IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 13, 2020, accepted October 14, 2020. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.3035327

Machine Learning in the Prevention, Diagnosis and Management of Diabetic Foot Ulcers: A Systematic Review

JACK TULLOCH[1], REZA ZAMANI[1], AND MOHAMMAD AKRAMI[2]

[1]Medical School, College of Medicine and Health, Exeter EX1 2LU, U.K.
[2]Department of Engineering, College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter EX4 4QF, U.K.

Corresponding author: Mohammad Akrami (m.akrami@exeter.ac.uk)

ABSTRACT Diabetic foot ulcers (DFUs) are a serious complication for people with diabetes. They result in increased morbidity and pressures on health system resources. Developments in machine learning (ML) offer an opportunity for improved care of individuals at risk of DFUs, to identify and synthesise evidence about the current uses and accuracy of ML in the interventional care and management of DFUs, and, to provide a reference for areas of future research. PubMed, Google Scholar, Web of Science and Scopus were searched using the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines for papers involving ML and DFUs. In order to be included, studies needed to mention ML, DFUs, and report relevant outcome measures regarding ML algorithm accuracy. Bias in included studies was assessed using the quality assessment tool for diagnostic accuracy (QUADAS-2). 37 out of 3769 papers were included after applying eligibility criteria. Included papers reported accuracy measures for multiple types of ML algorithms in DFU studies. Whilst varying across the ML algorithm used, all studies reported at least 90% accuracy compared to gold standards using a minimum of one reported ML algorithm for processing or recording data. Applications where ML had positive effects on DFU data analysis and outcomes include image segmentation and classification, raw data analysis and risk assessment. ML offers an effective and accurate solution to guide analysis and procurement of data from interventions which are designed for the care of DFUs in small samples and study conditions. Current research is limited, and, for the development of more applicable ML algorithms, future research should address the following: direct comparison of ML applications with current standards of care, health economic analyses and large scale data collection. There is currently no evidence to confidently suggest that ML methods in DFU diagnosis are ready for implementation and use in healthcare settings.

INDEX TERMS Diabetes, diabetic foot, machine learning, review, ulcers.

LIST OF ABBREVIATIONS

Acronym	Term
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under Curve
CNN	Convolutional Neural Network
CPU	Central Processing unit
D-ANN	Deep Artificial Neural Network
D-CNN	Deep Convolutional Neural Network
DFU	Diabetic Foot Ulcer
DPN	Diabetic Peripheral Neuropathy
DSC	Dice Similarity Coefficient
DT	Decision Tree
EMC	Expectation Maximisation Clustering
FCN	Fully Convolutional Networks
FFBP	Feed Forward Back Propagation
FL	Fuzzy Logic
FSC	Fuzzy Spectral Clustering
GA	Genetic Algorithm
GPU	Graphics Processing Unit
IoU	Intersection-over-Union
JSI	Jaccard Similarity Index
KMC	K-Means Clustering
k-NN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis

The associate editor coordinating the review of this manuscript and approving it for publication was Liangxiu Han.

LRC	Logistic Regression Classifier
MCC	Matthews correlation coefficient
ML	Machine Learning
MLR	Multivariate Linear Regression
MRM	Multivariate Regression Modelling
MSA	Mean-shift algorithm
NBC	Naïve Bayes Classifier
NPV	Negative Predictive Value
PCA	Principle Component Analysis
PNN	Probabilistic Neural Network
PPV	Positive Predictive Value
PSO	Particle Swarm Optimisation
R-ANN	Region Based Artificial Neural Network
R-CNN	Region Based Convolutional Neural Network
RF	Random Forest
ROI	Region of Interest
SOM	Self-Organising Map
SVD	Singular Value Decomposition
SVM	Support Vector Machine
T2DM	Type 2 Diabetes Mellitus

I. INTRODUCTION

Diabetes Mellitus (DM) is one of the leading worldwide causes of death and quality of life impairment. In 2019, approximately 463 million adults were living with diabetes, with this expected to rise to 700 million adults by the year 2045 [1]. Diabetic peripheral neuropathy (DPN) is a severe complication affecting 30-50% of people with diabetes which affects the sensory nerve supply to the feet which can cause infections, structural foot changes, and the development of diabetic foot ulcers (DFUs) [2]. Foot ulceration acts as a precursor to the development of gangrene and limb loss, and lower limb amputation is carried out more than 20 times as often in people with diabetes than in those without [3]. The World Health Organisation (WHO) estimates that a lower limb is lost every 30 seconds somewhere in the world because of diabetes [4]. DFUs have a 5-year mortality rate of more than 50%, rising to 80% in patients who have a diabetes-related amputation, making mortality rates worse than many common cancers [4]. The estimated cost to the UK's NHS is around £580 million a year, with £307 million spent on ulceration in the primary care setting [5]. This highlights the economic burden of the disease and the need for improvements to the current care paradigm.

The development of DFUs is the result of multiple contributing factors. The major underlying causes are peripheral neuropathy and ischaemia from peripheral vascular disease decreasing the protective factors of the tissues, in conjunction with some form of stress (e.g. pressure, shear and trauma) to the skin. Complications of foot ulcers include severe pain, infection, gangrene, osteomyelitis, amputation, and death [6]. Coexisting diabetic complications such as reduced peripheral sensation and lack of pain allow continued ambulation, facilitating further damage. Osteomyelitis may occur in advanced cases, carrying a high mortality rate [7].

Current management focusses on patient education, regular self-foot checks and annual diabetic foot assessments. These annual checks involve a patient history, peripheral vascular exam, and assessment of sensory nerve function for early identification of DPN. Studies looking at possibilities of reducing foot pressure or changing gait using pressure analysis provide an interesting insight into new technological advances for early detection and prevention of DFU [8], [9]. In evaluating risk, infrared thermography is another technological advance that can provide clinical information to assist in the early diagnosis and prevention of lesions in compromised zones of the foot [10]. Research indicates an isolated difference of just 2.2 °C compared to mean foot temperature is significant in identifying inflammatory processes [10]. Management of confirmed DFU patients involves offloading, control of infection or ischaemia, wound debridement and wound dressings if necessary [11]. The results from the tests used to guide management may take time to become available in clinical settings and progression of DFU severity is hard to distinguish visually, with many DFU classification systems taking into account the proportion and types of tissue which are visually distinguishable [12]. Granulation tissue is red/pink and is tissue that is healing, slough tissue is more yellow and represents infective tissue, whilst dark/black necrotic tissue indicates an area of tissue death. These tissues are the common basis for most visual tests carried out in healthcare with relation to DFU. As with all visual classification systems, there is an element of subjectivity in assessment, as well as the risk of confirmation bias. Numerous studies show that proper diagnosis and management of DFUs can greatly reduce or prevent serious complications [12], [13]. Despite various national and international guidelines, the management of DFUs remains inconsistent. Due to the importance of reliable and quick management in diabetic patients, machine learning (ML) has great potential to improve healthcare systems.

Constant advances in technology means the use of ML algorithms in healthcare is becoming an increasingly popular approach. Their ability to reduce human error, cost, number of personnel and time taken to complete tasks are valued features. ML algorithms are characterised by their ability to learn and adapt over time without being explicitly programmed. ML can be classified into supervised learning, which trains a model on known input and output data so that it can predict future outputs and unsupervised learning, which finds hidden patterns or intrinsic structures in input data [1]. Supervised learning uses classification and regression techniques to develop predictive models. Unsupervised learning in healthcare studies commonly utilises clustering, which allows the algorithms to find hidden patterns or groupings in data. These tasks can be applied to DFU care in predicting complication probability, screening, diagnosis, and guided management. However, the large systems require high amounts of raw data and high operational costs. Details of the different types of ML referenced are found in the list of ML definitions at the beginning of this review. ML provides an

opportunity for improved classification, extraction and analysis of data in DFU studies [14]. Studies utilising ML in diabetes have already shown promise in accuracy and reliability. Arcadu *et al.* [15] used deep learning (DL) to predict future progression in individuals with diabetic retinopathy, reaching a maximum sensitivity of 79±12% at 12 months and a maximum specificity of 77±12% at 6 months. Outside the field of diabetes, Wu *et al.* [16] utilised a novel deep convolution neural network (D-CNN) trained on a large data set to detect early gastric cancer during oesophagogastroduodenoscopy. The network outperformed all human comparators in the study, reaching an accuracy of 92.5%, a sensitivity of 94.0% and a specificity of 91.0% in identifying early gastric cancer lesions. ML therefore shows promise at improving diagnostic accuracy of tests in the healthcare field. Benefits of ML for use in DFU care are varied, and include improved clinical decision making based on ulcer classification and healing status, data analysis for risk and automated classification systems. ML could be applied to applications for mobile devices or to allow remote access. This literature review will identify current applications of ML in DFUs to identify future research needs. This will enable better use of resources by directing future studies towards areas of need. Studies on implementing ML into diabetic foot care focus mainly on imaging and segmentation or classification of DFU to improve diagnosis or management [17], [18]. They acknowledge the importance of prevention and diagnosis of DFUs as a means of improving overall life expectancy, quality of life and also in reducing healthcare-associated costs. Despite its many applications, there is a lack of coherence in the use of ML in DFUs and its potential to improve care. Currently there does not exist a single published systematic review comparing the uses and accuracy of ML in people with DFUs. The rationale for this study is to summarise and convey studied methods for improving DFU care in healthcare settings using ML. The advantages of ML over current diagnostic methods include the reduced impact of subjective bias and of human errors such as fatigue or negligence. Improvements in the time taken to complete tasks and accessibility are also considerations.

This systematic review aims to be the first to understand and compare the current applications of ML in the care of DFU patients and compare their diagnostic and prognostic accuracy. The review will discuss and compare papers using real-world participant data of people with DM, which use ML in the acquisition or interpretation of data from index tests. It will identify possible areas of future development for ML applications in DFU care and suggest improvements in research methodologies to allow the results of future studies to improve clinical treatment guidelines and aid the implementation of ML technologies into healthcare services.

II. METHODOLOGY

As a template for the methodology of this review, the Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) guidelines were used [19].

A. ELIGIBILITY CRITERIA

Titles and abstracts of papers in the search were initially screened using exclusion criteria. Studies which remained after this were assessed using their full-texts, using the eligibility criteria shown in detail in Table 1. Criteria was designed after review of preliminary papers by a focus group of the three authors stated above, and were deigned to be broad to avoid missing relevant literature. Characteristics included papers published in English, which used ML for the analysis or acquisition of data from patients with diagnosed type 1 (T1DM) or Type 2 diabetes mellitus (T2DM) with DFUs. Papers needed to report an outcome measure of importance relating to the accuracy of the ML method used in order to be included in this review. This was important to allow the comparison of uses between papers, and address the main research question regarding ML and DFU care.

B. LITERATURE SEARCH PROCESS

For this study, four electronic databases were analysed, which were as follows:

1. MEDLINE electronic database via PubMed
2. Google Scholar using Harzing's Publish or Perish software (Version 7.18)
3. Web of Science
4. Scopus

The search strategies were: computer search of databases, review of reference lists of included articles, and manual addition of relevant articles or exclusion of irrelevant articles. The search was performed between the 9th and 17th of January 2020, reporting papers focussed on the use of ML on diabetes-related ulcers. The categories of search terms used were:

1. Terms for diabetes
2. Terms for ulcers
3. Terms for ML
4. Terms for analysis of outcomes

The databases were searched employing Boolean logic using the precise keywords detailed in Table 1. Limits were applied to reduce the quantity of results returned whilst retaining relevant publications so as only to include: human studies, papers published in English, and papers with full-texts available. The use of limits was considered to cut down the number of irrelevant articles and limit keyword searches to the title, abstracts and introductions. This was agreed upon after discussion between the three authors to ensure it would not discriminately exclude relevant articles. Whilst this may have introduced some bias, the authors concluded it was unlikely that a relevant paper would not mention the above search terms in either its title, abstract or introduction.

C. STUDY SELECTION

In this review, only studies published in the English language were considered. The literature acquired by the above-mentioned search protocol were imported into Microsoft

TABLE 1. Eligibility criteria.

Inclusion Criteria	Exclusion Criteria
Published in English	Results that are conference abstracts where there is no available full-text
Studies reporting the use of machine learning on patient datasets with diabetic foot ulcers	Results that are singular case reports or extracts from books and published guidelines
	Papers with no abstract or full-text available
	Studies that did not report the use of machine learning on patients or datasets with diabetic foot ulcers
Studies that include a statistical analysis of results of machine learning	Studies that did not include a statistical analysis of results of machine learning
Studies that have full-text available	Studies that include participants or datasets who have not been officially diagnosed with type I or type II Diabetes Mellitus
Patients or datasets from a human diagnosed with type I or type II Diabetes Mellitus	Studies that only include pathological foot ulcers other than those associated with diabetes in origin
Reports an outcome measure of importance: a. Sensitivity of machine learning b. Specificity of machine learning c. Positive predictive value (PPV) of machine learning d. Negative predictive value (NPV) of machine learning e. Accuracy of machine learning procedures f. Reliability of machine learning procedures	Studies that did not report at least one outcome measure of interest for the purpose of this review

Inclusion and exclusion criteria for assessing studies

Excel to include the following data: title, authors, date of publication, place of publication, and full abstract. Using the software, duplicates were removed from the list of literature and remaining article abstracts were screened using eligibility criteria (table I). Forms of publications other than journal articles were excluded from the list of literature with proper inspection. Following this process, remaining articles were assessed in their full-text to exclude any irrelevant articles. Five additional publications were added manually, which were found in the references of previously selected systematic reviews and considered relevant for this study. Before screening started, the three authors discussed the use of eligibility criteria for screening to ensure agreement between assessors. Screening of titles and abstracts was done by the first author (JT), and a random sample of 100 papers was assessed by the second (MA) and third authors (RZ) for conformity to the eligibility criteria and research question to reduce the risk of selection bias. The final screening and selection of included papers was reviewed by the second (MA) and third authors (RZ). Overall the exclusion and inclusion of papers for the purposes of the review was agreed satisfactorily by all authors involved. The study selection procedure is presented in Fig. 1.

D. DATA EXTRACTION

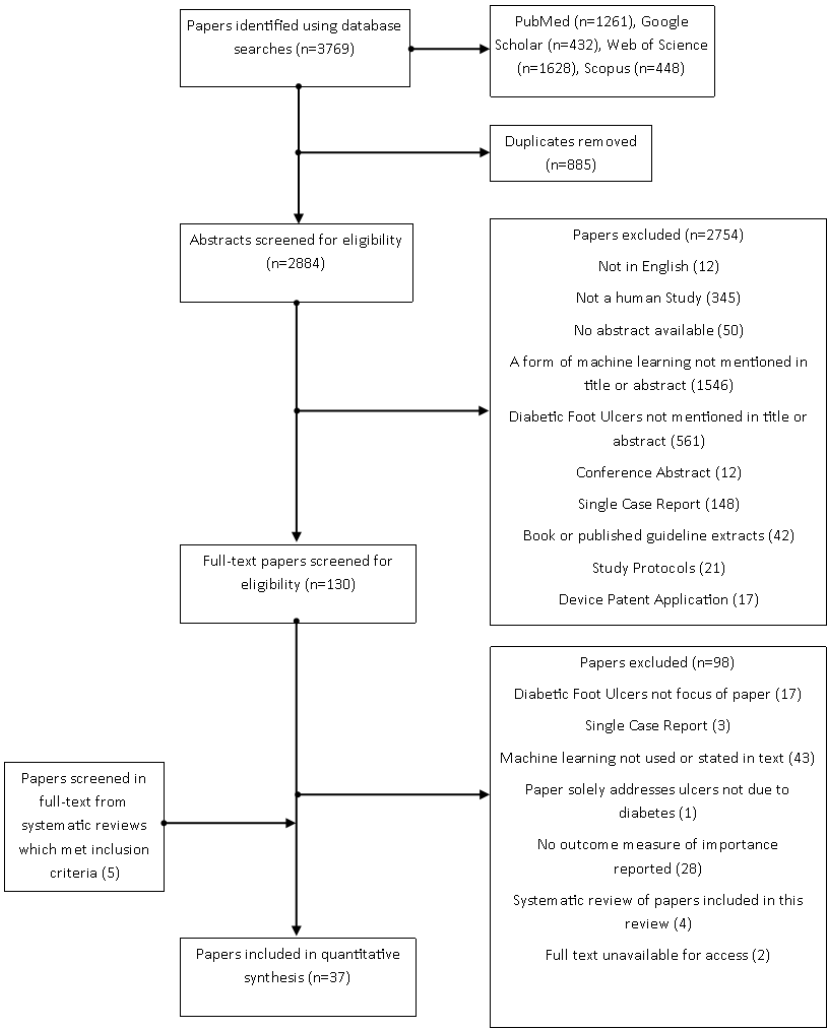
Data was extracted from the studies using a standardised form, and included: title, authors, date of publication, place of publication, dataset, sample size, ML methods used, applications of ML, and all relevant outcome measures reported.

E. ASSESSMENT OF METHODOLOGICAL QUALITY

Methodological quality of the studies was assessed using the quality assessment tool for diagnostic accuracy (QUADAS-2) checklist [20] and journal rating (SCImago) [21]. The QUADAS-2 tool was designed to evaluate the risk of bias and applicability of diagnostic accuracy studies and consists of four key domains:

1. Patient selection
2. Index test
3. Reference standard
4. Flow and timing

All four domains are assessed with regards to risk of bias and the first three in terms of concerns regarding applicability. The risks and concerns are rated as high, low and unclear. Unclear risk was determined when there was insufficient presented data in the study to draw a conclusion from. Not applicable (N/A) was used where the QUADAS domain does not apply due to the study methodology. High risks of bias in any category may indicate issues with the papers methodology, and across multiple categories may affect the reliability of the reported results in relation to the research question. High risk of bias regarding applicability in the tested domains may indicate the included data from the assessed paper does not accurately match the review question.



The method of literature screening and selection, including the number of results from the combined databases and the number and reason for exclusion at each stage of the process.

FIGURE 1. PRISMA-DTA flow diagram for literature selection.

F. DIAGNOSTIC ACCURACY MEASURES

The principal diagnostic accuracy measures reported in this review vary greatly across the included literature. Whilst all outcome measures report the accuracy of the ML algorithm used, this applies to the whole dataset of the individual paper, and no paper reports outcome measures per-patient or per-lesion. As such the reported measures seen in the results sections of this review must be examined with the knowledge that it applies to a whole dataset accuracy, and not an individual basis. The stated accuracy measures within this review include; accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the ROC curve (AUC), mean intersection-over-union (IoC) (AKA Jaccard index), Dice coefficient (AKA F-1 Score), kappa statistic and the Matthews correlation coefficient (MCC).

Accuracy is an umbrella term in measures, and can be reported in multiple ways. The reported accuracy demonstrates the ability of the ML algorithm to perform its task and determine the correct target condition/classification compared to its gold-standard. Sensitivity refers to the proportion of positives that are correctly identified whilst specificity measures the proportion of negatives that are correctly identified and are useful measures of accuracy. PPV and NPV assesses the likelihood that a patient has a specific disease and is more related to clinical scenarios. The AUC quantifies the overall ability of the test to discriminate between those individuals with the disease and those without the disease and is given a value between 0.5 and 1.0 where 0.5 means the test is no better at identifying true positives than chance. The Intersection-Over-Union (IoU), also known as the Jaccard Index, is one of the most commonly used metrics in semantic segmentation. IoU is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. The dice coefficient is linked to IoU and is calculated as 2 multiplied by the area of overlap divided by the total number of pixels in both images. Kappa statistic is used

as a quantitative measure of the magnitude of agreement between observers and can be measured in any situation in which two or more independent observers are evaluating the same thing, where a score of 1.0 indicates perfect agreement and <0 indicating a less than chance agreement. The MCC is used in machine learning as a measure of the quality of binary classifications using observed and predicted binary classifications. A coefficient of 1 represents a perfect prediction, 0 no better than random prediction and -1 indicates total disagreement between prediction and observation. As well as accuracy measures, other measures included are processing time and size of the model. These assess functions of ML algorithms other than the ability to give an accurate output. Processing time indicates the length of time for the algorithm to perform a given function as specified by the paper (i.e. time per-lesion or per-dataset), and can refer to any part of the ML process. Size of the model relates to the storage required for the ML algorithm and its stored data.

As the search criteria included umbrella-terms for accuracy, no papers were excluded for containing any form of accuracy measurement not mentioned specifically in the search parameters. Accuracy measures not reported in this review indicate a lack of relevant articles reporting these statistical measures as outcome measurements.

G. DATA SYNTHESIS AND ANALYSIS

Extracted data were organised into groups based on the type of ML method (supervised or unsupervised) and function (image analysis, data analysis, segmentation, or classification). Data were also organised based on the type of index test used in the study. This allowed direct comparison of data between studies. All outcome measures were extracted and analysed in a standard form, including all definitions of accuracy regardless of measure used by the included papers to record this. Data is presented in the results section by function, and all ML algorithms are recorded with their respective outcome measure values for comparison and understanding of ML efficacy across similar task functions.

III. RESULTS

A total of 3769 papers were identified. Titles were screened to eliminate duplications, leaving 2885, before evaluating abstracts and applying exclusion criteria. The remaining articles (n = 130) were reviewed in their full-text forms. These were refined to 37 studies based on the eligibility criteria (Table 1). The flowchart based upon the PRISMA-DTA methodology detailing this process is outlined below (See Figure 1). The 37 papers included in this review reported multiple forms of ML in their published outcomes.

A. STUDY CHARACTERISTICS

The 37 studies included in this paper have varied characteristics and demographics. The full details of study characteristics are found in Table 2. The publishing year for the included papers ranged from 2010 to 2020. Datasets varied between papers. One study used 255 patients' genetic data [22] and three studies used bacterial samples taken from DFUs [23]–[25]. Of the remaining 33 studies, four studies used thermographic images [26]–[29], and 22 studies included colour images [3], [17], [18], [30]–[48], three studies included both colour and thermographic images [49]–[51], two papers recorded dynamic foot pressure [32], [52], one study used hyperspectral imaging (HIS) [22] and one study examined 301 patient records for variables affecting amputation decisions [53]. Fifteen of the 37 studies focussed on using ML techniques for classification [3], [23]–[27], [29]–[32], [36]–[38], [53], [54]. Eight studies focussed on the segmentation of colour images [3], [17], [40], [42], [43], [45], [46], [48], and four studies focussed on the segmentation of thermal images [28], [49]–[51]. Nine studies utilised ML to conduct both segmentation and classification [31], [33]–[35], [39], [41], [44], [52], [55]. One study looked at risk analysis via regression [22].

ML algorithms varied across studies. The majority of the studies included a form of artificial neural network (ANN), k-Means, convolutional neural network (CNN) or support vector machine (SVM). SVM was the highest referenced form of ML, with 32% of studies including it in their analysis and results (n = 12). CNNs were present in 30% of studies (n = 11). K-Means and ANNs were both referenced in 19% of studies (n = 7) and K-Nearest Neighbour (k-NN) was mentioned in 14% of included studies (n = 5). The other ML algorithms were mentioned in under three studies. A graphical representation of the spread of ML techniques can be found below (See Figure 2).

All papers used at least one of the primary outcome measures identified by the eligibility criteria. Results are split into studies whose primary results considered outcomes regarding data classification (Table 3), colour image segmentation (Table 4), thermal image segmentation (Table 5), both segmentation and classification (Table 6) and other ML applications (Table 7).

B. METHODOLOGICAL QUALITY

The majority of studies were found to be low risk overall with regards to bias and applicability defined by the QUADAS-2 tool, indicating a good methodological quality, as shown in Table 8, and graphically in Figure 3. There was a high risk of bias identified in four studies [29], [34], [40], [43]. Adam *et al.* [29] was high risk for applicability due to flaws in patient selection. These included unjustified exclusions and non-random sample selection, resulting in test data that did not accurately fit the review question. The use of healthy subjects in the diabetic screening tests in the study by Siddiqui *et al.* [43] limits the ability to apply study data to the diabetic population. The risk of bias due to patient flow was high in two of the studies. In Godeiro *et al.* [34] out of a small data set of 30 images, only 15 were used to test the final classification system. In Dhane *et al.* [40] only images where the gold standard reference test showed high agreement were selected for testing. This is likely to have

TABLE 2. Characteristics of included studies.

Study	SCImago journal rating	Machine learning methods	Application	Dataset	Outcome measures	Reference test
[23]	Q4	PCA LDA ANN utilising FFBP	Data analysis (PCA & LDA) and classification (ANN) to recognise types of bacteria in the diabetic foot ulcer of patients	6 combined cultures Containing 3 different bacterial strains found on DFUs (wild and standard type)	Classification accuracy	N/A
[29]	Q2	SVM	Classification of thermogram features (SVM)	Thermographic images of feet (controlled conditions) Singapore 33 healthy subjects and 33 non-neuropathic diabetic patients	Classification accuracy	N/A
[30]	Q1	D-CNN CNN SVM KNN	Classification - Normal (healthy) skin versus abnormal (ulcer) skin	754 foot images Iraq healthy skin and skin with a diabetic ulcer from different patients	Precision Recall F1-score	Manually labelled by a DFU expert
[31]	Q3	PSO NBC Hoeffding tree classifier	Segmentation of diabetic wounds (PSO) Classification (Naïve Bayes & Hoeffding tree) - granulation tissue versus necrotic tissue versus slough tissue	Diabetic wound images collected from open source database Three DFU images used to test method	Accuracy Sensitivity Specificity	N/A
[32]	Q1	SVM	Feature selection and classification (SVM) - Healthy Controls (HC) - Diabetic Controls (DC) - Diabetics with Neuropathy (DN) Prediction for diabetic foot ulcers	84 subjects Egypt 56 diabetic patients with and without diabetic peripheral neuropathy 28 control non-diabetic subjects Recorded dynamic plantar pressure measurements during normal gait	Accuracy Precision	Diagnosis of DPN was performed using a 5.07/10g Semmes-Weinstein monofilament
[17]	Q1	CNN SVM	Diabetic wound segmentation	445 high resolution images 392 used for training classifiers Unseen set of 53 images used for testing	Sensitivity Specificity Dice index Mean IoC	N/A
[33]	Q1	CNN ANN RF SVM LDA	Feature extraction (GAP and SVD) Classification of severity stage of DFUs (ANN, Random forest and SVM) - 6 stages of Wagner Ulcer Grading scale	2400 images from a diabetic clinic containing all 6 Wagner stages 80% used as training dataset 20% used as testing dataset	Accuracy F-1 Score	N/A
[34]	Q1	CNN Watershed method Otsu method Gaussian algorithm	Image segmentation (Watershed algorithm, Otsu and GrabCut algorithm) Classification of DFU tissue (CNN) - granulation tissue versus necrotic tissue versus slough tissue	30 images of diabetic chronic wounds (predominantly hands and feet) Uncontrolled environment Brazilian patients 50% used as training dataset 50% used as testing dataset	Accuracy Specificity Sensitivity Dice coefficient	N/A
[35]	Q1	FCN CNN	DFU segmentation and classification (FCN) - Normal (healthy) skin versus abnormal (ulcer) skin	705 foot images 600 DFU images 105 healthy foot images Britain Partially controlled conditions 70% use as training dataset 10% used as validation dataset 20% used as testing dataset	Dice coefficient Specificity Sensitivity	Expert annotation of DFU and surrounding skin on foot images Performed by a specialised podiatrist and validated by a diabetic consultant
[36]	Q1	CNN	Feature extraction Classification - Normal (healthy) skin versus abnormal (ulcer) skin	397 foot images 292 DFU images 105 healthy foot images Britain Partially controlled conditions 85% use as training dataset 5% used as validation dataset 10% used as testing dataset	Sensitivity Specificity Precision Accuracy F-Measure AUC	Expert annotation of DFU and surrounding skin on foot images

TABLE 2. ***(Continued.)*** **Characteristics of included studies.**

[18]	Q2	CNN (SVM) Bayesian algorithm (BayesNet) RF MLP	Feature extraction (BayesNet, random forest, and multilayer perceptron) Classification (CNN) - Ischaemia versus non-ischaemia - infection versus non-infection	1459 DFU images Britain Partially controlled conditions 70% use as training dataset 10% used as validation dataset 20% used as testing dataset	Accuracy Sensitivity Precision Specificity F-Measure AUC	Expert annotation of DFU and surrounding skin on foot images Performed by two diabetic consultants
[37]	Q1	R-CNN	Classification - Normal (healthy) skin versus abnormal (ulcer) skin	1880 foot images 1775 DFU images 105 healthy foot images Android application for mobile phone use Britain Controlled conditions 80% use as training dataset 20% used as testing dataset	Speed Size of the model mAP Overlap Percentage	Expert annotation of DFU and surrounding skin on foot images Performed by a specialised podiatrist and validated by a diabetic consultant
[38]	Q1	Gradient boosted tree models	Classification - Normal (healing) wounds versus abnormal (non-healing) wounds	Basic demographic information on 53,354 patients and Information on 150,277 wounds used. 11% of wounds deemed non-healing 40 wound types and 37 anatomic locations Venous leg ulcers, pressure ulcers, and diabetic wounds on the lower extremities accounted for 48.8% of wounds America 52.7% male 80% use as training dataset 20% used as testing dataset	AUC	N/A
[28]	Q1	GA	Image segmentation (genetic algorithm)	100 thermal images, corresponding to patients' feet with and without inflammation Uncontrolled background	Misclassification error (ME) Region nonuniformity (NU) Relative foreground area error (RAE) Edge mismatch (EMM) Combination	N/A
[53]	Q1	Boosted classification and regression tree (Boosted C5.0 tree ensemble)	Classification - Risk of amputation	Hospital records of 301 diabetic foot patients examined retrospectively for explanatory variables of foot amputation decisions India 83 patients underwent amputation and 218 were managed conservatively Wagner grading and Doppler flow measurement used for classification of amputation risk 250 patients used as training dataset 51 patients used as testing dataset	Accuracy of predicted outcome	Simple classification and regression tree (C5.0 tree)
[49]	Q1	k-Means clustering EM clustering	Segmentation - Foot surface versus background image	RGB and thermographic images of 76 diabetic feet Controlled conditions Netherlands	Sensitivity Specificity	Live assessment form of the plantar surface of both feet was completed by wound care specialists.
[39]	Q2	R-CNN	Segmentation Classification - Ulcer versus necrosis versus no risk	First database for testing had a total of 108 thermographic images from 17 volunteers with 4 different backgrounds Second database for retraining R-CNN model contained 141 images from 47 new volunteers Uncontrolled environment	Sensitivity Specificity Accuracy	N/A
[40]	Q2	Fuzzy spectral clustering	Ulcer boundary demarcation and estimation Segmentation	70 selected Ulcer images of 64 patients 44 were diabetic Partially controlled	Accuracy Sensitivity Specificity Jaccard index Dice coefficient	Wound boundaries traced by two dermatologists and a surgeon All selected images have maximum agreement between clinicians

TABLE 2. ***(Continued.)*** **Characteristics of included studies.**

[41]	Q2	Bayesian classification SVM Fuzzy divergence based thresholding	Image segmentation (Fuzzy divergence based thresholding) Wound tissue classification (Bayesian classification and SVM - granulation tissue versus necrotic tissue versus slough tissue	74 wound images (burn (n = 12), diabetic ulcer (n = 24), malignant ulcer (n = 14), Pyoderma gangrenosum (n=8), venous ulcer (n=7), and pressure ulcer (n= 9)) 222 regions as granulation tissue, 451 regions as slough tissue, and 94 regions as necrotic tissue	Tissue-wise accuracy (granulation, slough, necrotic) Overall accuracy Kappa statistic	Ground truth images by clinical experts Medical experts identified total tissue regions using annotator software.
[50]	Q2	K-means clustering watershed method	Image segmentation for clinician annotation	26 wound images (thermographic and colour image) Uncontrolled conditions	Dice coefficient	As segmentation ground truth, each image has been manually segmented by medical expert
[42]	Q1	CNN	Wound segmentation	Colour images of PUs (n=400), DFUs (n=20) and VLUs (n=20) from a digital database Controlled conditions Japan Training data for the CNNs included 356 PU images Test data included 40 PU images DFU and VLU images were used to verify accuracy of best performing CNN	AUC Sensitivity Specificity Dice coefficient MCC Accuracy Processing time (GPU/CPU)	N/A
[43]	Q1	Undefined method (Feature extraction using optical image processing (HSV) incorporating plantar anthropometry)	Feature (ROI) identification to automatically identify suiTable pressure points to apply a monofilament along the plantar surface	70 foot sole images of different ethnicities, age groups and gender The database consists of Chinese, Asian, African and Caucasian foot images Healthy subjects (44 male and 26 female participants	Accuracy Manual versus automated test time	Qualified podiatrist independently marked the five pressure points on each foot. Drew a "circle of acceptance" with a diameter of 1cm bounding those points
[22]	Q2	ANN utilising back-propagation	Data analysis and risk assessment	255 patients (India) Genotyped data for TLR4 receptor SNPs and haplotypes 125 T2DM patients with DFU 130 controls	Accuracy of prediction of DFU risk in test data	Multivariate regression modelling
[52]	N/A	Otsu method SVM	Classification and Segmentation of pressure data as a foot image	84 subjects Dynamic pressure distribution is recorded using med logic foot pressure measuring system for both feet	Accuracy Precision	N/A
[44]	Q3	ANN utilising FFBP PCA	Image segmentation (PCA) Skin macules classification (ANN) - vascular changes versus petechiae versus trophic changes versus trauma macules	19 Mexican patients diagnosed with T2DM, but not diabetic foot 60% of the data was used to train the ANN 40% of the data was used to test the ANN.	Identification of data Reliability of results matching target data	N/A
[27]	N/A	k-NN SVM ANN Weka framework	Automatic classification of thermography images at ROIs	Infrared thermal images 56 DFU early stage patients Controlled conditions	Accuracy Positive prediction	N/A
[26]	N/A	ANN k-NN SVM	Classification of the data and assess the correct identification of the type of DFU	Infrared thermal images from plantar foot images America Dynamic cooling challenge of 39 active DFU patients was used. 14 had infected or ischaemic wound 25 had a healing wound Controlled environment	Accuracy Sensitivity Specificity	N/A
[3]	Q1	MSA K-Means clustering	Wound Area Determination (mean-shift) Color Segmentation (K-Means) - Healing versus necrosis versus slough Healing score validation	Image capture box to simplify image capture 12 patients over a period of one year Among the 12 patients, 9 of them were monitored over at least 2 consecutive visits 32 foot ulcer images were collected 28 images were used for the clinical validation of the healing score algorithm.	MCC Computing time KAC	Three experienced wound clinicians outlined the wound area of the wound independently Their delineations for a given wound were combined into one ground truth using a majority vote scheme at the pixel level
[45]	Q1	MSA K-mean clustering	Wound segmentation (mean shift) Colour Segmentation (K-means) - Healing versus necrosis versus slough	Image capture box Nexus 4 smartphone 30 images of simulated wounds 34 images of actual patient wounds America	MCC	Combined labelled wound images

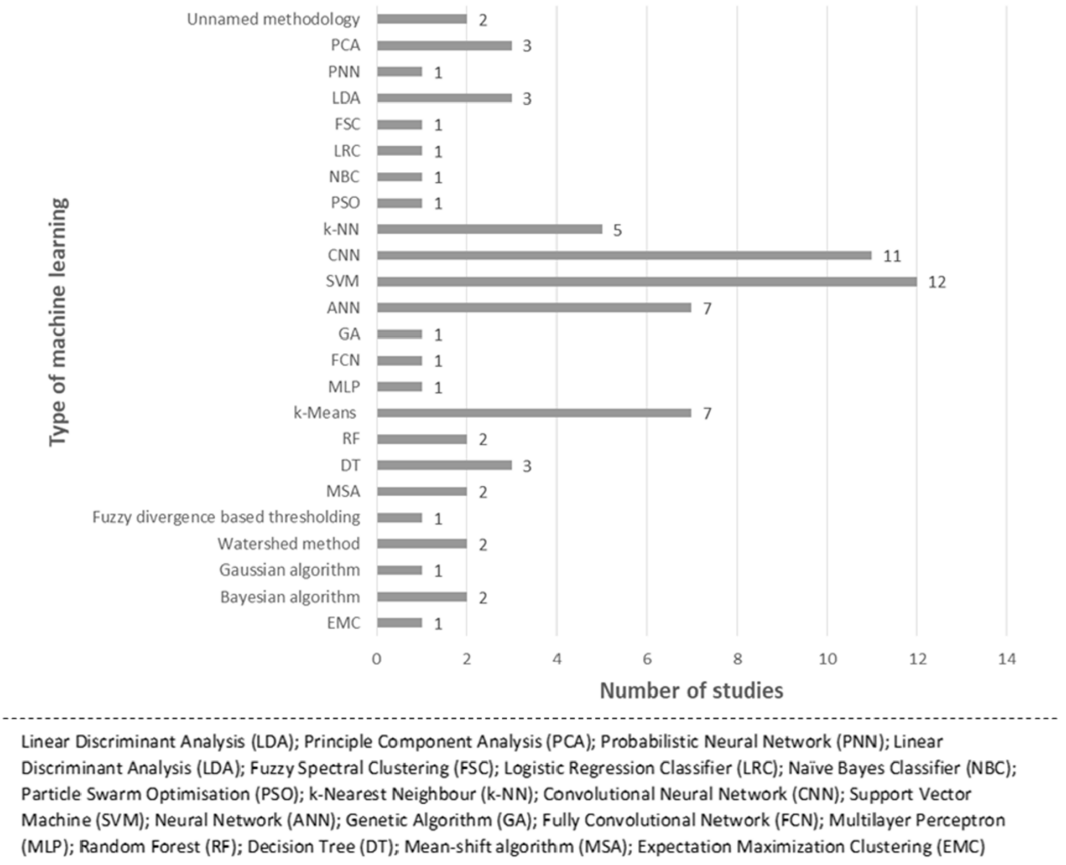
TABLE 2. *(Continued.)* **Characteristics of included studies.**

[51]	Q1	Unnamed machine learning methodology	Wound boundary measurement	124 patients The wounds of 87 patients were measured using the Swift Wound app (iPhone 6 devices) and a ruler The skin surface temperature of 37 patients was also measured using an infrared FLIR by one rater	inter-rater reliability	Three expert raters in clinical wound care measured 45 wounds using the smartphone application Two additional expert raters, measured an additional 42 wounds using the ruler method. All raters were blinded to each other's assessments
[46]	Q1	Novel Two-Stage SVM-Based Classification Single-stage SVM-based method Single-stage ANN-based method Two-stage SVM + CRF refinement	Wound area determination (SVM)	15 patients 100-ft ulcer images over a period of two years Nexus 5 smartphone Controlled conditions (image capture box)	Sensitivity Specificity Computation time	Three clinicians generated the ground truth labels of the wound boundaries for all 100 training images Using electronic pen, it took on average only 30 s for each clinician to label one wound image
[47]	Q1	SVM classifier K-Means clustering Fuzzy k-NN k-NN	Wound classification into tissue types - Granulation versus slough versus necrosis	850 significant tissue regions was selected among the 1200 samples extracted from the 50 typical wound images 48% granulation, 38% slough, 14% necrosis 850 significant region samples labelled by the experts was divided equally into training and test subsets	Sensitivity Specificity Success rate Overall accuracy	50 typical wound images were extracted from the database and submitted to a group of clinicians in order to label them according to the classical colour code using our graphical interface Tissue labelling was carried out twice, one month apart, by four clinicians
[55]	Q1	Mask-RCNN D-CNN ensemble model	Wound detection and boundary segmentation (mask-RCNN) Wound classification (deep CNN ensemble) - Into Wagner ulcer stages 0-5	Diabetic foot ulcer dataset (400 images) was collected from a diabetic clinic in Sri Lanka.	Accuracy	10-foot ulcer images to classify into severity stages by 10 clinicians using macule segmentation
[54]	N/A	PCA	Data analysis for risk prediction of DFU healing	HSI of subjects feet with diabetic foot ulcers Controlled conditions 43 volunteers (12 women; 31 men) Six of the 43 patients had type 1 diabetes and 37 had type 2 diabetes 24 ulcers healed by 12 weeks and a further 7 healed between 12 and 24 weeks. Ten ulcers did not heal within 24 weeks of follow-up.	AUC Sensitivity Specificity PPV	N/A
[48]	Q1	K-Means clustering	Segment foot region	60 feet: 30 diabetic and 30 control feet Controlled conditions 240 separate foot images from 30 people - four images acquired for each foot: two captured by operator 1 and two captured by operator 2	Jaccard similarity index	N/A
[24]	Q1	LRC	Data analysis and Classification of bacterial species isolated from DFU from substrates	1000 groups data are selected to calculate the accuracy of predicted parameters and information	Error of prediction	N/A
[25]	Q1	SVM KNN LDA PNN	Data analysis and Classification of microbial species from DFUs	Real patient samples then cultured e-nose technology sensor array picks up substrates of bacteria	Classification accuracy	N/A

This Table reports the characteristics of the 37 included studies. This includes: authors, year published, SCImago journal rating, machine learning methods used, application of machine learning, data set used, outcome measures and reference test used

introduced a risk of bias, as the images with higher agreement between experts are likely to be less complex and thus easier for the ML algorithm to classify.

In the majority of studies, there was a low risk of bias in reference to the index test used (97%). However, there were a high percentage (50%) with reference tests that were not applicable. This was matched regarding concerns over applicability, with 97% of the index tests being low concern and 50% of the reference tests being not applicable. Without a reference test, it is difficult to compare to the current available practices. There was a large percentage of studies where the sample selection domain was unclear in its risk of bias (43%) and applicability concerns (32%). This is mainly due to the lack of published details regarding the subject selection and lack of clinical details regarding the subjects. Studies with lack of data on which patients received the



This graph reports the number of studies out of the total (n=37) where each type of machine learning was referenced by outcome measure in their results.

FIGURE 2. **Graphical display of machine learning references in included studies.**

index/reference tests were rated as unclear for the risk of bias in the flow and timing domain. Blinding of the data was adequate in most studies, as the ML analysis occurs within a closed system. Overall, 66% of all responses showed low risks for bias and applicability concerns, which increases to 82% of all responses if those that were not applicable are excluded.

The majority of included studies (81%) are published by Q1 and Q2 rated journals as identified using the SCImago index (Table 2). The SCImago quartile rating refers to the impact of the journal in terms of referencing and citations. Q3 and Q4 rated journals are not cited or referenced as often as upper quartile journals, and has been used as a measure of methodological quality in systematic reviews. As this is a subjective rating it was not deemed appropriate for this paper to remove or assess bias of studies based on quartile rating as a sole factor.

C. CLASSIFICATION ALGORITHMS

Three studies used classification to delineate skin into a binary function of healthy skin or abnormal skin (ulcer) [30], [36], [37]. One study created a CNN (DFUNet), which achieved an overall accuracy of 92.5±2.9% [36]. An updated deep learning neural network with an SVM classifier (DFU_QUTNet+SVM) achieved a precision rate of 95.4% on 754 foot images [30]. The third study created a region-based CNN (R-CNN), which achieved a mean average precision (mAP) of 91.8% and an overlap percentage with the gold standard of 95.8% [37]. Two studies used classification to delineate skin into tissue types [38], [47]. Jung *et al.* [38] classified wound images into healing and non-healing using tree models, achieving an area under the curve (AUC) of 0.823 for diabetic neuropathic ulcers. Wannous *et al.* [47] classified wound images into granulation, slough and necrosis tissue types to classify the healing stage using SVM, achieving an overall accuracy of 88%.

Goyal *et al.* (2020) achieved an accuracy of 90.3% and 72.7% for ischaemia classification and infection classification respectively using a CNN [18]. One study classified patients into either a healthy control group, diabetic control group, or diabetic neuropathy group, achieving an average accuracy of classification of 95% [32]. In Yang *et al.* (2018) hyperspectral imaging (HSI) was combined with principle component analysis (PCA) and compared to an algorithm that used oxygen saturation (SpO2) to classify the risk of non-healing in DFUs [54]. The sensitivity of prediction of

TABLE 3. Studies that utilised machine learning algorithms for the process of classification.

Study	Dataset	ML methods	Outcome Measures						Purpose of application
			Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Other	
[23]	6 combined cultures containing 3 different bacterial strains found on DFUs (wild and standard types)	PCA LDA ANN (FFBP and SOM)					LDA = 100 (6hr), 99.7 (24hr) PCA =97.8 (6hr), 97.1 (24hr) LDA + ANN = 100 (6hr,24hr) PCA + ANN = 98.9 (6hr), 99.75 (24hr)		Classified types of bacteria present in sample
[29]	Thermographic images of feet (controlled conditions) Singapore 33 healthy subjects and 33 non-neuropathic diabetic patients	SVM	5-features = 81.81 2-features = 87.88	5-features = 96.97	5-feature = 96.43		5-features = 89.39		Classifies temperature changes across the plantar foot
[30]	754 foot images 1609 skin patches with 542 normal and 1067 abnormal (DFU) Iraq	D-CNN (DFU_QUTNet, GoogleNet, VGG16, and AlexNet) SVM KNN						*Precision (%); Recall (%); F1-score (%)* DFU_QUTNet = 94.2; 92.6; 93.4 DFU_QUTNet+KNN = 93.8; 92.7; 93.2 DFU_QUTNet+SVM = 95.4; 93.6; 94.5	Classifies data into healthy skin and abnormal skin (DFU)
[32]	84 subjects Dynamic plantar pressure measurements during normal gait 56 diabetic patients with and without diabetic peripheral neuropathy 28 control non-diabetic subjects Egypt	SVM					HC/DC = 96.4 HC/ DN = 96.4 DC/DN = 94.6	*Precision (%)* HC/DC = 96.4 HC/ DN = 96.7 DC/DN = 95.2	Classifies plantar pressure data into three binary classification stages: Healthy controls / diabetic controls (HC/DC) Healthy controls / diabetic neuropathy (HC/DN) Diabetic controls / diabetic neuropathy (DC/DN)
[36]	397 foot images (partially controlled conditions) 292 DFU images 105 healthy foot images Britain	CNN (DFUNet, GoogLeNet, AlexNet, LeNet)	93.4 ±3.3	91.1 ±4.4			92.5 ±2.9	*Precision (%); F-Measure; AUC* Proposed DFUNet = 94.5 ±3.2; 0.939 ±0.024; 0.961	Classifies data into normal skin (healthy) and abnormal skin (ulcer)
[18]	1459 DFU images (partially controlled conditions) Britain	Ensemble CNN (SVM) BayesNet (Bayesian algorithm) RF MLP	Ensemble CNN (Ischaemia classification) = 88.6±3.5 Ensemble CNN (infection classification) = 70.9±4.4	Ensemble CNN (Ischaemia classification) = 92.1±2.1 Ensemble CNN (infection classification) = 74.4±5.0			Ensemble CNN (Ischaemia classification) = 90.3±1.2 Ensemble CNN (infection classification) = 72.7±2.5	*Precision (%); F-Measure; AUC* Ensemble CNN (Ischaemia classification) = 91.8±1.9%; 0.902±0.014; 0.904 Ensemble CNN (infection classification) = 73.5±3.6%; 0.722±0.028; 0.731	Feature extraction using RF and MLP. Then classifies data into either ischaemia and non-ischaemia or infection and non-infection
[37]	1880 foot images (controlled conditions) 1775 DFU images 105 healthy foot images Britain	CNN Faster R-CNN (deep learning)						*Speed (ms); Size of the model (MB); mAP; Overlap Percentage (%)* Faster R-CNN = 48; 52.2; 91.8; 95.8	Segmentation and localisation of DFU images
[38]	Basic demographic information on 53,354 patients (52.7% male) and Information on 150,277 wounds used (11% of wounds deemed non-healing) 40 wound types and 37 anatomic locations Venous leg ulcers, pressure ulcers, and diabetic wounds on the lower extremities accounted for 48.8% of wounds America	Gradient boosted tree models						*AUC* Gradient boosted tree model (all wound types) = 0.842 Gradient boosted tree model (diabetic neuropathic ulcers) = 0.823 Gradient boosted tree model (lower extremity diabetic wounds) = 0.799	Classified data from patient records and images of wounds into normal (healing) wounds and abnormal (non-healing) wounds

TABLE 3. ***(Continued.)*** **Studies that utilised machine learning algorithms for the process of classification.**

[53]	Variables of foot amputation decisions from hospital records of 301 diabetic foot patients (83 patients underwent amputation and 218 were managed conservatively) India Wagner grading and Doppler flow measurement used for classification of amputation risk	Tree ensembles					Simple C5.0 tree = 94 (on test data) Boosted C5.0 tree ensemble = 100 (on training data), 96 (on test data)		Classifies risk of amputation using hospital record data of patients with DFUs
[27]	Infrared thermal plantar foot images (controlled conditions) 56 DFU early stage patients America	k-NN SVM ANN Weka (k-NN)					5-feature k-NN = 92.5 Weka 3-feature k-NN = 93.4		Automatic classification of ROIs in thermography images
[26]	Infrared thermal plantar foot images (controlled environment) Dynamic cooling challenge 39 active DFU patients 14 (infected or ischaemic) 25 (healing) America	ANN k-NN SVM	ANN = 75 SVM = 50 5-feature k-NN = 80	ANN = 83 SVM = 100 5-feature k-NN = 100			ANN = 81.25 SVM = 87.5 5-feature k-NN = 81.25		Classification of the data and assess the correct identification of the status of DFU risk and localisation using thermography
[47]	50 wound images 850 significant tissue regions from 1200 samples extracted (48% granulation, 38% slough, 14% necrosis)	SVM K-Means clustering Fuzzy k-NN k-NN	SVM classifier = 77 k-NN = 63 Fuzzy k-NN = 66 k-Means = 39	SVM classifier = 92 k-NN = 86 Fuzzy k-NN = 87 k-Means = 58			SVM classifier = 88 k-NN = 80 Fuzzy k-NN = 81 k-Means = 68	*Success rate (%); Average overlap score (%)* SVM classifier = 84; 79.3 (on test data) k-NN = 5 Fuzzy k-NN = 77 k-Means = 58 *Average overlap score (%) in second test data* 3D multi-view strategy = 73.8% 2D single-view strategy = 69.9	Wound classification into tissue types: Granulation , slough and necrosis
[54]	HSI of feet with diabetic foot ulcers Controlled conditions 43 volunteers (12 women, 31 men; Six T1DM, 37 T2DM)	PCA	PCA = 87.5 SpO2 = 50	PCA = 88.2 SpO2 = 88.2	PCA = 91.3 SpO2 = 85.7			*AUC* PCA = 0.66 SpO2 = 0.88	Classification of risk status and DFU healing prediction using HSI
[24]	1000 groups data are selected to calculate the accuracy of predicted parameters and information	LRC						*Error of prediction (%)* LRC = 2.8 (predicted intention status)	Data analysis and Classification of wound status.
[25]	Bacterial substrates from real patient sampled then cultured on 4 mediums: Blood agar (1,750 data), Mueller Hinton (1,750 data), Mac Conkey (1,000 data), Mixed media (4,500 data)	SVM KNN LDA PNN					LDA +PNN = 100 (Blood agar), 96.3 (Mueller Hinton), 100 (Mac Conkey), 96.8 (Mixed media) LDA+KNN = 100 (Blood agar), 100 (Mueller Hinton), 100 (Mac Conkey), 99.6 (Mixed media) LDA + SVM = 100 (Blood agar), 100 (Mueller Hinton), 100 (Mac Conkey), 98.7 (Mixed media)		Data analysis and Classification of microbial species from DFUs using substrate data

Relevant results and details of studies that involved machine learning algorithms to perform classification of data

healing at 12 weeks using PCA (87.5%) was significantly greater than that of SpO2 (50.0%).

Three included studies classified thermal images into regions of interest (ROI) and ulcer areas. Adam *et al.* used an SVM classifier; yielding a classification accuracy of 89.39%, sensitivity of 81.81% and specificity of 96.97% using five features [37]. Vardasca *et al.* (2018) used a 5 class k-nearest neighbour (k-NN) algorithm, achieving a classification accuracy of 92.5%. In 2019, Vardasca *et al.* [26] used a 5 class k-NN algorithm to localise DFUs, achieving best results of 81.25% accuracy, 80% specificity and 100% sensitivity using 5 feature k-NN.

Three studies [23]–[25] looked at the ability of ML to classify data from bacteria isolated from DFUs for wound

TABLE 4. Studies that utilised machine learning algorithms for the process of colour image segmentation.

Study	Dataset	ML methods	Outcome Measures						Purpose of application
			Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Other	
[17]	445 high resolution DFU images 392 used for training classifiers Unseen set of 53 images used for testing	CNN (U-Net) Patch-based CNN SVM	SVM = 80.6 Patch-based CNN = 90 CNN (U-Net) = 91.7	SVM = 89.6 Patch-based CNN = 94.7 CNN (U-Net) = 97.3				*Mean IoU; Dice index* SVM = 0.472; 0.596 Patch-based CNN =0.660; 0.770 CNN (U-Net) = 0.761; 0.845	Diabetic wound segmentation
[40]	70 selected Ulcer images of 64 patients (Partially controlled) 44 diagnosed diabetic	Fuzzy spectral clustering	Proposed Method = 87.3±0.10	Proposed Method = 95.7±0.97			Proposed Method = 91.5±0.35	*JSI; Dice index* Proposed Method = 79±1.08; 86.7±0.97	Ulcer boundary demarcation and estimation by segmentation
[42]	Colour images of pressure ulcers (PU) (n=400), DFUs (n=20) and venous leg ulcers (VLU) (n=20) from a digital database (controlled conditions) Japan Training data for the CNNs included 356 PU images DFU and VLU images were used to verify accuracy of best performing CNN	CNN (best performing CNN = U-Net)	U-Net (on PU dataset) = 99.3 U-Net (on DFU dataset) = 85.8	U-Net (on PU dataset) = 99.3 U-Net (on DFU dataset) = 98.8			U-Net (on PU dataset) = 98.8 U-Net (on DFU dataset) = 97.82	*AUC; Dice coefficient; MCC; Processing time (s)* U-Net (on PU dataset) = 0.997; 0.936; 0.937; 2.61 (GPU) and 11s (CPU) U-Net (on DFU dataset) = 0.982; 0.850; 0.846; 2.19 (GPU) and 7.09 (CPU)	Training a system for DFU wound segmentation using PU data set to teach the CNN
[43]	70 foot sole images of different ethnicities, age groups and gender (Chinese, Asian, African and Caucasian) Healthy subjects (44 male and 26 female participants	Feature extraction using optical image processing (HSV) incorporating plantar anthropometry					First phase = 96 Final system = 100	*Test time (s)* Automated system = 47 (per foot) Manual SWME = 180 (per)	Segmentation feature (ROI) identification to automatically identify suiTable pressure points to apply a monofilament along the plantar surface of the foot
[3]	DFU images 12 patients over a period of one year (controlled conditions) Among the 12 patients, 9 of them were monitored over at least 2 consecutive visits 32 foot ulcer images were collected 28 images were used for the clinical validation of the healing score algorithm.	MSA K-Means						*MCC; computing time (s); KAC* Study designed algorithm: 0.68; 6 (CPU); range: 0.42-0.81	Wound area determination using mean shift and color Segmentation of 3 tissue types (healing, necrosis, slough)
[45]	DFU images captured on a Nexus 4 smartphone 30 images of simulated wounds 34 images of actual patient wounds America	MSA K-Means						*MCC* Fixed optimal parameter settings = 0.403 Customised parameter settings = 0.736	Wound area determination using mean shift and color Segmentation of 3 tissue types (healing, necrosis, slough)
[46]	15 patients 100 foot ulcer images over a period of two years Taken on a Nexus 5 smartphone (Controlled conditions)	Novel Two-Stage SVM-Based Classification Single-stage SVM-based method Single-stage ANN-based method Two-stage SVM + CRF refinement	Single-stage SVM-based method = 68.3 Single-stage ANN-based method = 66.4 Cascaded Two-Stage SVM-Based Classification = 71.4 Two-stage SVM + CRF refinement = 73.3	Single-stage SVM-based method = 86.9 Single-stage ANN-based method = 83.7 Cascaded Two-Stage SVM-Based Classification = 92.8 Two-stage SVM + CRF refinement = 94.6				*Computation time (s)* Single-stage SVM-based method = 15.4 Single-stage ANN-based method = 16.1 Cascaded Two-Stage SVM-Based Classification = 18.8 Two-stage SVM + CRF refinement = 20.5	Wound area determination

TABLE 4. ***(Continued.)*** **Studies that utilised machine learning algorithms for the process of colour image segmentation.**

[48]	60 patients: 30 diabetic and 30 control 240 separate foot images from 30 people - four images acquired for each foot: two captured by operator 1 and two captured by operator 2 (controlled conditions)	K-Means clustering						*JSI* In diabetic feet: Intraoperator = Operator 1: 0.89 (range: 0.84-0.93) Intraoperator = Operator 2: 0.91 (range: 0.84-0.95) Interoperator = mean: 0.89 (range: 0.83-0.94) In control feet: Intraoperator = Operator 1: 0.94 (range: 0.87-0.98) Intraoperator = Operator 2: 0.93 (range: 0.87-0.97) Interoperator = mean: 0.93 (range: 0.90-0.98)	Foot region segmentation in a mobile phone application

Relevant results and details of studies that involved machine learning algorithms to perform segmentation of colour images

healing status. One study evaluated the pathogens of wound infection from gas sensor data, using a logistic regression classifier (LRC), and achieved an error of prediction of 2.8% [24]. The other two studies used linear discriminant analysis (LDA) in conjunction with another ML classifier. Yusuf *et al.* found that LDA and a single ML classifier successfully identified poly and single microbial species with above 98% accuracy [25]. Abdullah *et al.* achieved a minimum accuracy of 97% using principle component analysis and LDA, in conjunction with an ANN classifier to classify bacteria present [23].

One study used variables from hospital records to classify risk of amputation using decision tree ensembles [53]. The simple classifier used two parameters and obtained an accuracy of 94% in the test group. The second classifier was a more complex computer-derived construct that showed 100% accuracy in the principle group and an accuracy of 96% during testing. All outcomes were measured and results for these studies are described in Table 3.

D. COLOUR IMAGE SEGMENTATION ALGORITHMS

Four studies address smartphone applications of ML for image segmentation [3], [45], [46]. MSA was used [3], [45] to colour segment images into three tissue types (healing, necrosis and slough). The former study averaged a Matthews Correlation Coef?cient value of 0.68, compared to wound area delineation by clinicians [3]. This study also used Krippendorff's alpha coefficient (KAC) to measure the agreement of ratings given by clinicians, where a value of 1 indicates perfect agreement. The agreement values ranged from 0.42 to 0.81. The latter study analysed accuracy and reliability of wound segmentation using ML using MCC, where the fixed optimal parameter setting scored 0.403. In contrast, the customised parameter settings scored 0.736 [45]. Wang *et al.* [46] used four different SVMs to determine wound area in DFU images. The two-stage binary classification system was compared to other ML strategies. All ML strategies used achieved a specificity greater than 92%. The two-stage SVM-based classi?er provided the best sensitivity rate to determine wound boundary on DFU images containing wound regions (73.3%). Yap *et al.* (2017) assessed the reliability of an app to standardise the image capture of DFUs [48]. Interoperator reliability was highly rated (JSI = 0.89 (range: 0.83-0.94)).

Two studies looked at uses of CNNs for segmentation of diabetic wounds. Cui *et al.* (2019) compared SVM and two CNNs (U-Net and Patch-based CNN) and found a deep learning CNN (U-Net) performed highly in all outcome measures. U-Net achieved a 91.7% sensitivity and 97.3% specificity. It also achieved a 0.761 mean IoU and 0.845 on the Dice index, which are used to evaluate the overlapping extent of the ground truth and the predicted segmentation (values close to 1 indicate agreement). Ohura *et al.* (2019) compared four CNN systems, and identified U-Net as the best performing CNN, achieving a sensitivity of 85.8% and a specificity of 98.8% on a DFU dataset. U-Net also achieved an AUC score of 0.982, a Dice coefficient of 0.850, and an MCC of 0.846. The time taken to process each image was calculated as 2.61 seconds (GPU) and 7.09 seconds (CPU).

Only one study utilised segmentation of diabetic plantar surfaces to automatically identify areas for application of a Semmes-Weiss Monofilament (SWME) for DPN checks [43]. The machine achieved a 96% accuracy rate on 70 images, which increased to 100% on the same subjects after adapting conditions for imaging. The average time for the automated system was 47 seconds per foot compared to 180 seconds per foot manually. In this study by Dhane *et al.*, fuzzy spectral clustering (FSC) is used to demarcate and segment ulcer borders. The proposed method achieves a sensitivity of 87.3±0.10%, a specificity of 95.7±0.97% and an overall accuracy of 91.5±0.35%. JSI and the Dice index are used to analyse the predicted overlap and achieve scores of 79±1.08 and 86.7±0.97 respectively. All outcomes measured and results for these studies are described in Table 4.

TABLE 5. Studies that utilised machine learning algorithms for the process of thermal image segmentation.

Study	Dataset	ML methods	Outcome Measures						Purpose of application
			Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Other	
[28]	100 thermal images, corresponding to patients' feet with and without inflammation (Uncontrolled background)	GA Otsu's method						*ME; NU; RAE; EMM; Combination* GA = 0.0028776; 0.034485; 0.000437; 0.013672; 0.0129 Otsu's method = 0.0054818; 0.039365; 0.012102; 0.020619; 0.0194	Thermographic images are segmented using an asymmetry analysis, combined with a genetic algorithm, for early detection of foot ulcers
[49]	Colour and thermographic images of 76 diabetic feet (Controlled conditions) Netherlands	k-Means clustering EMC	All machine learning methods used = 98.1±1 EM-LDC for three clusters = 98.4±0.4	All machine learning methods used = 99.1±1 EM-LDC for three clusters = 98.4±0.4					Image segmentation Into foot surface and background to identify ROI on thermography
[50]	Colour and thermographic images of 26 wound images (Uncontrolled conditions)	SLIC (K-means clustering) Compact watershed method						*Dice coefficient (%)* SLIC = 98.25 Watershed method = 97.75	Image segmentation of diabetic foot images for clinician annotation and classification SLIC resulted in the highest Dice similarity coefficient. We
[51]	124 patients Wounds of 87 patients were measured using the Swift Wound app (iPhone 6 devices) and a ruler The skin surface temperature of 37 patients was also measured (thermographic imaging)	Not stated – segmentation algorithm						*ICC (95% CI)* App surface area (n=45) = 1.00 (0.99–1.00) Ruler (n = 42) = length: 0.92 (0.86–0.96), Width: 0.97, (0.95–0.99) App surface area by novice (n = 12) 0.99 (0.99–1.00)	Smartphone application that enables non-contact wound surface area and temperature measurements.

Relevant results and details of studies that involved machine learning algorithms to perform segmentation of thermal images

E. THERMAL IMAGE SEGMENTATION ALGORITHMS

Four studies looked at the use of ML for segmentation of thermal images. In one study, thermographic images are segmented using a genetic algorithm (GA) for the early detection of foot ulcers [28]. Compared to Otsu's method, the GA achieved a smaller relative overall combination value than the other techniques. GA achieved misclassification error (ME), region nonuniformity (NU), relative foreground area error (RAE), and edge mismatch (EMM) scores of 0.0028776, 0.034485, 0.000437 and 0.013672 respectively. GA achieved a combination score of 0.0129 compared to Otsu's method, which achieved a score of 0.039365. Image segmentation into foot surface and background in order to identify ROIs on thermography was performed by one study, which used an expectation maximisation clustering (EMC) algorithm to segment data [49]. All ML algorithms used had an average sensitivity of 98.1±1% and specificity of 99.1±1%, with EMC performing best at 95% accuracy in a test data group.

Another study used image segmentation of diabetic foot images for clinician annotation and classification [50]. Using a k-Means clustering algorithm (SLIC) resulted in the highest Dice similarity coefficient (98.25%), indicating a high level of crossover between SLIC and gold standard manual segmentation. The final study used a segmentation method embedded in a smartphone application for non-contact wound surface area and temperature measurement [51]. High inter-rater reliabilities were observed using the application across all wound sizes (Intraclass correlation coefficients (ICCs) = 1.00 for area measurements); indicating wound app area measurements are accurate. The data shows the reliability of using an ML algorithm for segmentation of wound images (Table 5).

TABLE 6. Studies that utilised machine learning algorithms for both segmentation and classification.

Study	Dataset	ML methods	Outcome Measures						Purpose of application
			Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Other	
[31]	Diabetic wound images collected from open source database Three DFU images used to test method	PSO NB classifier Hoeffding tree classifier	NB Classifier = 100 Hoeffding tree = 100	NB Classifier = 87.5 Hoeffding tree = 77.7			NB Classifier = 90.90 Hoeffding tree = 81.81		Segmentation of diabetic wounds (PSO) and tissue classification (necrotic, granulation, slough)
[33]	2400 images from a diabetic clinic containing DFUs of all 6 Wagner classes	CNN ANN RF SVM LDA GAP					Densenet201 + GAP + SVD + ANN = 96.22% AlexNet + SVM = 58.75% AlexNet + LDA + SVM = 70.63% VGG-19 + SVD + Softmax = 65.41%	*F-1 Score* Densenet201 + GAP + SVD + ANN = 0.9610 AlexNet + SVM = 0.5749 AlexNet + LDA + SVM = 0.6956 VGG-19 + SVD + Softmax = 0.6502	Feature extraction and classification of severity stage of DFUs (CNN, ANN, Random forest and SVM) (6 stages of Wagner Ulcer Grading scale)
[34]	30 images of diabetic chronic wounds (predominantly hands and feet) (Uncontrolled environment) Brazil	CNN Watershed algorithm Otsu method GrabCut algorithm	U-Net_CSR = 91.28±7.40 SegNet_CSR = 85.91±15.51 FCN-32s_CSR = 65.80±21.82 FCN-8s_CSR = 14.38±1.93	U-Net_CSR = 98.76±2.30 SegNet_CSR = 95.99±6.25 FCN-32s_CSR = 94.03±12.87 FCN-8s_CSR = 98.65±1.90			U-Net_CSR = 96.10±4.08 SegNet_CSR = 92.34±9.71 FCN-32s_CSR = 82.13±13.28 FCN-8s_CSR = 31.14±6.69	*Dice coefficient* U-Net_CSR = 0.9425±0.0598 SegNet_CSR = 0.8879±0.1415 FCN-32s_CSR = 0.7585±0.1710 FCN-8s_CSR = .1833±0.0354	Image segmentation (Watershed algorithm, Otsu and GrabCut algorithm) and classification of DFU tissue (CNN) (necrotic, granulation, slough)
[35]	705 foot images (Partially controlled conditions) 600 DFU images 105 healthy foot images Britain	FCN CNN	FCN-AlexNet = 97.9 FCN-32s = 90.4 FCN-16s = 90.0 FCN-8s = 85.4	FCN-AlexNet = 98.5 FCN-32s = 98.9 FCN-16s = 98.8 FCN-8s = 99.0				*Dice coefficient* FCN-AlexNet = 0.869 FCN-32s = 0.899 FCN-16s = 0.897 FCN-8s = 0.873	DFU segmentation and classification (FCN) (Normal (healthy) skin versus abnormal (ulcer) skin)
[39]	First database for testing had a total of 108 images from 17 volunteers with 4 different backgrounds Second database for retraining R-CNN model contained 141 images from 47 new volunteers (Uncontrolled environment)	Retrained mask R-CNN model	Ulcer = 90.29 Necrosis = 88.00 Without risk = 91.32	Ulcer = 90.28 Necrosis = 91.84 Without risk = 89.16			Ulcer = 90.28 Necrosis = N/A Without risk = N/A		Segmentation and classification (Ulcer, necrosis, no risk)
[41]	74 wound images (burn (n = 12), diabetic ulcer (n = 24), malignant ulcer (n = 14), Pyoderma gangrenosum (n=8), venous ulcer (n=7), and pressure ulcer (n= 9)) 222 regions as granulation tissue, 451 regions as slough tissue, and 94 regions as necrotic tissue	Bayesian classification SVM Fuzzy divergence based thresholding					SVM with 3rd polynomial kernel = Total: 86.13, granulation: 87.84, slough: 90.90, necrotic: 79.78 Bayesian classifier = Total: 81.15, granulation: , slough: , necrotic:	*Kappa statistic* SVM with 3rd polynomial kernel = 0.793 Bayesian classifier = 0.704	Image segmentation (Fuzzy divergence based thresholding) Wound tissue classification (Bayesian classification and SVM) (necrotic, granulation, slough)
[52]	84 subjects Dynamic pressure distribution recorded using foot pressure measuring system	Otsu method SVM					SVM = 96.4	*Precision* SVM = 96.7	Classification and Segmentation of pressure data as a foot image

TABLE 6. *(Continued.)* **Studies that utilised machine learning algorithms for both segmentation and classification.**

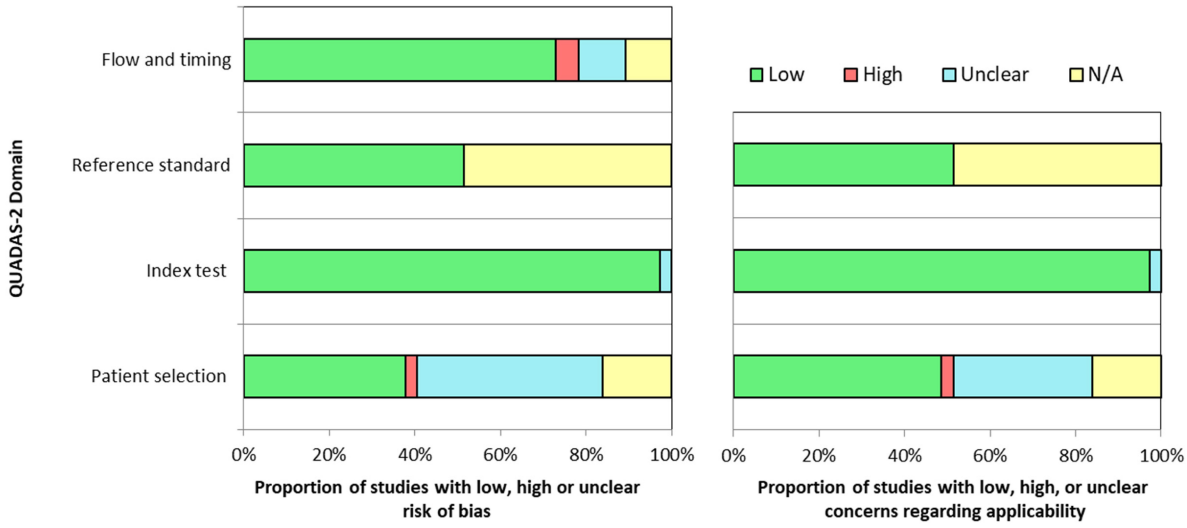
[44]	19 Mexican patients diagnosed with T2DM, but not diabetic foot	ANN (utilising FFBP) PCA					ANN = 97.5		Image segmentation (PCA) Skin macules classification (ANN) (vascular changes versus petechiae versus trophic changes versus trauma macules)
[55]	Diabetic foot ulcer dataset (400 images) was collected from a diabetic clinic in Sri Lanka.	R-CNN D-CNN ensemble model					CNN = 97.5		Wound detection and boundary segmentation (mask-RCNN) Wound classification (deep CNN ensemble) (Wagner ulcer stages 0-5)

Relevant results and details of studies that involved machine learning algorithms to perform segmentation and classification

TABLE 7. **Studies that utilised machine learning algorithms for data and risk analysis.**

Study	Dataset	ML methods	Outcome Measures						Purpose of application
			Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)	Other	
[22]	255 patients (India) Genotyped data for TLR4 receptor SNPs and haplotypes 125 T2DM patients with DFU 130 controls	ANN (back-propagation) MLR					ANN = 83 MLR = 76		Data analysis of TLR4 polymorphisms in DFU patients, looking at predictability for the risk of DFU in T2DM patients

Relevant results and details of studies that involved ML for roles other than segmentation and classification



A graph to show the proportion of studies included in this review which received a rating of low risk, high risk, Unclear risk, or not applicable for the four domains. Graphs show the risk of bias and concerns regarding applicability separately.

FIGURE 3. **Graphical representation of QUADAS quality assessment.**

F. SEGMENTATION AND CLASSIFICATION ALGORITHMS

Three studies used ML techniques for the segmentation and classification of DFU images into tissues (necrotic tissue, granulation tissue and slough tissue) [31], [34], [41]. The study by Babu *et al.* [31] used particle swarm optimisation (PSO) to segment data and enable ROIs to be extracted for the classifiers. Classification was completed via Naïve Bayes and Hoeffding tree processes, which achieved respective accuracies of 90.90% and 81.81%. In the study by Godeiro *et al.* [34] different CNNs were used to classify the images into the three tissue types. U-Net_CSR performed the best in all outcome measures, achieving a classification

TABLE 8. QUADAS quality assessment of included studies.

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
[23]	Unclear	Low	N/A	N/A	Unclear	Low	N/A
[29]	High	Low	N/A	Low	High	Low	N/A
[30]	N/A	Low	Low	Low	N/A	Low	Low
[31]	Unclear	Low	N/A	N/A	Unclear	Low	N/A
[32]	N/A	Low	Low	Low	N/A	Low	Low
[17]	Low	Low	N/A	Low	Low	Low	N/A
[33]	Low	Low	N/A	Low	Unclear	Low	N/A
[34]	Unclear	Low	N/A	High	Unclear	Low	N/A
[35]	Low	Low	Low	Low	Low	Low	Low
[36]	Low	Low	Low	Low	Low	Low	N/A
[18]	Low	Low	Low	N/A	Low	Low	Low
[37]	N/A	Low	Low	Low	N/A	Low	Low
[38]	Low	Low	N/A	Low	Low	Low	N/A
[28]	Unclear	Low	N/A	Low	Low	Low	N/A
[53]	Low	Low	Low	Low	Low	Low	Low
[49]	Unclear	Low	Low	Low	Low	Low	Low
[39]	Low	Low	N/A	Low	Unclear	Low	N/A
[40]	Low	Low	Low	High	Low	Low	Low
[41]	Low	Low	Low	Low	Low	Low	Low
[50]	Unclear	Low	Low	Low	Unclear	Low	Low
[42]	N/A	Low	N/A	Low	N/A	Low	N/A
[43]	Unclear	Low	Low	Low	High	Low	Low
[22]	Low	Low	Low	Low	Low	Low	Low
[52]	Unclear	Unclear	N/A	Unclear	Unclear	Unclear	N/A
[44]	Low	Low	N/A	Unclear	Low	Low	N/A
[27]	Unclear	Low	Low	Low	Low	Low	Low
[26]	Low	Low	N/A	Low	Low	Low	N/A
[3]	Unclear	Low	Low	Low	Unclear	Low	Low
[45]	Unclear	Low	N/A	N/A	Unclear	Low	N/A
[51]	N/A	Low	Low	Low	N/A	Low	Low
[46]	Unclear	Low	Low	Low	Unclear	Low	Low
[47]	N/A	Low	Low	Low	N/A	Low	Low
[55]	Unclear	Low	Low	Low	Low	Low	Low
[54]	Low	Low	Low	Low	Low	Low	Low
[48]	Unclear	Low	N/A	Low	Low	Low	N/A
[24]	Unclear	Low	N/A	Unclear	Unclear	Low	N/A
[25]	Unclear	Low	N/A	Unclear	Unclear	Low	N/A

Low = Low Risk High = High Risk Unclear = Unclear Risk N/A = Not applicable

accuracy of 96.10±4.08%, 98.76±2.30% specificity and a sensitivity of 91.28±7.40%. Classification into the three aforementioned tissue output classes was also achieved using an SVM after utilising fuzzy divergence for image segmentation [41]. In this study, it was observed that a 3rd order SVM methodology achieved classification accuracies for granulation, slough, and necrotic tissues of 86.94%, 90.47%, and 75.53%, respectively. The system also achieved the highest total accuracy of all tested methods (86.13%), with the highest kappa statistic value (0.793), which assesses agreement between results accounting for chance (where 1 indicates perfect agreement).

Goyal *et al.* [35] classified DFU images into normal (healthy skin) and abnormal (ulcer) skin, using fully

convolutional networks (FCNs). Four different FCNs were tested, with very high speci?city for all regions. FCN-16s and FCN-32s had higher sensitivities, with values of 90.0% and 90.4%, respectively. A similar methodology was used in Maldonado *et al.* [39] where the application of a region-based CNN (R-CNN) was used to segment and classify thermographic diabetic foot images into ulcer tissue, necrotic tissue and tissue with no risk of ulceration. The results illustrated a detection accuracy of 90.29% for ulcers and 88% for necrosis. This demonstrates that the system is capable of successfully detecting and visualising temperature di?erences in samples. Toledo Peral *et al.* [44] classified skin macules (small, flat, coloured areas of skin) using an ANN for early diagnosis of foot changes in DFU, achieving a 97.5% accuracy rate in differentiating between skin macules.

Two studies focussed on neural networks to classify DFU images into six categories, based on the Wagner diabetic foot ulcer grade classification system [33], [55]. In Wijesinghe *et al.* [55] a D-CNN was utilised for both the process of wound detection and segmentation, as well as wound classification, achieving an accuracy of 97.5% on 400 test images. 2400 images were assessed in the paper by Gamage *et al.* [33] adapting an existing CNN (DenseNet-201) by utilising other ML classifiers, such as global average pooling (GAP) and singular value decomposition to boost the power of the network. The best performing ML model in this study (DenseNet-201+GAP+SVD+ANN) achieved an accuracy of 96.22% and an F-1 score of 0.9610. The F-1 Score is a measure of a test's accuracy, where a score of 1 reflects perfect precision and recall.

One study applied SVM to the classification and segmentation of dynamic plantar foot pressure measured by innovative pressure insoles [52]. This study achieved an accuracy and precision of 96.4% and 96.7%, respectively. The data suggests plantar pressure may be a good screening tool in DFU risk, however inaccuracies in method reporting limit the power of this study. A large random sample size repeat of this study design would improve its reliability for the general population. All outcomes measured and results for these studies are in Table 6.

G. DATA ANALYSIS ALGORITHMS

One study compared multivariate linear regression (MLR) to an ANN that utilised back-propagation for the process of data analysis of genetic data (TLR4 haplotypes) in DFU patients [22]. Assessing ML for predicting the risk of development of DFU in type II diabetes (T2DM), the ANN model was able to predict with 83% accuracy compared to MLR (74% accuracy). All outcomes measured and results for these studies are described in Table 7.

IV. DISCUSSION

A. RESULTS OF THIS REVIEW

1) IMAGE ANALYSIS

This review aimed to identify the current uses and effectiveness of ML in DFU care. Firstly, the main application of ML was in the analysis and classification of images. The main ML algorithms used were CNNs and SVMs, as they showed a high overall accuracy, and the neural network architecture allows for the easy addition of more classifiers, which can boost the accuracy of the network. Classification is a common task in the treatment of DFUs, with many different applications. Multiple studies applied ML algorithms to the classification of slough, necrotic and granulation tissue [34, 41]. This has important implications for the management of chronic non-healing ulcers. However, tissue identification is limited by the lighting of the images obtained, potentially leading to misclassification. Therefore, to improve the classification accuracy in these studies, testing image capture should be undertaken in a controlled environment. Yap *et al.* [48] tested a user app designed to standardise images for segmentation and classification. It showed a high inter- and intra-operator reliability. The app is however, designed for mobile phones, whereas the CNNs and D-CNNs necessary for classification require more memory in order to run. Image analysis by ML is beneficial due to its speed and the autonomous nature of the screening. Manual identification of ulcer severity is challenging, as vision may vary between consultants. Accordingly, automatic severity stage classification is of signi?cant benefit and all the ML techniques analysed here are capable of achieving over 90% accuracy when classifying DFU images. In the study by Godeiro *et al.* [34] U-Net_CSR was able to achieve a classification accuracy of 96.10±4.08%, which was the best out of any CNN network for classification. However, the CNN in this study was trained on 15 images and tested on another set from the same sample, creating potential bias in patient flow. As the image sample sizes were small and the test data may have been more similar to the training data, there may have been a falsely high sense of accuracy.

DFU imaging is time consuming and inaccuracies lead to failed healing. ML therefore offers the ability to accurately classify in a faster timeframe. However, neural networks are only as good as the set of training data on which they are initialised. In order to be accurate in a clinical setting, it would need to be trained on a larger set of data, which would still require annotation to establish ground truth. Whilst ML has been shown to be highly accurate in these studies, dataset limitations mean they may not retain this classification accuracy in a wider dataset. To improve the clinical applicability of these studies, the neural networks and SVMs should be trained on larger, more variable datasets, as the gold standard of training. This shows the need for the development of larger image databases to allow for this training. Research into the use of ML for DFU image processing should build databases for the populations in need. This change in training data means studies of these neural networks and SVM classifiers would be more reflective of clinical practice.

Other than classifying into percentage of each tissue or a binary classifier such as ulcer/no ulcer, the integration of already existing classification systems can help support decision making. By increasing the number of classes in the classification system, these systems were able to classify

images by the ulcer grade classification system. As many different models exist for the classification of DFUs, there is an opportunity to program an ML algorithm that can apply different numbers of classes (such as K-Means) to allow classification using the appropriate system for each patient. Overall, the ability of ML to automatically classify DFU images is highly beneficial for patients and staff alike, through enabling access and reducing clinician labour. The main limitation is that no existing study has used a sufficient training dataset. As such, these systems require testing on patients in clinical settings to determine their accuracy.

2) THERMOGRAPHY

Four studies from this review looked at the use of ML for segmentation of thermal images. Many studies addressing thermographic imaging also included asymmetric analysis. A difference of just 2.2°C has been shown to be the cut off for a possible inflammatory process like ulcer formation [10]. However, asymmetric analysis for DFU would be unable to compare two feet in diabetic patients with diabetic foot deformities. Although asymmetric analysis was successful for 35/37 diabetic feet, the study excluded major deformities. This system would therefore have functional limitations in clinical practice. Further, thermographic imaging requires expensive equipment. This means this tool would have to be integrated into the health system, rather than as part of self-screening. Moreover, thermal imaging requires adjustments for environmental conditions. The majority of studies utilised a period of acclimatisation for temperature normalisation and used a black cloth backing to prevent the background confusing the algorithm. However, one study utilised an FLIR camera, with a point-of-care wound measurement app, which performed well without acclimatisation [51]. Overall, thermography is a very useful tool in the evaluation of the diabetic foot and reliably detects DFUs early [10], [51]. It is limited by an inability to use asymmetry detection on patients with foot deformities and has a high associated cost, prohibiting use in self-monitoring.

3) DATA ANALYSIS

In one study, classification of risk was calculated after analysis of dynamic plantar foot pressure to calculate the ROI and rank the risk [52]. This methodology managed to achieve accuracy and precision of 96.4% and 96.7% respectively. Whilst the data suggests plantar pressure may be a good screening tool, inaccuracies in method reporting limit the power of this study. Using a large random sample of diabetic subjects would improve the study's reliability to classify DFU risk in the general population. Smart insoles for diagnosis would be improved by an ML classifier, as risk classification could influence management. One other study compared MLR to an ANN, which utilised back-propagation for data analysis of genetic data in DFU patients [22]. Assessing ML predicting the risk of development of DFU in T2DM, the ANN model was able to predict with 83% accuracy compared to MLR, which predicted with 74% accuracy. Data analysis enabled the ANN to predict the relationship between genetic material and the risk of DFU. There are many applications for this use of ML, as an ANN can analyse data and process into risk classes with any other type of data. This suggests that the application of ML techniques on large datasets are a beneficial implementation of ML in diabetic patients, as they are able to calculate many variables from input data in a way a human clinician cannot. Unsupervised learning techniques can therefore act to screen patient data from DFU patients and diabetic patients to both calculate risk and probability, but identify links that may also provide insight into future clinical investigations, treatments and preventative measures. The lack of research in this topic is likely due to the lack of large datasets, and the creation of these is of vital importance to future research.

B. LIMITATIONS OF INCLUDED RESEARCH

Whilst filters are not recommended in systematic review methodology, the limits applied by the authors in this review are unlikely to have impacted the articles retrieved. The limits were deemed satisfactory so as to restrict the number of irrelevant articles whilst not impacting on the retrieval of relevant articles. Limits used included human studies only which is relevant to the review question and eligibility criteria. Using the limits to include papers published in English and only papers with full-texts available was necessary due to the scope of this review and is unlikely to have impacted the.

Some of the limitations in this data are caused by methodological bias, as described in the results section (see methodological quality). There are, however, further limitations in the research. Deep networks, which are becoming more popular, require large amounts of RGB colour and thermal DFU images, as classification accuracy is dependent on the quality and size of the dataset. Further, the database requires annotation by medical experts to act as a gold standard, which acts as a barrier in setting up training data. Neural networks are often subject to overfitting, where they learn their training data too well. This means ML is not applicable to populations beyond those trained, for example other ethnicities. This highlights the importance of using diverse data matched to the target population.

V. CONCLUSION

DFUs are a concern for the growing population of diabetic patients around the world. Although the current principles that guide healthcare are comprehensive, there is still an important gap between our current and desired management outcomes. This is the ?rst review to assess the published literature with respect to the efficacy of applications of ML in the diagnosis, prevention and management of DFUs. Current studies relating ML to DFUs are promising, but there is a need for studies to ensure dataset populations are based on those most in need (i.e. outpatient clinics, rural areas, developing countries).

ML offers a way in which the care of DFU patients can be substantially improved; from large scale data analysis, to at-home image analysis for personalised treatment and

management. This review has shown that the primary DFU uses of ML algorithms currently are related to segmentation and classification of images using different modalities (colour image, thermography, etc.) Promise is shown in the ability of neural networks and SVM to be able to achieve high levels of accuracy and specificity, even when not trained on large datasets. This has the potential to allow for accurate testing for patients from their homes, either using smartphone apps or by uploading images to a neural network.

Finally, although the papers included in this review favour the ability of ML to aid healthcare tasks, the limitations of the studies create problems in generalising results from small study groups to the more varied, general population. Future research should facilitate the development of large databases for the target populations to ensure reliability. These databases should contain data regarding diabetic foot care, diabetic care and control, and images of diabetic ulcers in order to allow future ML algorithms to train on data-sets specific to the healthcare population they would inevitably be used on. The development of these databases both at a regional and national level is integral in increasing the available data pool for training such algorithms, and improving the overall accuracy of these systems when confronted by input data which is varied by the patient demographic. Further, ML techniques should also be assessed in randomised-control trials to allow comparison to current standards of care. Comparing the effects of ML on task and data analysis when compared to time and cost to current practices of care is vital for proving the effectiveness of implementing ML into a public healthcare system. Randomised control trials offer a robust methodological approach to help comparison of ML and current standards, and more of these studies need to be done to further the embedding of ML in healthcare. As none of the reported studies included a health economics analysis, this is an essential measure for future studies to report. Implementation of new systems and methods into healthcare is costly, and the effects of these must be calculated to aid national and local services manage funding. These analyses must assess the cost of generating and storing training data from diabetic patients, implementing the ML algorithms to test data, creation of interfaces for interaction with patients and healthcare professionals, as well as a host of other considerations. With more of these analyses from future research, health organisations will be able to more accurately evaluate the decision to integrate ML systems into clinical practice.

DECLARATIONS OF INTEREST

None

FUNDING

None

REFERENCES

[1] K. Bhuyan, "Identification of socioeconomic variables responsible for diabetic heart disease among Bangladeshi adults," *ARC J. Diabetes Endocrinol.*, vol. 5, no. 2, pp. 1–8, 2019. [Online]. Available: https://www.arcjournals.org/pdfs/ajde/v5-i2/1.pdf, doi: 10.20431/2455-5983.0502001.

[2] S. Tesfaye, "Neuropathy in diabetes," *Medicine*, vol. 43, no. 1, pp. 26–32, 2015.

[3] L. Wang, P. C. Pedersen, D. M. Strong, B. Tulu, E. Agu, R. Ignotz, and Q. He, "An automatic assessment system of diabetic foot ulcers based on wound area determination, color segmentation, and healing score evaluation," *J. Diabetes Sci. Technol.*, vol. 10, no. 2, pp. 421–428, Mar. 2016.

[4] D. G. Armstrong, J. Wrobel, and J. M. Robbins, "Guest editorial: Are diabetes-related wounds and amputations worse than cancer?" *Int. Wound J.*, vol. 4, no. 4, pp. 286–287, Dec. 2007.

[5] M. Kerr, G. Rayman, and W. J. Jeffcoate, "Cost of diabetic foot disease to the national health service in England," *Diabetic Med.*, vol. 31, no. 12, pp. 1498–1504, Dec. 2014.

[6] S. P. Pendsey and U. D. Foot, "Understanding diabetic foot," *Int. J. Diabetes Developing Countries*, vol. 30, no. 2, p. 75, 2010.

[7] Z. Aziz, W. K. Lin, A. Nather, and C. Y. Huak, "Predictive factors for lower extremity amputations in diabetic foot infections," *Diabetic Foot Ankle*, vol. 2, no. 1, p. 7463, Jan. 2011.

[8] M. Fernando, R. Crowther, P. Lazzarini, K. Sangla, M. Cunningham, P. Buttner, and J. Golledge, "Biomechanical characteristics of peripheral diabetic neuropathy: A systematic review and meta-analysis of findings from the gait cycle, muscle activity and dynamic barefoot plantar pressure," *Clin. Biomech.*, vol. 28, no. 8, pp. 831–845, Oct. 2013.

[9] L. A. Lavery, K. R. Higgins, D. R. Lanctot, G. P. Constantinides, R. G. Zamorano, K. A. Athanasiou, D. G. Armstrong, and C. M. Agrawal, "Preventing diabetic foot ulcer recurrence in high-risk patients: Use of temperature monitoring as a self-assessment tool," *Diabetes Care*, vol. 30, no. 1, pp. 14–20, Jan. 2007.

[10] L. A. Lavery, K. R. Higgins, D. R. Lanctot, G. P. Constantinides, R. G. Zamorano, D. G. Armstrong, K. A. Athanasiou, and C. M. Agrawal, "Home monitoring of foot skin temperatures to prevent ulceration," *Diabetes Care*, vol. 27, no. 11, pp. 2642–2647, Nov. 2004.

[11] D. Coppini, "New NICE guidelines on diabetic foot disease prevention and management," *Practical Diabetes*, vol. 32, no. 8, p. 286, 2015.

[12] T. B. Santema, E. A. Lenselink, R. Balm, and D. T. Ubbink, "Comparing the Meggitt-Wagner and the University of Texas wound classification systems for diabetic foot ulcers: Inter-observer analyses," *Int. Wound J.*, vol. 13, no. 6, pp. 1137–1141, Dec. 2016.

[13] N. C. Schaper, J. Apelqvist, and K. Bakker, "The international consensus and practical guidelines on the management and prevention of the diabetic foot," *Current Diabetes Rep.*, vol. 3, no. 6, pp. 475–479, Nov. 2003.

[14] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017.

[15] F. Arcadu, F. Benmansour, A. Maunz, J. Willis, Z. Haskova, and M. Prunotto, "Deep learning algorithm predicts diabetic retinopathy progression in individual patients," *NPJ Digit. Med.*, vol. 2, no. 1, pp. 1–9, Dec. 2019.

[16] L. Wu, W. Zhou, X. Wan, J. Zhang, L. Shen, S. Hu, Q. Ding, G. Mu, A. Yin, X. Huang, and J. Liu, "A deep neural network improves endoscopic detection of early gastric cancer without blind spots," *Endoscopy*, vol. 51, no. 6, pp. 522–531, Jun. 2019.

[17] C. Cui, K. Thurnhofer-Hemsi, R. Soroushmehr, A. Mishra, J. Gryak, E. Dominguez, K. Najarian, and E. Lopez-Rubio, "Diabetic wound segmentation using convolutional neural networks," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 1002–1005.

[18] M. Goyal, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang, and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," *Comput. Biol. Med.*, vol. 117, Feb. 2020, Art. no. 103616.

[19] M. D. F. McInnes, D. Moher, B. D. Thombs, T. A. McGrath, P. M. Bossuyt, T. Clifford, J. F. Cohen, J. J. Deeks, C. Gatsonis, L. Hooft, and H. A. Hunt, "Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement," *JAMA*, vol. 319, no. 4, p. 388, Jan. 2018, doi: 10.1001/jama.2017.19163.

[20] P. F. Whiting, A. W. Rutjes, M. E. Westwood, S. Mallett, J. J. Deeks, J. B. Reitsma, M. M. Leeflang, J. A. Sterne, and P. M. Bossuyt, "QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies," *Ann. Internal Med.*, vol. 155, no. 8, pp. 529–536, 2011.

[21] SCImago. (2007). *SJR: SCImago Journal & Country Rank*. Accessed: Aug. 10, 2020. [Online]. Available: http://www.scimagojr.com

[22] K. Singh, V. K. Singh, N. K. Agrawal, S. K. Gupta, and K. Singh, "Association of toll-like receptor 4 polymorphisms with diabetic foot ulcers and application of artificial neural network in DFU risk assessment in type 2 diabetes patients," *BioMed Res. Int.*, vol. 2013, pp. 1–9, Jul. 2013.

[23] A. A. Abdullah, T. W. Jing, C. A. Sie, N. Yusuf, A. Zakaria, M. I. Omar, A. Y. M. Shakaff, A. H. Adom, L. M. Kamarudin, Y. E. Juan, A. Othman, and M. S. Yassin, "Rapid identification method of aerobic bacteria in diabetic foot ulcers using electronic nose," *Adv. Sci. Lett.*, vol. 20, no. 1, pp. 37–41, Jan. 2014.

[24] Z. Yuan, J. Huang, Z. Zhao, A. Zahid, H. Heidari, R. Ghannam, and Q. H. Abbasi, "A compact wearable system for detection and estimation of open wound status in diabetic patient," in *Proc. IEEE Asia Pacific Conf. Postgraduate Res. Microelectron. Electron. (PrimeAsia)*, Oct. 2018, pp. 60–63.

[25] N. Yusuf, A. Zakaria, M. I. Omar, A. Y. M. Shakaff, M. J. Masnan, L. M. Kamarudin, N. Abdul Rahim, N. Z. I. Zakaria, A. A. Abdullah, A. Othman, and M. S. Yasin, "*In-vitro* diagnosis of single and poly microbial species targeted for diabetic foot infection using e-nose technology," *BMC Bioinf.*, vol. 16, no. 1, p. 158, Dec. 2015.

[26] R. Vardasca, C. Magalhaes, A. Seixas, R. Carvalho, and J. Mendes, "Diabetic foot monitoring using dynamic thermography and AI classifiers," in *Proc. QIRT Asia*, Tokyo, Japan, 2019, pp. 1–5.

[27] R. Vardasca, L. Vaz, C. Magalhaes, A. Seixas, and J. Mendes, "Towards the diabetic foot ulcers classification with infrared thermal images," in *Proc. Int. Conf. Quant. Infr. Thermogr. (QIRT)*, Berlin, Germany, 2018, pp. 293–296. [Online]. Available: http://qirt.org/archives/qirt2018/papers/008.pdf, doi: 10.21611/qirt.2018.008.

[28] N. Kaabouch, W.-C. Hu, Y. Chen, J. W. Anderson, F. Ames, and R. Paulson, "Predicting neuropathic ulceration: Analysis of static temperature distributions in thermal images," *J. Biomed. Opt.*, vol. 15, no. 6, 2010, Art. no. 061715.

[29] M. Adam, E. Y. K. Ng, S. L. Oh, M. L. Heng, Y. Hagiwara, J. H. Tan, J. W. K. Tong, and U. R. Acharya, "Automated characterization of diabetic foot using nonlinear features extracted from thermograms," *Infr. Phys. Technol.*, vol. 89, pp. 325–337, Mar. 2018.

[30] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, "DFU_QUTNet: Diabetic foot ulcer classification using novel deep convolutional neural network," *Multimedia Tools Appl.*, vol. 79, pp. 15655–15677, Jun. 2019.

[31] K. Babu and S. Sabut, and D. K. Nithya, "Efficient detection and classification of diabetic foot ulcer tissue using PSO technique," *Int. J. Eng. Technol.*, vol. 7, no. 3.12, pp. 1006–1010, 2018. [Online]. Available: https://www.researchgate.net/profile/Sukanta_Sabut/publication/332683220_Efficient_Detection_and_Classification_of_Diabetic_Foot_Ulcer_Tissue_using_PSO_Technique/links/5cfe2537299bf13a384a5825/Efficient-Detection-and-Classification-of-Diabetic-Foot-Ulcer-Tissue-using-PSO-Technique.pdf

[32] F. S. Botros, M. F. Taher, N. M. ElSayed, and A. S. Fahmy, "Prediction of diabetic foot ulceration using spatial and temporal dynamic plantar pressure," in *Proc. 8th Cairo Int. Biomed. Eng. Conf. (CIBEC)*, Dec. 2016, pp. 43–47.

[33] C. Gamage, I. Wijesinghe, and I. Perera, "Automatic scoring of diabetic foot ulcers through deep CNN based feature extraction with low rank matrix factorization," in *Proc. IEEE 19th Int. Conf. Bioinformat. Bioeng. (BIBE)*, Oct. 2019, pp. 352–356.

[34] V. Godeiro, J. S. Neto, B. Carvalho, B. Santana, J. Ferraz, and R. Gama, "Chronic wound tissue classification using convolutional networks and color space reduction," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2018, pp. 1–6.

[35] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, and J. Spragg, "Fully convolutional networks for diabetic foot ulcer segmentation," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 618–623.

[36] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H. Yap, "DFUNet: Convolutional neural networks for diabetic foot ulcer classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 5, pp. 728–739, Oct. 2020.

[37] M. Goyal, N. D. Reeves, S. Rajbhandari, and M. H. Yap, "Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1730–1741, Jul. 2019.

[38] K. Jung, S. Covington, C. K. Sen, M. Januszyk, R. S. Kirsner, G. C. Gurtner, and N. H. Shah, "Rapid identification of slow healing wounds," *Wound Repair Regener.*, vol. 24, no. 1, pp. 181–188, Jan. 2016.

[39] H. Maldonado, R. Bayareh, I. A. Torres, A. Vera, J. Gutiérrez, and L. Leija, "Automatic detection of risk zones in diabetic foot soles by processing thermographic images taken in an uncontrolled environment," *Infr. Phys. Technol.*, vol. 105, Mar. 2020, Art. no. 103187.

[40] D. Dhane, M. Maity, T. Munglea, C. Bar, A. Achar, M. Kolekar, and C. Chakraborty, "Fuzzy spectral clustering for automated delineation of chronic wound region using digital images," *Comput. Biol. Med.*, vol. 89, pp. 551–560, Oct. 2017.

[41] R. Mukherjee, D. D. Manohar, D. K. Das, A. Achar, A. Mitra, and C. Chakraborty, "Automated tissue classification framework for reproducible chronic wound assessment," *BioMed. Res. Int.*, vol. 2014, pp. 1–9, Jul. 2014.

[42] N. Ohura, R. Mitsuno, M. Sakisaka, Y. Terabe, Y. Morishige, A. Uchiyama, T. Okoshi, I. Shinji, and A. Takushima, "Convolutional neural networks for wound detection: The role of artificial intelligence in wound care," *J. Wound Care*, vol. 28, no. 10, pp. S13–S24, Oct. 2019.

[43] H.-U.-R. Siddiqui, M. Spruce, S. R. Alty, and S. Dudley, "Automated peripheral neuropathy assessment using optical imaging and foot anthropometry," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1911–1917, Aug. 2015.

[44] C. L. T. Peral, F. J. R. Becerril, G. V. Martínez, A. V. Hernández, L. L. Salas, and J. G. Martínez, "An application for skin macules characterization based on a 3-stage image-processing algorithm for patients with diabetes," *J. Healthcare Eng.*, vol. 2018, pp. 1–13, Dec. 2018.

[45] L. Wang, P. C. Pedersen, D. M. Strong, B. Tulu, E. Agu, and R. Ignotz, "Smartphone-based wound assessment system for patients with diabetes," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 477–488, Feb. 2015.

[46] L. Wang, P. C. Pedersen, E. Agu, D. M. Strong, and B. Tulu, "Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2098–2109, Sep. 2017.

[47] H. Wannous, Y. Lucas, and S. Treuillet, "Enhanced assessment of the wound-healing process by accurate multiview tissue classification," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 315–326, Feb. 2011.

[48] M. H. Yap, K. E. Chatwin, C.-C. Ng, C. A. Abbott, F. L. Bowling, S. Rajbhandari, A. J. M. Boulton, and N. D. Reeves, "A new mobile application for standardizing diabetic foot images," *J. Diabetes Sci. Technol.*, vol. 12, no. 1, pp. 169–173, Jan. 2018.

[49] C. Liu, J. J. van Netten, J. G. van Baal, S. A. Bus, and F. van der Heijden, "Automatic detection of diabetic foot complications with infrared thermography by asymmetric analysis," *J. Biomed. Opt.*, vol. 20, no. 2, Feb. 2015, Art. no. 026003.

[50] R. Niri, Y. Lucas, S. Treuillet, and H. Douzi, "Smartphone-based thermal imaging system for diabetic foot ulcer assessment," in *Proc. Journées d'Etude Sur la TéléSanté*, Paris, France, 2019, pp. 1–6.

[51] S. C. Wang, J. A. E. Anderson, R. Evans, K. Woo, B. Beland, D. Sasseville, and L. Moreau, "Point-of-care wound visioning technology: Reproducibility and accuracy of a wound measurement app," *PLoS ONE*, vol. 12, no. 8, Aug. 2017, Art. no. e0183139.

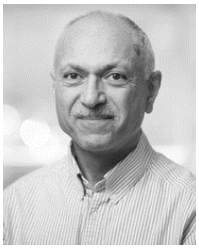
[52] M. D. Sudarvizhi, M. Nivetha, P. Priyadharshini, and J. R. Swetha, "Identification and analysis of foot ulceration using load cell technique," vol. 6, no. 3, pp. 7792–7797, Mar. 2019. [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/60514746/IRJET-V6I3937201909 07-116496-1diomp7.pdf?1567849916=&response-content-disposition=inline%3B+filename%3DIRJET_IDENTIFICATION_AND_ANALYSIS_OF_FOO.pdf&Expires=1604197009&Signature=Q3H8EQ2vCfPsPchD6zZ7a-XENZm76H4tEg8IiBwpbZCKGTyptFAnqL5rU585wBJEe52~Mw1zkvOHSw5B4tw7~tb4zR3AXB-l5z5jiMvF4395pdw8D-n14Qj2gWfgIF78yp33UuxOpRUHp8hcQhltwJTQK5VXy7lZ0Pzs-dBc~OrjI2u~u7NGrK7NYVdamQpAvmzJZi83SURHOaOLpfOd8dVzFHas4wpO76ot62rPaHNkOdlOREKrr86nJhshxx5FkwP56b-evucexkw5KUbcO1dvPX98MsKQ9m8E3dtysTGy7ZBNwoguMB5Av19 Bqp4h2S4XUjCAjQWARnC–uA0Q__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[53] P. U. Kasbekar, P. Goel, and S. P. Jadhav, "A decision tree analysis of diabetic foot amputation risk in Indian patients," *Frontiers Endocrinol.*, vol. 8, p. 25, Feb. 2017.

[54] Q. Yang, S. Sun, W. Jeffcoate, D. Clark, A. Musgove, F. Game, and S. Morgan, "Investigation of the performance of hyperspectral imaging by principal component analysis in the prediction of healing of diabetic foot ulcers," *J. Imag.*, vol. 4, no. 12, p. 144, Dec. 2018.

[55] I. Wijesinghe, C. Gamage, I. Perera, and C. Chitraranjan, "A smart telemedicine system with deep learning to manage diabetic retinopathy and foot ulcers," in *Proc. Moratuwa Eng. Res. Conf. (MERCon)*, Jul. 2019, pp. 686–691.

JACK TULLOCH received the Intercalated B.Sc. degree (Hons.) in sports and exercise medical sciences from Exeter University, Exeter, U.K., in 2020. He is currently pursuing the B.M.B.S. degree from the Exeter University Medical School, Exeter, U.K.



REZA ZAMANI received the B.Sc. degree in chemistry from City University London, in 1983, and the B.Sc. degree in psychology from the University of East London, in 1988, the M.Sc. degree in neuroscience from King's College London, in 1986, and the Ph.D. degree in neural basis of learning and memory from Open University, in 1991. His postgraduate career consisted of a number of Research fellowships at the universities of Manchester and Liverpool, investigating neuronal intracellular calcium metabolism and beta-amyloid toxicity, before moving to the USA, where he spent four years at the University of Charlottesville, looking at synaptic transmission and plasticity in female brain. Then, he moved to U.K., and carried on his research in animal models of Alzheimer's disease at Cardiff, in 1998, and the role of nicotinic transmission in synaptic network plasticity at the National Institute for Medical Research, London, in 2000. He joined the University of Exeter, as a member of Academic Faculty and as a Lecturer of Biomedical Sciences. He is currently an Associate Professor of Medical Sciences and Neuroscience with the University of Exeter, U.K.

MOHAMMAD AKRAMI received the M.Sc. and Ph.D. degrees in mechanical engineering from the University of Manchester, U.K., in 2012 and 2017, respectively. Since then, he started a new position as a Lecturer of Mechanical Engineering with the University of Exeter, U.K. His research interests include the development of engineering analyses of the human joints, applications of computational modeling to predict the human movements, morphology, and rehabilitation.