulcer has been cropped from the image to remove noise.

F1-Score however there is a marked increase in Recall which means more of the ground truths were detected. There was an increase in the number of false positives (152). Setting the CS to @.50 increased all mean average values above 0.50 as shown in Table 9. This time 93 false positives were reported. With a CS @.75 the mean average results for Precision, Recall and F1-Score increase further to just below .70 as indicated in Table 10 with 45 false positives reported. In the final experiment, there were further increases for precision (0.7762) and F1-Score (0.6956) however recall dropped from

TABLE 8. Faster R-CNN Inference Results Using Cropped Images with IOU@.50 and CS@.30.

Class	Precision	Recall	F1 Score
CategoryI	0.1666	0.8000	0.2757
CategoryII	0.3852	0.7096	0.4994
CategoryIII	0.1923	0.4545	0.2702
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.7037	0.7402	0.7214
DTI	0.7037	0.6333	0.6666
Mean Average	**0.4203**	**0.6675**	**0.4866**

TABLE 9. Faster R-CNN Inference Results Using Cropped Images with IOU@.50 and CS@.50.

Class	Precision	Recall	F1 Score
CategoryI	0.3076	0.8000	0.4443
CategoryII	0.5000	0.6666	0.5714
CategoryIII	0.2608	0.5454	0.3528
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.7500	0.7402	0.7450
DTI	0.7307	0.6333	0.6785
Mean Average	**0.5098**	**0.6771**	**0.5584**

TABLE 10. Faster R-CNN Inference Results Using Cropped Images with IOU@.50 and CS@.75.

Class	Precision	Recall	F1 Score
CategoryI	0.3750	0.6000	0.4615
CategoryII	0.6595	0.6666	0.6630
CategoryIII	0.5714	0.7272	0.6399
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.8378	0.8051	0.8211
DTI	0.9545	0.7000	0.8076
Mean Average	**0.6796**	**0.6997**	**0.6786**

TABLE 11. Faster R-CNN Inference Results Using Cropped Images with IOU@.50 and CS@.90.

Class	Precision	Recall	F1 Score
CategoryI	0.5000	0.6000	0.5454
CategoryII	0.7763	0.6344	0.6982
CategoryIII	0.6666	0.5454	0.5999
CategoryIV	0.0000	0.0000	0.0000
Unstageable	0.9384	0.7922	0.8591
DTI	1.0000	0.6333	0.7754
Mean Average	**0.7762**	**0.6410**	**0.6956**

0.6997 to 0.6410. As would be expected with a higher precision the number of false positives reported fell to 19.

A) Precision-Recall Curve for Cropped Images @IOU.50 and @CS.50

The Precision-Recall ROC curve in Fig. 10 shows the model's performance. This time the AUC values for category I, II, III, IV, DTI, and unstageable were 0.6253, 0.8552, 0.5051, 0.0000, 0.9299 and 0.8194 respectively. Compared with the results in Table 5 there was a 0.1593 improvement for category I, a 0.2256 improvement for category II, a 0.3072 improvement for category III, 0.2608 improvement for unstageable and a 0.3281 improvement for DTI.

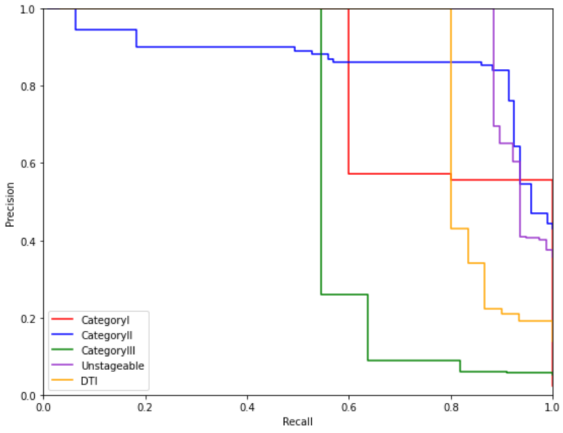

FIGURE 10. Model trained on the Medetec and Google dataset with cropped images.

VI. DISCUSSION

This paper presented an end-to-end platform that classifies and documents pressure ulcers automatically. The results demonstrated that the Faster R-CNN, trained on a custom set of pressure ulcer images, using its RPN was able to effectively detect objects and apply localisation with losses of 0.0593, and 0.0598 respectively. The BoxClassifierLoss was able to produce a similar loss for localisation (0.0564) but classification loss was higher (0.2015). Collecting the required pressure ulcer images for each class (typically 1500 objects per class is required when using transfer learning) proved to be difficult in this study as there are no publicly available datasets. Images from the internet were sourced but the quality and distribution between classes was poor.

Despite this limitation, the evaluation dataset achieved an overall mAP of 0.7743 which is considered a good result in object detection. Table 2 also showed the mAP results for large, medium, and small objects. Larger objects produced higher mAP values than smaller objects as you would expect (0.8045 and 0.1620 respectively). Therefore, close images of pressure ulcers produce better results than those taken at longer distances - a point we will return to later in the discussion. Similar results were reported for Recall as indicated in Table 3. Recall increased in line with the number of detections (AR@1 0.7308 and AR@100 0.8249) showing a strong correlation between large and small objects (0.8496 and 0.4212 respectively). Again, this suggested the model is better at recalling larger objects than smaller ones.

The trained model was evaluated in a clinical setting. The results from the first evaluation were disappointing and there are several reasons for this. First, the number of images collected from the trial was small with a significant imbalance across all classes. Category III performed the worst with an F1-Score of 0.2926. This is reasonable given that only eleven category III instances were recorded during the trial and only 432 tags were used in training. The best performing category was Unstageable with an F1-Score of 0.5205.

DTIs and unstageable (which are often larger in appearance) produced better results than smaller pressure ulcers

(category I and II) which are more difficult to analyse because of their size. This was in line with the training results discussed earlier. However, this did not fully explain the poor results. The images collected from Google and Medetec for training were pre-processed to maximise the appearance of a pressure ulcer in an image. In the trial, however, there was significant variance in how photographs were taken (i.e., distance and lighting). Larger representations of pressure ulcers were better detected (although in several instances they were miss-classified). Photographs of pressure ulcers taken at larger distances were often missed and recorded as a false negative.

To address this issue the 216 images were cropped to remove unwanted information. With the CS @.30, a marked improvement in both unstageable and DTI classifications was observed with F1-Scores of 0.7214 and 0.6666, respectively. With CS @.50 the results improved further with similar improvements @.75 and @.90. The best-balanced results reported was @.75 with category I=0.4615, category II=0.6630, category III=0.6399, category IV=N/A (note no Category IV pressure ulcers were seen during the trial hence the N/A value to indicate this category could not be evaluated), unstageable=0.8211 and DTI=0.8076. In comparison with the best results obtained from uncropped images (@.50 - mAP=0.5341, mAR=0.3639, mAF1=0.4170) and the results obtained @.75 in this evaluation (mAP=0.6796, mAR=0.6997, mAF1=0.6786), cropping the images significantly improved overall performance.

We accept the results are not clinically relevant. However, they are encouraging. The model was trained on poor-quality images obtained from the Internet and despite the limitations reported, we were able to develop a pressure ulcer categorisation and reporting system that produced reasonably good results. It is hoped that this evidence will convince clinical organisations that a better model could be developed if high-quality pressure ulcer images are openly shared with the research community. There are obviously several other issues that need to be addressed, particularly with the mobile app. For example, there needs to be a feature that can automatically zoom and crop a pressure ulcer - this is something that will address in future work.

VII. CONCLUSION AND FUTURE WORK

Pressure ulcers are a significant challenge for patients and healthcare professionals. While training and guidelines are given to assess, treat, and report their occurrence there are inconsistencies in the type of ulcers reported, data collection and classification systems used. This paper considered the issue and reported the results from a clinical trial conducted by Mersey Care NHS Foundation Trust who evaluated the efficacy of an automated pressure ulcer categorisation and reporting system. District nurses in the study took photographs of pressure ulcers using their mobile phones and transmitted them over a 4/5G network to servers at LJMU. A total of 1016 images were collected over eight months. This number was reduced to 216 following quality checks to remove blurry images, images that contain patient or staff identifiable information, images that did not contain pressure ulcers and images that looked similar.

While the results from the evaluation are encouraging the main challenge was getting access to a sufficient number of high-quality images with equal distributions across all categories. Empirically, we found that transfer learning requires a minimum of 1500 tagged objects per class to produce results in the 90s. This means that for the six classes we would need 9000 tags for training a new model. To achieve this a widespread push across all NHS trusts in the UK would be required which was beyond the scope of this study.

Nonetheless, given the challenges, we believe the results highlight the benefits of the end-to-end platform and its ability to detect, categorise, and report pressure ulcers. This contributes to the biomedical field and provides new insights into the use of deep learning and mobile platforms for pressure ulcer management that warrants further investigation. While work exists in the digital analysis of pressure ulcers using different machine learning methods, to the best of our knowledge the study in this paper is the first comprehensive NHS clinical trial of its kind that combines deep learning and an enterprise mobile platform to analyse, categories, and report pressure ulcers in real-time in domiciliary settings. The work builds on existing research where current methods are only capable of classifying a limited range of pressure ulcer conditions (usually the most visually distinctive) in very controlled environments.

In future work, the focus will be on obtaining NIHR funding to carry out a much larger study. There will also be a focus on understanding NHS data access policies and leveraging resources to obtain a much larger corpus of pressure ulcer images across the UK. Following sufficient imagery, we will also focus on segmentation to measure pressure ulcers and their constituent tissue types. Additional development work will be undertaken to help clinicians standardise the photography of pressure ulcers in the community to improve the predictive performance of the model.

We believe that the application of this technology also has huge potential in many other wound care settings. The mobile application makes this particularly attractive to wound care in countries where there is no NHS-level of service, i.e. in Africa (for example in Uganda healthcare is very inaccessible so applications like this with appropriate recommendations on how to treat wounds would be a welcomed intervention to the many poor people who live there). This is also an area we will be looking at in future work.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for providing excellent feedback during the review process—their comments have significantly helped to improve the quality of the article, also would like to thank the NIHR for adopting the study in their Clinical Research Network (CRN) portfolio and providing clinical support, also would like to thank Mersey Care NHS Foundation Trust for running the

clinical trial—in particular, Pauline Parker and Ann Bennett, and also would like to thank all the district nurses and the 50 patients that took part in the study and the specialist nurses that evaluated the detections generated by the AI models.

REFERENCES

[1] M. Stephens and C. A. Bartley, "Understanding the association between pressure ulcers and sitting in adults what does it mean for me and my carers? Seating guidelines for people, carers and health social care professionals," *J. Tissue Viability*, vol. 27, no. 1, pp. 59–73, Feb. 2018.
[2] *Pressure Ulcers: Revised Definition and Measurement*, London, U.K.: NHS Improvement, Jun. 2018.
[3] C. Dealey, J. Posnett, and A. Walker, "The cost of pressure ulcers in the United Kingdom," *J. Wound Care*, vol. 21, no. 6, pp. 261–266, Jun. 2012.
[4] J. F. Guest, G. W. Fuller, P. Vowden, and K. R. Vowden, "Cohort study evaluating pressure ulcer management in clinical practice in the U.K. following initial presentation in the community: Costs and outcomes," *BMJ Open*, vol. 8, no. 7, Jul. 2018, Art. no. e021769.
[5] P. Browning, "The house of lords debates wound care strategy," *J. Wound Care*, vol. 26, no. 12, pp. 707–711, Dec. 2017.
[6] R. White, D. Bennett, C. Bree-Aslan, and F. Downie, "Pressure ulcers, negligence and litigation," *Age and Ageing*, vol. 11, no. 1, pp. 8–14, 2015.
[7] P. Avsar, Z. Moore, and D. Patton, "Dressings for preventing pressure ulcers: How do they work?" *Wounds U.K.*, vol. 30, no. 1, pp. 33–39, Jan. 2021.
[8] K. Y. Woo, D. Beeckman, and D. Chakravarthy, "Management of moisture-associated skin damage: A scoping review," *Adv. Skin Wound Care*, vol. 30, no. 11, pp. 494–501, 2017.
[9] J. F. Guest, N. Ayoub, T. McIlwraith, I. Uchegbu, A. Gerrish, D. Weidlich, K. Vowden, and P. Vowden, "Health economic burden that wounds impose on the national health service in the U.K.," *BMJ Open*, vol. 5, no. 12, Dec. 2015, Art. no. e009283.
[10] T. Sato, T. Abe, and S. Ichioka, "Factors impairing cell proliferation in the granulation tissue of pressure ulcers: Impact of bacterial burden," *Wound Repair Regener.*, vol. 26, no. 3, pp. 284–292, May 2018.
[11] J. Schiffman, M. S. Golinko, A. Yan, A. Flattau, M. Tomic-Canic, and H. Brem, "Operative debridement of pressure ulcers," *World J. Surgery*, vol. 33, no. 7, pp. 1396–1402, Jul. 2009.
[12] G. Stansby, L. Avital, K. Jones, and G. Marsden, "Prevention and management of pressure ulcers in primary and secondary care: Summary of NICE guidance," *BMJ*, vol. 348, p. g2592, Apr. 2014.
[13] I. L. Smith, J. Nixon, S. Brown, L. Wilson, and S. Coleman, "Pressure ulcer and wounds reporting in NHS hospitals in England part 1: Audit of monitoring systems," *J. Tissue Viability*, vol. 25, no. 1, pp. 3–15, Feb. 2016.
[14] D. McCaughan, L. Sheard, N. Cullum, J. Dumville, and I. Chetter, "Patients' perceptions and experiences of living with a surgical wound healing by secondary intention: A qualitative study," *Int. J. Nursing Stud.*, vol. 77, pp. 29–38, Jan. 2018.
[15] M. Lumbers, "An overview of 'Pressure ulcers: Revised definition and measurement,'" *Brit. J. Community Nursing*, vol. 24, no. 5, pp. 216–223, May 2019.
[16] I. Cho, H.-A. Park, and E. Chung, "Exploring practice variation in preventive pressure-ulcer care using data from a clinical data repository," *Int. J. Med. Informat.*, vol. 80, no. 1, pp. 47–55, Jan. 2011.
[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[18] M. Treveil, N. Omont, C. Stenac, K. Lefevre, D. Phan, J. Zentici, A. Lavoillotte, M. Miyazaki, and L. Heidmann, *Introducing MLOps*. Sebastopol, CA, USA: O'Reilly Media, 2020.
[19] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, Jan. 1988.
[20] T. D. Jones and P. Plassmann, "An active contour model for measuring the area of leg ulcers," *IEEE Trans. Med. Imag.*, vol. 19, no. 12, pp. 1202–1210, May 2000.
[21] L. Bertelli, B. Sumengen, B. S. Manjunath, and F. Gibou, "A variational framework for multiregion pairwise-similarity-based image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1400–1414, Aug. 2008.
[22] F. J. Veredas, H. Mesa, and L. Morente, "Efficient detection of wound-bed and peripheral skin with statistical colour models," *Med. Biol. Eng. Comput.*, vol. 53, no. 4, pp. 345–359, Apr. 2015.
[23] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[24] D. M. Dhane, V. Krishna, A. Achar, C. Bar, K. Sanyal, and C. Chakraborty, "Spectral clustering for unsupervised segmentation of lower extremity wound beds using optical images," *J. Med. Syst.*, vol. 40, no. 9, p. 207, Sep. 2016.
[25] D. P. Ortiz, D. Sierra-Sosa, and B. G. Zapirain, "Pressure ulcer image segmentation technique through synthetic frequencies generation and contrast variation using toroidal geometry," *Biomed. Eng. OnLine*, vol. 16, no. 1, Dec. 2017.
[26] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
[27] P. Plassmann and T. D. Jones, "MAVIS: A non-invasive instrument to measure area and volume of wounds," *Med. Eng. Phys.*, vol. 20, no. 5, pp. 332–338, Jul. 1998.
[28] B. Albouy, Y. Lucas, and S. Treuillet, "3D modeling from uncalibrated color images for a complete wound assessment tool," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 3323–3326.
[29] A. Yee, J. Harmon, and S. Yi, "Quantitative monitoring wound healing status through three-dimensional imaging on mobile platforms," *J. Amer. College Clin. Wound Specialists*, vol. 8, nos. 1–3, pp. 21–27, 2016.
[30] B. Gunsel, A. M. Ferman, and A. M. Tekalp, "Temporal video segmentation using unsupervised clustering and semantic object tracking," *J. Electron. Imag.*, vol. 7, no. 3, pp. 592–605, 1998.
[31] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. Vis.*, vol. 1, nos. 1–2, pp. 261–268, 1992.
[32] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annu. Rev. Biomed. Eng.*, vol. 2, no. 1, pp. 315–337, Aug. 2000.
[33] S. Wang and R. M. Summers, "Machine learning and radiology," *Med. Image Anal.*, vol. 16, no. 5, pp. 933–951, Jul. 2012.
[34] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift Für Medizinische Physik*, vol. 29, no. 2, pp. 86–101, May 2019.
[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[36] S. Zahia, D. Sierra-Sosa, B. Garcia-Zapirain, and A. Elmaghraby, "Tissue classification and segmentation of pressure injuries using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 159, pp. 51–58, Jun. 2018.
[37] R. H. L. E. Silva and A. M. C. Machado, "Automatic measurement of pressure ulcers using support vector machines and GrabCut," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105867.
[38] S. Sakakibara, A. Takekawa, C. Takekawa, S. Nagai, and H. Terashi, "Construction and validation of an image discrimination algorithm to discriminate necrosis from wounds in pressure ulcers," *J. Clin. Med.*, vol. 12, no. 6, p. 2194, Mar. 2023.
[39] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, pp. 1–21, Dec. 2015.
[40] S. Thomas, "Medetec wound database: Stock pictures of wounds," Tech. Rep., Feb. 2014.
[41] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, Dec. 2016.
[42] J. Zhou, P. Ke, X. Qiu, M. Huang, and J. Zhang, "ChatGPT: Potential, prospects, and limitations," *Frontiers Inf. Technol. Electron. Eng.*, vol. 2023, pp. 1–6, Feb. 2023.
[43] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, J. Wrobel, and H. Lee, "A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2415–2418.
[44] F. Li, C. Wang, X. Liu, Y. Peng, and S. Jin, "A composite model of wound segmentation based on traditional methods and deep neural networks," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–12, May 2018.
[45] X. Zhao, Z. Liu, E. Agu, A. Wagh, S. Jain, C. Lindsay, B. Tulu, D. Strong, and J. Kan, "Fine-grained diabetic wound depth and granulation tissue amount assessment using bilinear convolutional neural network," *IEEE Access*, vol. 7, pp. 179151–179162, 2019.

[46] D. Y. T. Chino, L. C. Scabora, M. T. Cazzolato, A. E. S. Jorge, C. Traina-Jr., and A. J. M. Traina, "Segmenting skin ulcers and measuring the wound area using deep convolutional networks," *Comput. Methods Programs Biomed.*, vol. 191, Jul. 2020, Art. no. 105376.

[47] B. Pandey, D. Joshi, A. S. Arora, N. Upadhyay, and H. S. Chhabra, "A deep learning approach for automated detection and segmentation of pressure ulcers using infrared-based thermal imaging," *IEEE Sensors J.*, vol. 22, no. 15, pp. 14762–14768, Aug. 2022.

[48] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 21–37.

[49] J. C. Gardiner, P. L. Reed, J. D. Bonner, D. K. Haggerty, and D. G. Hale, "Incidence of hospital-acquired pressure ulcers—A population-based cohort study," *Int. Wound J.*, vol. 13, no. 5, pp. 809–820, Oct. 2016.

[50] S. Seo, J. Kang, I. H. Eom, H. Song, J. H. Park, Y. Lee, and H. Lee, "Visual classification of pressure injury stages for nurses: A deep learning model applying modern convolutional neural networks," *J. Adv. Nursing*, Feb. 2023.

[51] B. Aldughayfiq, F. Ashfaq, N. Jhanjhi, and M. Humayun, "YOLO-based deep learning model for pressure ulcer detection and classification," *Healthcare*, vol. 11, no. 9, p. 1222, 2023.

[52] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[53] A. Khalil, M. Elmogy, M. Ghazal, C. Burns, and A. El-Baz, "Chronic wound healing assessment system based on different features modalities and non-negative matrix factorization (NMF) feature reduction," *IEEE Access*, vol. 7, pp. 80110–80121, 2019.

[54] Y. Liu, Z. Ma, X. Liu, S. Ma, and K. Ren, "Privacy-preserving object detection for medical images with faster R-CNN," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 69–84, 2022.

[55] R. Sa, W. Owens, R. Wiegand, M. Studin, D. Capoferri, K. Barooha, A. Greaux, R. Rattray, A. Hutton, J. Cintineo, and V. Chaudhary, "Intervertebral disc detection in X-ray images using faster R-CNN," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 564–567.

[56] T. Chen, W. Zheng, H. Ying, X. Tan, K. Li, X. Li, D. Z. Chen, and J. Wu, "A task decomposing and cell comparing method for cervical lesion cell detection," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2432–2442, Sep. 2022.

[57] C. W. Chang, M. Christian, D. H. Chang, F. Lai, T. J. Liu, Y. S. Chen, and W. J. Chen, "Deep learning approach based on superpixel segmentation assisted labeling for automatic pressure ulcer diagnosis," *PLoS ONE*, vol. 17, no. 2, Feb. 2022, Art. no. e0264139.

[58] R. Zhang, D. Tian, D. Xu, W. Qian, and Y. Yao, "A survey of wound image analysis using deep learning: Classification, detection, and segmentation," *IEEE Access*, vol. 10, pp. 79502–79515, 2022.

[59] D. Ramachandram, J. L. Ramirez-GarciaLuna, R. D. J. Fraser, M. A. Martínez-Jiménez, J. E. Arriaga-Caballero, and J. Allport, "Fully automated wound tissue segmentation using deep learning on mobile devices: Cohort study," *JMIR mHealth uHealth*, vol. 10, no. 4, Apr. 2022, Art. no. e36977.

[60] M. Swerdlow, O. Guler, R. Yaakov, and D. G. Armstrong, "Simultaneous segmentation and classification of pressure injury image data using Mask-R-CNN," *Comput. Math. Methods Med.*, vol. 2023, pp. 1–7, Feb. 2023.

[61] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[62] A. Kairys, R. Pauliukiene, V. Raudonis, and J. Ceponis, "Towards home-based diabetic foot ulcer monitoring: A systematic review," *Sensors*, vol. 23, no. 7, p. 3618, Mar. 2023.

[63] O. Y. Dweekat, S. S. Lam, and L. McGrath, "Machine learning techniques, applications, and potential future opportunities in pressure injuries (Bed-sores) management: A systematic review," *Int. J. Environ. Res. Public Health*, vol. 20, no. 1, p. 796, Jan. 2023.

[64] S. Chairat, S. Chaichulee, T. Dissaneewate, P. Wangkulangkul, and L. Kongpanichakul, "AI-assisted assessment of wound tissue with automatic color and measurement calibration on images taken with a smartphone," *Healthcare*, vol. 11, no. 2, p. 273, 2023.

[65] B. Ay, B. Tasar, Z. Utlu, K. Ay, and G. Aydin, "Deep transfer learning-based visual classification of pressure injuries stages," *Neural Comput. Appl.*, vol. 34, no. 18, pp. 16157–16168, Sep. 2022.

[66] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*. Zurich, Switzerland: Springer, Sep. 2014, pp. 740–755.

[67] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[68] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[69] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*.

[70] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[71] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. ICML*, 2010, pp. 1–11.

[72] C. Olston, N. Fiedel, K. Gorovoy, J. Harmsen, L. Lao, F. Li, V. Rajashekhar, S. Ramesh, and J. Soyke, "TensorFlow-serving: Flexible, high-performance ML serving," 2017, *arXiv:1712.06139*.

[73] R. T. Fielding and R. N. Taylor, "Principled design of the modern web architecture," *ACM Trans. Internet Technol.*, vol. 2, no. 2, pp. 115–150, May 2002.

PAUL FERGUS is currently a Professor in machine learning with Liverpool John Moores University. He has spent just under 30 years studying artificial intelligence (AI) from early days of symbolic AI using Prolog and Lisp to his current research practices in deep learning. His research interests include machine learning for detecting and predicting preterm births and the detection of foetal hypoxia, electroencephalogram seizure classification and bioinformatics (polygenetic obesity, type II diabetes, and multiple sclerosis), and conservation. He is also looking at the use of machine learning to solve different conservation-related problems. He has competitively won external grants to support his research from EPSRC, HEFCE, Royal Academy of Engineering, Innovate U.K., Knowledge Transfer Partnership, North West Regional Innovation Fund, and Bupa. He has published over 200 peer-reviewed papers in these areas and coauthored a book titled *Applied Artificial Intelligence: Mastering the Fundamentals*. Before his academic career, he was a senior software engineer in industry for six years developing bespoke solutions for a number of large organizations.

CARL CHALMERS is currently a Senior Lecturer with the Department of Computer Science, Liverpool John Moores University. He is also leading a three-year project on smart energy data and dementia in collaboration with Mersey Care NHS Trust. As a part of the project a six month patient trial is underway with NHS with future trials planned. The current trail involves monitoring and modeling the behavior of dementia patients to facilitate safe independent living. In addition, he is also working in the area of high performance computing and cloud computing to support and improve existing machine learning approaches, while facilitating application integration. His research interests include advanced metering infrastructure, smart technologies, ambient assistive living, machine learning, high performance computing, cloud computing, and data visualization. His current research interests include remote patient monitoring and ICT-based healthcare.

WILLIAM HENDERSON is currently a Clinical Research and Innovation Practitioner with Mersey Care NHS Foundation Trust. He supports both local and National Institute for Health Research (NIHR) studies. He is committed to delivering and supporting access to high quality research to a diverse range of patients, careers, and staff. He strongly believes that research should be accessible to everyone in the NHS regardless of their mental or physical health diagnosis, helping to ensure evidence-based practice is at the forefront of everything Mersey Care NHS Foundation Trust does.

DANNY ROBERTS is currently the Quality Improvement Lead with the Centre for Perfect Care, Mersey Care NHS Foundation Trust. His research interests include improving patient safety and quality with primary car. He was a runner up in the Tissue Viability Nurse of the year category in *The British Journal of Nursing* awards, in 2021.

ATIF WARAICH is currently the Director of the School of Computer Science and Mathematics, Liverpool John Moores University (LJMU). Prior to joining LJMU, in June 2017, he was the Head of the Division of Digital Media and Entertainment Technology, Manchester Metropolitan University (MMU), and also the Enterprise Lead, the Founder, and the Director of the Manchester Usability Laboratory. His research interests include the use of technology to enhance learning, specifically he is interested in how game like environments can be used to promote learning and to motivate learners to engage in their studies, the security and application of 5G wireless technologies, and the application of usability and behavior modification techniques (including gamification) to augmented and virtual reality environments. He is also the Founder and the Director of the Liverpool Immersive Experience (LIVE) Laboratory. He has been a Principal Investigator on an Innovate U.K. funded Project, ''Rail Incident Manager. From 2011 to 2013, he was a PI on LIFE+ Environmental Policy and Governance as part of a larger EU project with Greater Manchester Waste Disposal Authority (GMWDA). He was the lead academician with LJMU on the €5M Liverpool 5G testbed project funded by Innovate U.K. He is also the lead academician on the €4M Liverpool 5G Create Programme. He is currently a Co-Investigator on the U.K.–Malaysia University Consortium (U.K.–MUC) and Digital Catalyst for Graduate Employability projects with international partners in Malaysia and Indonesia. He has acted as a reviewer for numerous conferences and journals and a reviewer for the HEA in funding bids for technology enhanced learning. He is a member of the British Computer Society.