ORIGINAL ARTICLE

Design and comparison of machine learning-based computer-aided diagnostic techniques to aid diagnosis of diabetes and detection of ulcer-prone regions in the feet using thermograms

Nirmala Vasan Balasenthilkumaran[1] · Barath Ram S[1] · Siddharth Gorti[1] · Sivakumar Rajagopal[1] · Rahul Soangra[2]

Abstract

Purpose According to the International Diabetes Federation, diabetes mellitus (DM) afflicted nearly 463 million people in the world by 2019. Approximately one in every ten individuals is diabetic and this number is projected to increase in the near future. Nearly, 15% of all diabetes cases lead to ulceration of the feet which culminate in amputation if left untreated. Therefore, there is a need for an accurate method to diagnose diabetes and diabetic foot ulcers (DFUs).

Methods In this work, we present machine learning (ML) and image processing-based algorithms to detect and diagnose DFUs and diabetes using thermal images of the feet. DFUs are diagnosed by detecting diabetes and pre-ulcerative foot hotspots or ulcer-prone regions. These hotspots are detected using asymmetry analysis and image processing whereas diabetes is detected by analysing various features of the foot thermogram. In order to find the most optimal diagnostic method for diabetes diagnosis, five classifiers utilizing artificial neural networks (ANN), quadratic support vector machine (QSVM), linear discriminant, logistic regression and Gaussian naïve Bayes algorithms were modelled, tested and evaluated.

Results ANN-based classifier was determined to be most accurate for diagnosing diabetes yielding an accuracy of 93.3% and an F1 score of 0.95. The algorithm to detect ulcer-prone regions of the feet illustrated a detection accuracy and an F1 score of 87.1% and 0.86, respectively.

Conclusion The high accuracies of the proposed algorithms enable its use as a screening tool before blood tests for diagnosis of diabetes and diabetic foot ulcers, facilitating treatment.

Keywords Diabetic foot ulcers · Diabetes · Image processing · Machine learning · Computer-aided diagnosis · Diagnosis · Neural networks

Introduction

Diabetes mellitus (DM), commonly referred to as diabetes, is generally used to define a group of metabolic disorders characterized by increased blood sugar levels over a prolonged period. Increased blood glucose levels are also known as hyperglycaemia which can lead to failure of various organs such as nerves, heart, kidney, eyes and blood vessels (American Diabetes Association 2009). The number of people afflicted by diabetes mellitus is predicted to increase to 700 million by 2045 (Saeedi et al. 2019). The rise in sedentary lifestyles and unhealthy dietary habits are the main cause of the increase in the prevalence of DM worldwide. One of the consequences of DM is ulceration in the legs (especially in the foot) known as diabetic foot ulcers (DFUs). It occurs in nearly 15% of patients affected by DM. Diabetic foot ulcers are a devastating complication of diabetes and are characterized by ulceration, neuropathy and peripheral arterial disease of the lower limb coupled with a lack of protective sensation (Alexiadou and Doupis 2012). They are one of the most common foot injuries to result in lower extremity amputations. Hotspots are identified by comparing corresponding regions of the contralateral foot (regions with the temperature at least 2 °C hotter than the contralateral site). These hot regions are sometimes accompanied by inflammation. Approximately 20% of the overall healthcare expenditure on diabetes can be attributed to DFUs (Skrepnek et al., 2017). Often, if left untreated, they spread severely and can lead to needing to

✉ Sivakumar Rajagopal
  rsivakumar@vit.ac.in

1 Vellore Institute of Technology, Vellore, Tamil Nadu, India

2 Chapman University, Orange, USA

amputate the limb and in certain extreme cases, even lead to death. DFUs are preventable through early diagnosis, treatment of the signs of DFUs or detection of potential ulcerative hotspots. However, early detection of DFUs would require frequent examination by medical professionals which is not feasible for the patient or the medical industry because it is usually recommended that the feet need to be examined daily for signs of pain, skin complexion changes, swelling and cuts or bruises. Increased temperatures may also indicate other diabetic foot complications such as neuropathic ulcers, Charcot foot or osteomyelitis (Boulton and Whitehouse 2020). Conversely, a decrease in foot temperatures indicates vascular insufficiency in the feet.

Thermal imaging is a useful modality for identifying these hotspots in diabetic foot patients. Rapid temperature changes on the plantar surface of the foot in diabetic patients can be considered as an early sign of ulcer formation (Davenport and Kalakota 2019). Increased temperatures are sometimes present up to a week before ulcer formation. Detection of such pre ulcerative hotspots will aid in the early detection of diabetes. These temperature changes are unnoticed by the patient as diabetes sometimes damages the peripheral nerves and hence the patient may be unable to feel pain or perceive the temperature increase in the foot (Quinn et al. 2019). Infrared thermography (IRT) or thermal imaging is the process where a thermal camera is used to create an image using infrared radiations. Thermal cameras are designed to detect radiations in the long-infrared range (9–14 μm) and produce images of the radiations detected; these images are called thermograms. The black body radiation low states that, every object with a temperature above absolute zero emits infrared radiations (Usamentiaga et al. 2014). An increase in temperature of the body increases the amount of radiation; therefore, thermography allows one to see variations in temperature (Armstrong and Lavery 1998).

Early detection, management and treatment are of vital importance for a better prognosis of diabetes and DFUs. Early detection of hotspots in the form of ulcer-prone regions will aid treatment and help to avoid surgeries and amputations. Careful and routine inspection of feet is one of the most efficient measures for preventing diabetic foot complications. However, several studies have discovered that primary care physicians do not regularly examine the feet of diabetic patients during health check-ups and inadequately detect diabetic feet (Davenport and Kalakota 2019). Detection of ulcer-prone regions in diabetic patients will enable them to undertake treatment and facilitate a better prognosis of the disease.

Artificial intelligence (AI), machine learning (ML) and other automated technologies are increasingly beginning to be applied to healthcare. Currently, several research studies suggest that AI can even outperform humans at key healthcare tasks, such as diagnosing diseases. Today many ML-based algorithms are outperforming radiologists in spotting malignant tumours in CTs and MRIs and helping researchers in conducting clinical trials (Davenport and Kalakota 2019). Machine learning can be defined as a statistical technique that determines relationships used to fit models of data. ML is one of the most common forms of AI and can be of several types. Computer-aided diagnosis (CAD) refers to a method of assisting clinicians in diagnosis using computer-generated outputs. These systems have been used to assist radiologists and conventionally relied on manually engineered features and domain knowledge; however, current approaches employ artificial intelligence to discover patterns and latent features in the radiological images.

Computer-aided diagnostic techniques involving the use of thermography to detect diabetic foot ulcers have been proposed by many researchers. Thermograms are usually segmented, processed and then classified. Quinn et al. (2019) detected ulcerative foot hotspots using a methodology involving algorithms such as CNN for feet detection, K means clustering for background removal, intensity-based registration for alignment, subtraction and shape-based classification to detect hotspots from thermograms. Bennetts et al. (2013) used k means clustering to explore differentiation of regional peak plantar pressure in diabetic patients. In his research, Adam et al. (2018) used discrete wavelet transforms (DWT), higher-order spectra (HOS) techniques and an SVM classifier to diagnose diabetic foot using plantar foot thermograms. Vilcahuaman et al. (2015) proposed methods to detect diabetic foot hyperthermia by assessing the mean temperature of the left and right feet obtained by infrared thermography. Liu et al. (2015) proposed a system to automatically detect diabetic foot using infrared thermography and asymmetry analysis-based algorithms. Netten et al. (2013) designed a system implementing high-resolution infrared thermal imaging to detect signs of diabetic foot complications. Hutting et al. (2020) proposed methods to monitor thermal asymmetry (difference in mean plantar temperatures of the affected and unaffected feet) and assess the severity of diabetic foot infections. Maldonado et al. (2020) developed a system to detect high-risk zones such as ulcers and necrosis in diabetic foot patients by calculating mean values of temperature increments and decrements. Astasio-Picado et al. (2018) used thermal images to analyse the temperature variability of the foot by segmenting the sole into four areas of interest.

Methods

In this work, we developed two algorithms, one algorithm focuses on the detection of ulcer-prone regions or potential hotspots and the other algorithm is used to diagnose DM from foot thermograms. The algorithms were developed in

MATLAB R2020A. Image processing toolbox was used for pre- and post-processing of images, the deep learning toolbox was used to model the artificial neural networks and finally, the classification learner app of the machine learning toolbox was used to model the other machine learning models. Confusion matrices and receiver operating characteristic (ROC) curves were plotted to view classification results.

Dataset preparation

A labelled dataset prepared by Hernández-Contreras et al. (2019) is used to train and test the classifiers. The dataset consists of foot thermal images of 122 diabetic patients and 45 non-diabetic subjects. Two images, namely, the thermogram of the left foot and the thermogram of the right foot, are available for each subject. However, for real-time implementation, the thermograms were combined into one image using Adobe Photoshop and background noise was added to the image. So finally, a thermal image consisting of thermograms of both feet corrupted with noise is used as the input. This dataset is used to train the classifiers to detect diabetes and is also used as input to test the performance of the hotspot detection (ulcer-prone regions) algorithm. The dataset was validated by Dr. Sathish K (consultant radiologist). He identified potential ulcer-prone regions in all the foot thermal images and these ulcer-prone regions constitute hotspots.

Algorithm for detection of pre-ulcerative hotspots

The algorithm designed for the detection of ulcer-prone regions primarily utilizes image processing operations along with asymmetry analysis and an ANN. A block diagram outlining the proposed methodology is included in Fig. 1.

Image processing refers to the method of analysing and performing various operations on an image to enhance the quality and extract relevant information from the image (Kapoor and Prasad 2010). To achieve the desired quality,

Fig. 1 Block diagram delineating the hotspot detection algorithm

the foot thermogram images are post-processed using various image processing techniques such as segmentation, resizing, thresholding and erosion. Segmentation refers to the process of dividing an image into numerous segments to extract and isolate the desired region of interest from the whole image for further processing. Image resizing includes altering the dimensions of images while retaining all relevant information. Thresholding estimates threshold values and divides images into distinct regions. Erosion is a morphological operation that shrinks the objects of interest removing unwanted anomalies and perimeter of the region using a predefined image probe to achieve useful contrasting results from the image.

Segmentation and pre-processing

The input images are passed onto a pixel-based segmentation algorithm which detects thermogram pixels and filters out noise and other artefacts. Pixels that contain information of thermograms have similar colour, intensity and average colour values. This property is exploited to design the segmentation algorithm. The algorithm computes the average colour value for each pixel, compares it with a threshold and retains important pixels filtering out noise and other artefacts. The average colour value is obtained by averaging the red, green and blue pixel values. A threshold of an average colour value of 10 pixels was used after inspecting the pixel values of the foot and testing out multiple thresholds ranging from 0 to 30. A segmentation algorithm with a threshold of 10 exhibited the best performance (in terms of noise removal and retaining pixel values corresponding to the foot). Then, indices corresponding to the extremities of the feet are obtained and two images corresponding to the left foot and the right foot are segmented from the input image. The images are finally filtered using a low pass averaging filter of window size 3, enhanced and resized to a common size (155×60).

Conversion to temperature-weighted images

An ANN is created to convert every pixel of a thermogram to its temperature equivalent. The red, green and blue pixel values are used as input and the temperature is obtained as the output. A 3–10-1 neural network architecture is used for fitting (Fig. 2). The network was trained and tested using twelve thermal images and their corresponding temperature values (these values were provided in the dataset); this resulted in 65,520 sets of values. Seventy percent of these values were used for training, 15% for validation and the final 15% were used for testing. Validation was conducted at 26 epochs. ANN are multi-layer interconnected computational systems inspired by the biological neural networks in animals and are used to recognize patterns. Every neuron
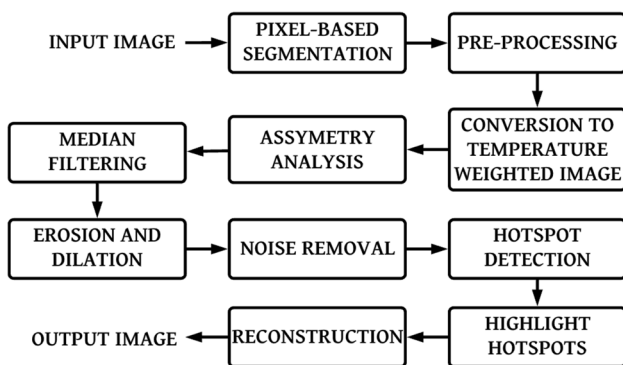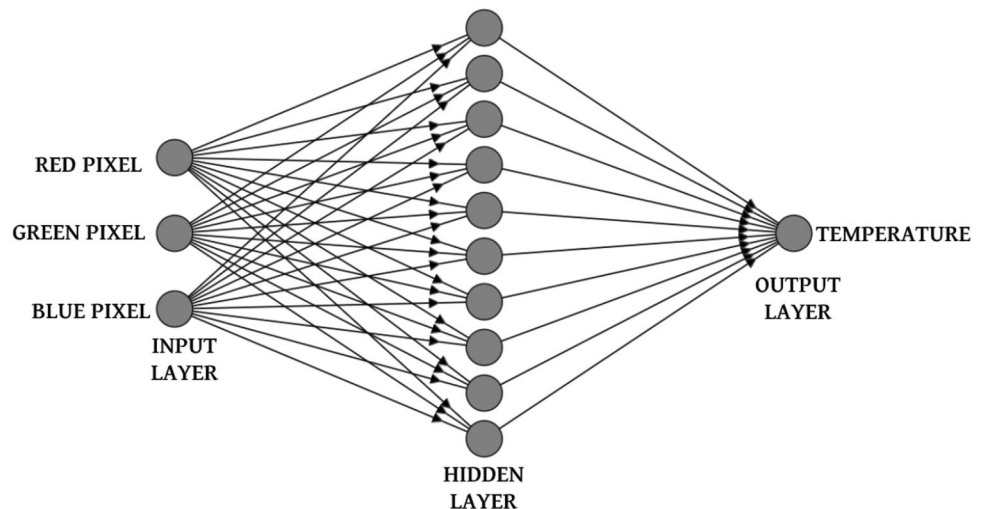
Fig. 2 Neural network architecture for RGB to a temperature conversion

is interconnected to other neurons and every connection has a weight or a real number associated with it which controls the signal between these connected neurons. An ANN can alter its structure based on the input data and is achieved by adjusting the weight of the connection. ANN can be used for classification, regression and clustering.

Asymmetry analysis

Generally, the ulcerated tissues and the ulcer-prone regions in the foot are hotter than the healthy tissues. This property is used to visualize the ulcer-prone tissues as hotspots in the foot thermograms. This detected asymmetry in the temperature distribution in the thermograms verifies the existence of hotspots in the patient's feet. Corresponding regions of the temperature-weighted images of the left and right feet were compared and asymmetry analysis was performed. The images were aligned by detecting the feet extremities in the segmented left and right foot thermograms, resizing the feet to common dimensions and finally, flipping any one of the two processed foot thermal images (left or right, depending on the difference image to be obtained). This was accomplished by first documenting the location (in terms of pixel indices) of the great toe, heel extremity and left and right extremities and then cropping the images using these indices to remove the outermost regions. The images are resized to a common size. The feet are finally aligned by flipping any one of the two processed thermal images. The temperature differences were identified by subtracting the two thermal images after alignment. Two different images are obtained for each subject, the first image is obtained by subtracting the flipped and processed temperature-weighted image of the right foot from the processed temperature-weighted image of the left foot and the second difference image is obtained by subtracting the flipped and processed temperature-weighted image of the left foot from the processed temperature-weighted

image of the right foot. A temperature difference greater than 2.2 °C signifies a hotspot or an ulcer-prone region and therefore, 2.2 °C is used as a threshold to create binary difference images. Binary thresholding was performed on the subtracted images and this finally results in the formation of different images. The image resembles a black and white mask with the white regions corresponding to regions with high-temperature differences.

Image processing for hotspot detection

The difference images were then subject to median filtering using a 5×5 median filter to remove minute inconsequential temperature differences and other background noise. Morphological operations such as erosion and dilation were then performed using line-shaped structuring elements (a 1×6 array of ones is used as the structuring element) to remove false hotspots that arise due to improper alignment of both feet. Bright regions with an area of fewer than 200 pixels were identified, filtered and removed. The biological nature of foot ulcers such as their points of origin and spread is taken into account to remove false hotspots in post-processing. This is accomplished by dividing the image into 16 different regions, identifying the hotspots in these regions and assessing its origin and spread. Hotspots that originate in the centre of the foot are removed. The difference image corresponding to the left foot is flipped. The masks used to reconstruct the images are finally complete. The white regions of the masks correspond to ulcer-prone regions and this black and white mask is used for reconstruction.

Reconstruction

A reconstruction algorithm that compares the input thermal images of both feet with the black and white mask is developed. Hotspots are obtained from the bright regions of the

mask and are highlighted in the original image. The input thermal images are now modified to highlight hotspots. The white regions in the output (processed thermal image) indicate ulcer-prone regions.

Figure 3 consists of outputs obtained after different steps of the image processing algorithm used to highlight potential ulcerative regions or hotspots. Figure 3a and b represent the input and segmented images, respectively. Figure 3c and d represent the temperature-weighted and initial difference images, respectively. Figure 3e and f represent the different images after thresholding and after applying the median filter. Figure 3g represents the mask used for reconstruction and finally, Fig. 3h represents the output image.

Evaluation

All the thermal images in the dataset were manually passed as input to the hotspot detection algorithm and the output was observed for each input. The hotspots identified by the algorithm were compared with the ulcer-prone regions identified by the radiologist and the results were documented; the detected hotspots were categorized into four categories, namely, true positives (*TP*), true negatives (*TN*), false positives (*FP*) and false negatives (*FN*). The algorithm was tested using 164 thermograms of which 76 thermal images consisted of various ulcer-prone regions. A confusion matrix was plotted and used to evaluate the performance of the

Fig. 3 **a** Input image from our modified dataset. **b** Output of the pixel-based segmentation algorithm. **c** Temperature-weighted images obtained using an ANN. **d** Output of image subtraction and asymmetry analysis. **e** Output on thresholding the difference images. **f** Output after application of a median filter. **g** Output after application of morphological operations and post-processing algorithms. **h** Output after applying the reconstruction algorithm utilizing the hotspots mask
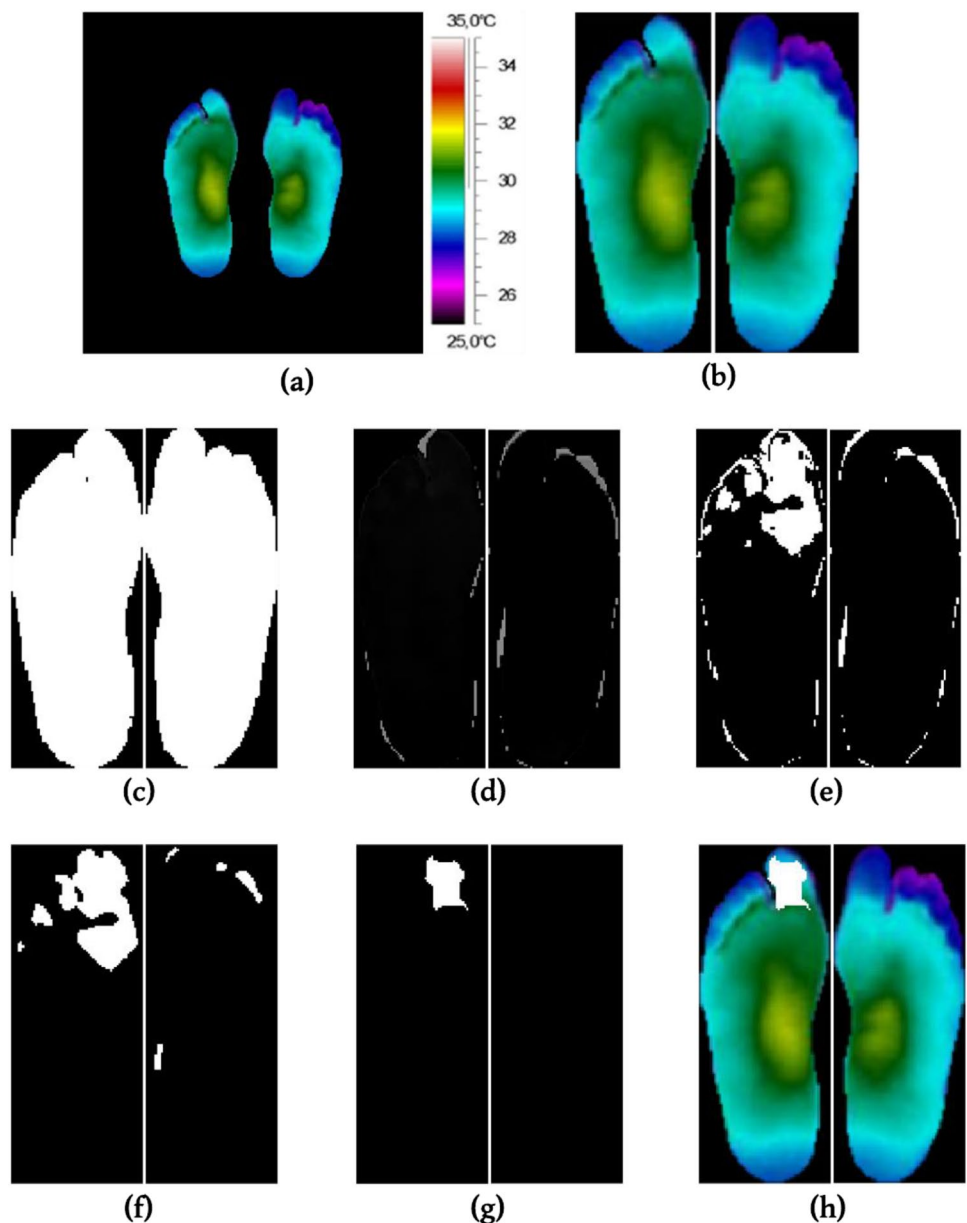
hotspot detection algorithm. The parameters used for evaluation include specificity, precision, recall, F1 score and classification accuracy. Precision represents the correctly identified positives and is represented by Eq. 1, whereas specificity represents the correctly identified negatives and is denoted by Eq. 2. Recall quantifies the true positives detected out of all positives present and is given in Eq. 3. F1 score refers to the harmonic mean of precision and recall thereby giving an estimate of both values and is represented in Eq. 4. Finally, classification accuracy, represented by Eq. 5, signifies the number of correct predictions made out of all predictions. Higher values of these scores indicate better performance.

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

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

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

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

$$Classification\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Detection of diabetes

Diabetes is detected by classifying the foot thermograms. Machine learning-based classifiers were modelled to diagnose diabetes using foot thermograms. Thermograms of a subject's feet are used as input by the algorithm. A block diagram enumerating the proposed system is included in Fig. 4.

Pre-processing

The thermal images were resized, enhanced, filtered and segmented using the same pixel-based segmentation algorithm
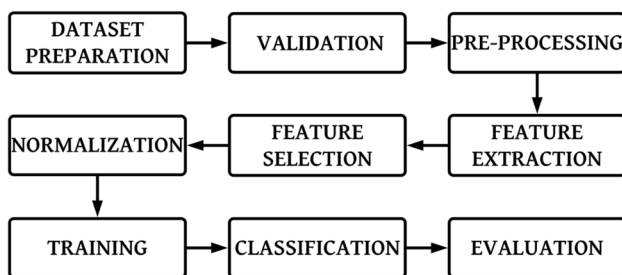

Fig. 4 Block diagram explaining the classification process used by our proposed system

and the same filters discussed in the previous section. Segmentation results in two images, one image corresponds to the thermogram of the left foot and the other image corresponds to the thermogram of the right foot. These thermograms are converted to temperature-weighted images using the ANN described in the pre-ulcerative foot hotspot detection algorithm and represented in Fig. 2. These temperature-weighted images are subject to the feature extraction process.

Feature extraction

Feature extraction used to classify images and features gives us information about the characteristics of the image that is being processed (Kumar and Bhatia 2014). Feature extraction refers to the process of carefully extracting desired features from an input image which makes it easier for the classifier to classify these input images based on their features. Two different types of features are commonly extracted, i.e. first-order statistical features and grey-level co-occurrence matrix (GLCM) features from the foot thermograms (Rasyid et al. 2018). The first-order statistical features extracted include mean, standard deviation and maximum value of temperature. Variables such as contrast, correlation, energy and homogeneity are measured in GLCM and are mainly used to describe the texture of the input image. It additionally includes information on the spatial frequency of the input image (Gogoi et al. 2015). Here, a total of forty-one features were extracted from the temperature-weighted images as a part of the feature extraction process. The temperature-weighted images were compressed to provide a clearer representation of the temperature distribution in a region and the compressed images were also used for feature extraction. Furthermore, the outputs at different stages of the pre-ulcerative foot hotspot detection algorithm were used for feature extraction. The features extracted include information on average temperature, standard deviation, maximum value, GLCM features such as energy, homogeneity, contrast and correlation from the temperature-weighted images and the compressed temperature-weighted images. Other features include the number of pixels corresponding to hotspot regions detected before and after processing, mean value, maximum value and standard deviation of the temperature-weighted image after asymmetry analysis and subtraction. Features are extracted from every input thermogram and their corresponding segmented and temperature-weighted images, these features are used as input parameters by the different ML-based models for classification.

The features extracted were normalized using a normalization function to increase the accuracy of the classifier. The normalized value of an entry 'i' of a feature 'X' is given in Eq. 6.

$$X_i = \frac{X_i - min(X)}{max(X) - min(X)} \tag{6}$$

Feature selection

Feature selection succeeds the feature extraction process. The main objective of the algorithm is to eliminate redundant features and to select the desired features from a raw input image to build an efficient predictive model by increasing classification accuracy and decreasing computational time (Maheta and Shroff 2015). Training the model with a lot of features can reduce classification accuracy even though the features are relevant and contain the desired information about the input image. The chi-squared test-based feature selection process helps us to understand the significance of all the features extracted and using these values, the user keeps the most significant features discarding the other features (Spencer et al. 2020). The chi-squared method is relatively an easier process for univariate feature selection for image classification as the feature interactions are not taken into account when this method is applied for feature selection. All forty-one normalized features were subject to the feature selection process. The best twenty-five ranked features were selected and used as inputs for the classification process.

Classification

The twenty-five features selected were used to train five classifiers. The machine learning algorithms used for classification include an ANN performing pattern recognition, logistic regression, linear discriminant, QSVM and Gaussian naïve Bayes. The classifiers were modelled, trained, validated and evaluated using the deep learning and machine learning toolboxes in Matlab. Confusion matrices and ROC curves were generated for each classifier and were used for evaluation.

ANN image classier applies the process of learning to classify inputs into different classes by finding common features in them. ANN can be used for predicting output values by identifying relations in non-linear problems using the training dataset. The weights of the connection are optimized during training using the back-propagation algorithm. A major advantage of using ANN is that the model of the system can be built from the existing data itself.

Linear discriminant analysis (LDA) is a widely used linear classification technique and has been used extensively to identify linear combinations in features that are extracted from images. It characterizes and classifies the features into various categories with the primary objective of reducing the redundant features to increase the accuracy and efficiency in classifying the foot thermogram image

and to produce accurate and understandable classification results (Zhao et al. 2019).

Naive Bayes classifiers are primarily based on the Bayes theorem and the principle that a pair of features used for classification are independent (Jahromi and Taheri 2017). The name arises because it ignores the prior distribution parameters and assumes the independence of the features extracted and selected from the input images. The advantages of this classifier include fast computation, capability to handle both continuous and discrete data and requirement of lesser training data. The Gaussian naïve Bayes follows Gaussian normal distribution and is more efficient when handling continuous data.

Support vector machines (SVM) are predominantly binary classification algorithms based on predictive analysis which assign the input feature data to classes by constructing a hyperplane in a high-dimensional feature space. A quadratic support vector machine (QSVM) uses a quadratic function as the hyperplane for classification. It is generally considered as one of the best and most robust classifiers among the supervised learning-based classifiers (Ali et al. 2015). They are relatively memory efficient but are also characterized by higher computational and training times.

Logistic regression is widely used as a supervised classification algorithm to predict make predictions of categorically dependent variables by analysing relationships between a set of given independent variables in a model by estimating their probabilities using a logistic function also known as the sigmoidal function, which is an S-shaped curve that takes a value between '0' and '1' and converts these probabilities into binary values for further prediction (MurtiRawat et al. 2020). The error is calculated after each training example and is minimized by modifying the weights. Its major advantage is that it is comparatively faster than other supervised classification techniques and it is also very easy to train and implement.

Evaluation

The ML-based diabetes detection algorithm was tested using foot thermograms from 118 diseased subjects and 45 healthy subjects. Seventy percent of the dataset were used for training, 15% were used for validation and 15% were used for testing. The confusion matrices plotted provide information about how well the classifier worked; it includes true positive, true negative, false positive and false negative values. These values are used to calculate various scores used to evaluate the performance of each classifier. Similar to how the previous algorithm was evaluated, the metrics specificity, precision, recall, classification accuracy and F1 score were used to evaluate the classifiers.

Fig. 5 Confusion matrix of the hotspot detection algorithm

Results

The performance of the hotspot detection algorithm, diabetes diagnosis algorithm and the performance of different classifiers are discussed in this section.

Artificial neural network to determine the temperature

The artificial neural network created to fit RGB pixel values and convert them into temperature equivalent values is evaluated by passing ten different thermograms as input and finding the correlation between the output of the neural network and the known temperature values corresponding to different pixels of the image. Only the non-zero temperature values of the thermograms were considered for evaluation. A correlation score of 0.9264 was obtained on evaluation which means that the temperature values predicted by the artificial neural network are 92.64% accurate.

Hotspot's detection algorithm

The algorithm compares the temperatures of both feet and applies various image processing algorithms as discussed to highlight ulcer-prone regions. The algorithm was tested using 164 thermograms of which 76 thermal images consisted of various ulcer-prone regions. The confusion matrix obtained on evaluating the algorithm is included in Fig. 5. Moderate detection accuracy of 87.1% is obtained. The precision, recall, specificity and F1 score of the algorithm are determined to be 0.84, 0.89, 0.85 and 0.86, respectively. Figures 6 and 7 show how the ulcer-prone regions are highlighted by the proposed algorithm for different input images and Fig. 8 shows the output for a healthy test subject.

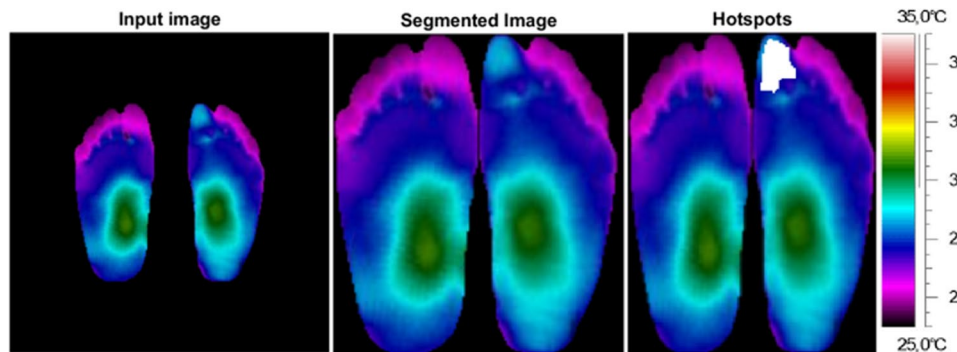**Fig. 6** Hotspots on the left great toe detected by the proposed algorithm

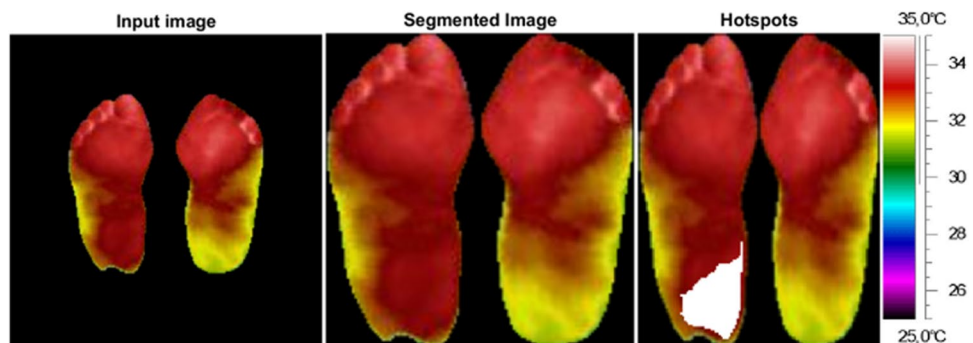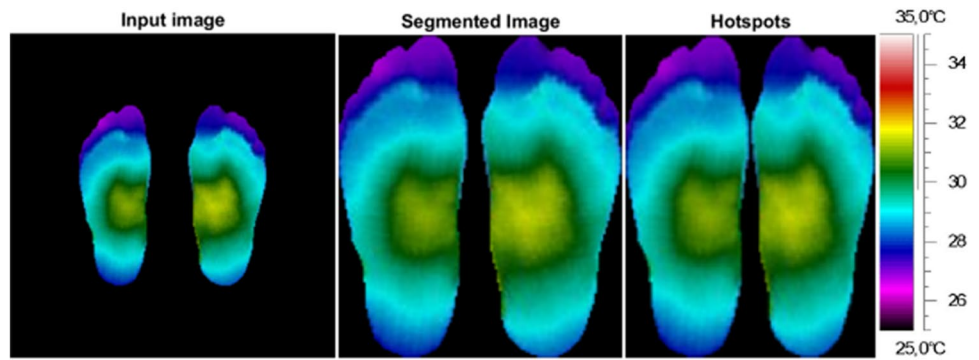**Fig. 7** Hotspots on the right heel detected by the proposed algorithm

Fig. 8 Absence of hotspots in a healthy test subject

Feature selection

Forty-one features were extracted from the input thermogram and the features were subject to the feature selection process which utilizes a chi-sq test-based algorithm. The algorithm ranked the features and the best twenty-five features were used for classification. The features used for classification included a combination of average temperature value, the standard deviation of feet temperature values, maximum feet temperature value, standard deviation and maximum feet temperature value of compressed temperature-weighted images, results of the hotspot detection algorithm and the GLCM feature homogeneity. The five best features include an average temperature of left foot, mean temperature value of both feet, the standard deviation of temperature values of left foot, the standard deviation of temperature values of compressed temperature-weighted images and the maximum temperature of the compressed left foot temperature-weighted image. A heatmap consisting of correlation values of the first five features is included in Fig. 9.
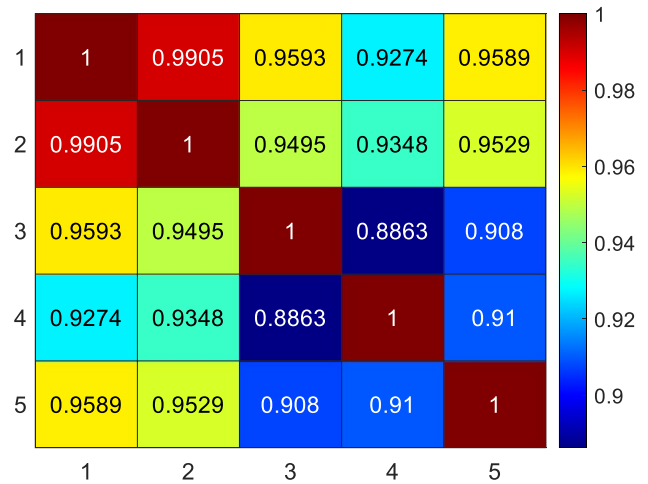
Classifiers used for diabetes detection

Five algorithms, namely, logistic regression, Gaussian naïve Bayes, QSVM, ANN and linear discriminant, were used for the classification process. The algorithms were trained and tested using the dataset. The dataset after cleaning consisted of 163 thermal images of which 118 individuals suffered from diabetes mellitus. Five-fold validation was applied and the dataset was divided into three parts, 70% for training, 15% for validation and the final 15% for testing. The best twenty-five features selected were extracted from the thermal images, normalized and then used as input to train the classifiers. The confusion matrices obtained on evaluating the classifiers are included in Figs. 10, 11, 12, 13 and 14 and the ROC curves along with the area under the curve (AUC) values obtained on evaluating the model are included in Figs. 15, 16, 17, 18 and 19.
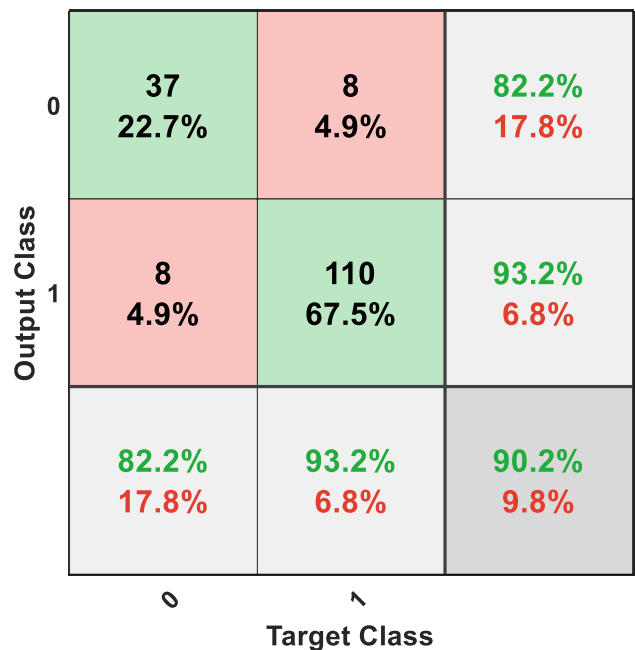

Fig. 9 Heatmap corresponding to the correlation values of the first five features

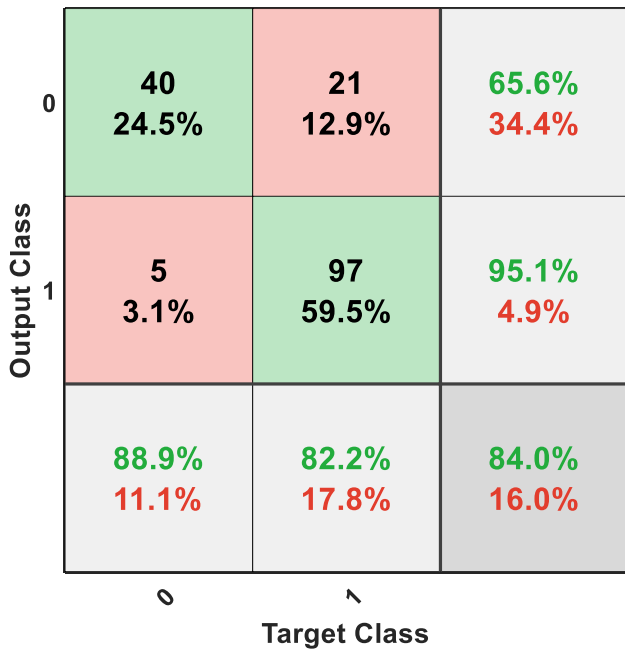**Fig. 10** Confusion matrix of the logistic regression classifier

Fig. 11 Confusion matrix of the Gaussian naïve Bayes classifier
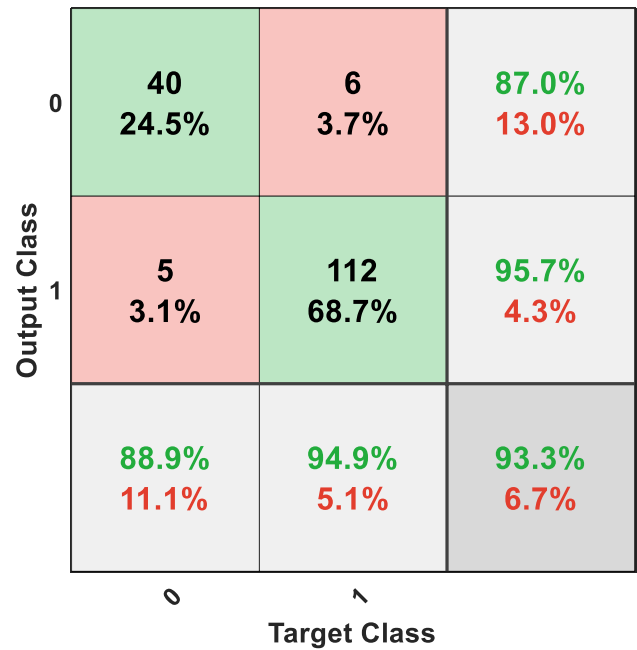

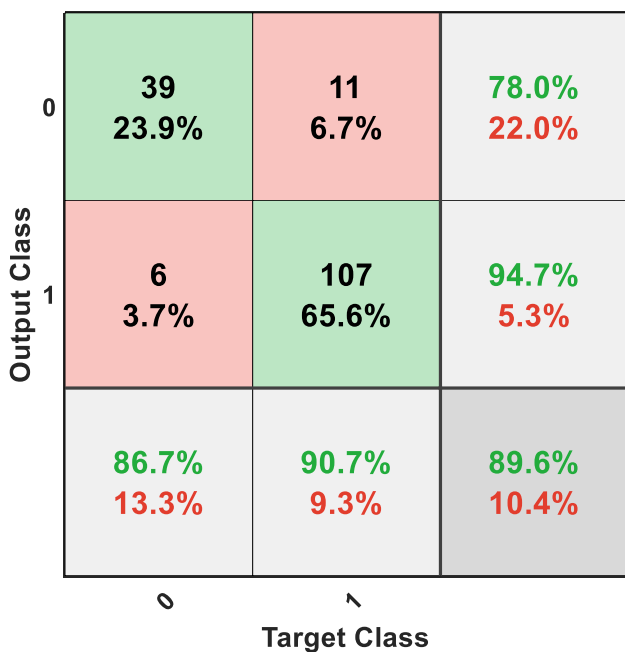**Fig. 13** Confusion matrix of the ANN classifier
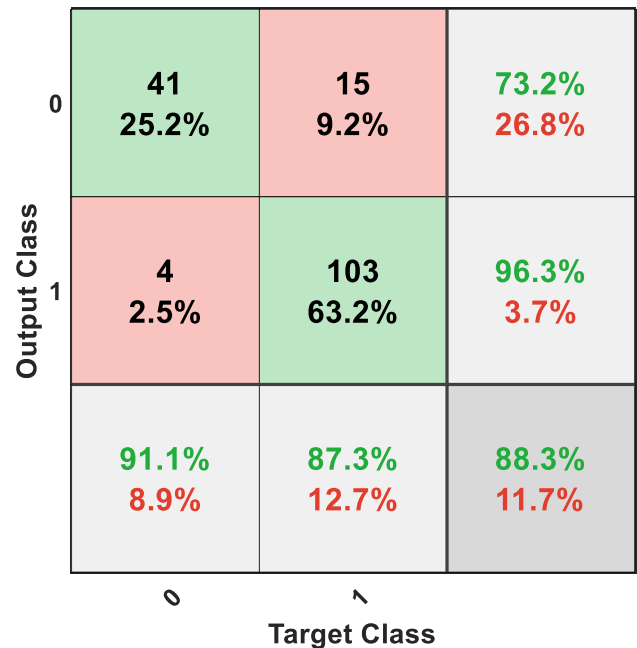

Fig. 12 Confusion matrix of the QSVM classifier

Fig. 14 Confusion matrix of the linear discriminant classifier

Discussion

The five models created were evaluated by calculating classification accuracy, precision, specificity, recall and F1 score. A comparison of the performance of different classifiers is included in Table 1. The ANN-based classifier exhibited the best performance yielding a classification accuracy of 93.3% and an F1 score of 0.95.

Firstly, we decided to use an ANN for the transposition of pixel intensities into temperature after testing multiple algorithms including linear and non-linear regression, polynomial interpolation methods such as Newton polynomial interpolation, Lagrange polynomial interpolation, spline
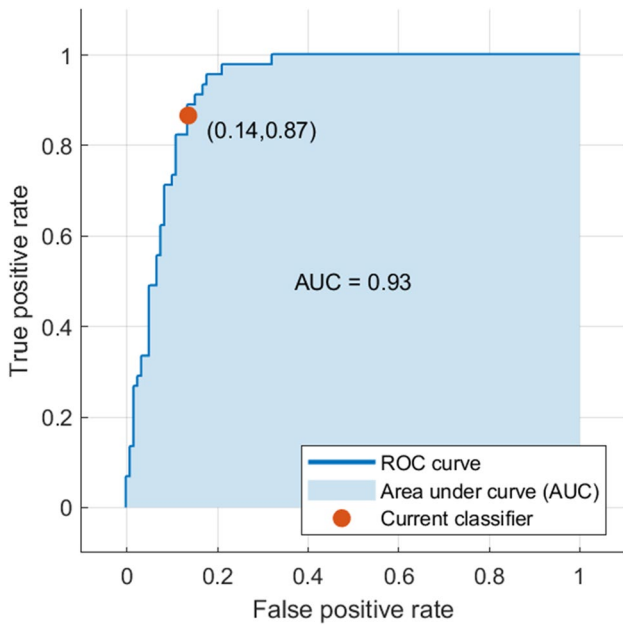
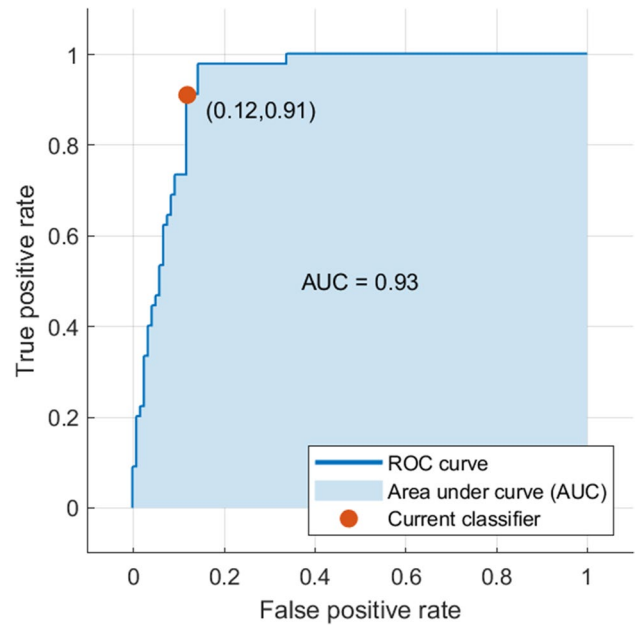**Fig. 15** ROC curve of the logistic regression classifier

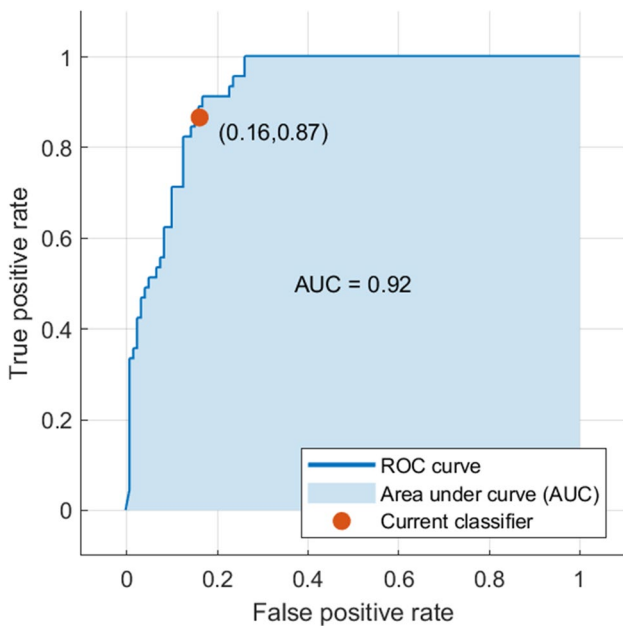**Fig. 17** ROC curve of the QSVM classifier

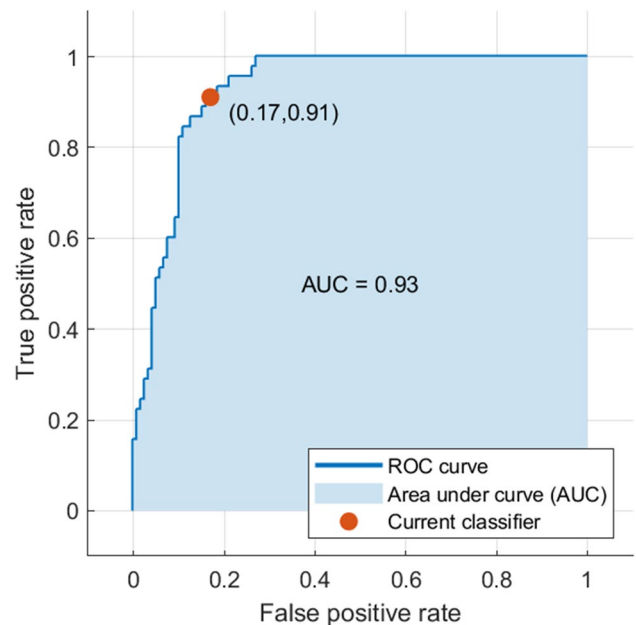**Fig. 16** ROC curve of the Gaussian naïve Bayes classifier

Fig. 18 ROC curve of the ANN classifier

interpolation and cubic spline interpolation, curve fitting and the usage of least mean squares for error reduction. These other methods resulted in a maximum correlation of just over 70% and led to an increased error rate in hotspot identification. However, the usage of an ANN increased the accuracy of pixel transposition by nearly 20% and rectified most of the errors in hotspot identification that were observed when the other methods were used.

Secondly, as the hotspot detection algorithm is highly sensitive to temperature changes, it suffers from an increased false-positive count resulting in an accuracy of 87.1%. Improper alignment of feet before subtraction attributes to almost 50% of the false positives. Hence, with a few modifications to improve the accuracy of the algorithm (deep learning techniques such as R-CNN can be used for better
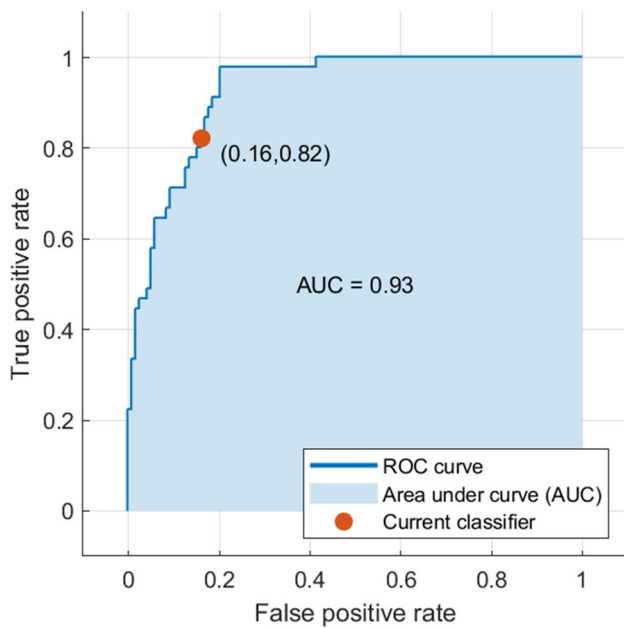
Fig. 19 ROC curve of the linear discriminant classifier

alignment of both feet and CNN-based classifiers can be implemented for classifying temperature changes) and after conducting additional tests using images from different thermal cameras, the algorithm can be clinically be used to detect regions prone to ulceration and can thereby be used to detect diabetic foot ulcers in the early stages.

To better evaluate the performance of the proposed model, we compare the results of the proposed algorithm with other existing hotspot detection algorithms. In Table 2, we compare our algorithm with other works related to the diagnosis of DFU or detection of ulcer-prone regions. Our algorithm outperforms most of the pre-existing algorithms proposed in the literature and hence can potentially serve as a viable alternative to the existing algorithms. As the proposed algorithm is primarily based on image processing, it offers other advantages such as ease of implementation, lower system requirements, lower storage requirements and faster computational time.

Finally, the classification accuracy of the classifiers used to detect diabetes is limited by a small-sized dataset. The accuracy can be increased by increasing the size of the dataset and by exploring other features that can be used for classification. The algorithm can be further tested and improvised by

Table 1 Comparison of the performance of different classifiers utilized in the proposed design

Classifier	Accuracy	Specificity	Precision	Recall	F1 Score
Logistic regression	90.2%	0.82	0.93	0.93	0.93
Gaussian naïve Bayes	84%	0.89	0.95	0.82	0.88
QSVM	89.6%	0.87	0.95	0.91	0.93
ANN	93.3%	0.89	0.96	0.95	0.95
Linear discriminant	88.3%	0.91	0.96	0.87	0.91

Table 2 Comparison between the proposed system and other works related to the detection of DFU or ulcer-prone regions using ML and image processing

Authors	Diagnostic inputs used	Algorithms used for diagnosis	Results
Vardasca et al. (2019)	Foot thermograms	Classified using kNN, ANN and SVM	Maximum classification accuracy of 81.25% (kNN)
Cruz-Vega et al. (2020)	Foot thermograms	Classified using DFT Net, AlexNet and GoogleNet CNN structures, ANN and SVM	Maximum classification accuracy of 85.3% (DFT Net CNN)
Quinn et al. (2019)	Foot thermograms	Segmented using R-CNN and k means clustering followed by intensity-based image registration and asymmetry analysis	Classification accuracy of 60%
Maldonado et al. (2020)	Foot thermograms	Segmented using R-CNN followed by classification	Detection accuracy of 90% with an error rate of 7.05% in labelled areas for ulcers
Sivayamini et al. (2017)	Foot infrared images	Image segmentation using discrete wavelet transform (DWT) followed by GLCM feature extraction and optimization using cuckoo search and particle swarm techniques	Maximum classification accuracy of 79.109% (particle swarm optimization)
Proposed system	Foot thermograms	Pixel-based image segmentation and image processing followed by asymmetry analysis and pot processing	Final detection and labelling accuracy of 87.1%

acquiring foot thermograms using different thermal cameras and from individuals of different races, regions, lifestyles and origins. The performance of different classifiers is compared in Table 1. It is also important to note that the same thermograms which were misdiagnosed by the hotspot detection algorithm were also misclassified by the diabetes detection classifier. Therefore, accuracy can also be improvised by properly acquiring and processing the input thermograms. Classifiers utilizing ANN were found to be most accurate followed by classifiers utilizing logistic regression, QSVM, linear discriminant and finally Gaussian naïve Bayes algorithms. The performance of QSVM and logistic regression is almost identical with minor differences. QSVM exhibits slightly higher precision and specificity scores whereas logistic regression exhibits slightly higher accuracy and recall scores and both classifiers exhibit equal F1 scores. With a few tweaks, the proposed algorithm can be clinically used to diagnose diabetes.

To better evaluate the performance of the diabetes detection classifier, we compare the results of our best classifier (ANN) with other diabetes detection CAD algorithms proposed by other researchers. In Table 3, we compare our model with other works related to the diagnosis of diabetes. The ANN-based classifiers outperform most of the pre-existing algorithms proposed in the literature and hence can serve as an alternative to existing technologies. Finally, DFUs can be detected in the early stages by correlating the results of the hotspot detection algorithm with the results of the diabetes detection algorithms.

The limitations of our work are as follows. DM cannot be diagnosed with certainty when only foot thermograms are used. These tests need to be coupled with blood glucose tests for an accurate diagnosis. Therefore, right now, our algorithm can only act as a diagnostic aid and be used as a screening tool before blood testing. Our algorithm also fails at diagnosing DM in the early stages; the algorithm is based on detecting high-temperature regions and ulcer-prone regions in the feet. Hence, it cannot aid in the early diagnosis of the disease as these changes in the feet occur only after years of suffering from DM. Finally, our hotspot detection algorithm is validated only for the detection of regions with an increased risk of ulceration. Ulcer-prone tissues are hotter than normal tissues. A dataset consisting of foot thermograms of subjects suffering from DM and at risk of foot ulceration must be acquired and be used to test the proposed hotspot detection algorithm. Only, when this test yields successful results can our algorithm be used to detect pre-ulcerative hotspots for the early diagnosis of DFUs.

Conclusion

A non-irradiant, non-invasive, fast and accurate method was designed, developed and evaluated for the diagnosis of diabetes and detection of ulcer-prone regions in the feet. In

Table 3 Comparison between the proposed system and other works related to diagnosis of diabetes using ML or image processing

Authors	Diagnostic inputs used	Algorithms used for diagnosis	Results
Thirunavukkarasu et al. (2020b)	Facial thermograms	Classified using SVM, K nearest neighbours (kNN) and LDA algorithms after feature extraction	Maximum classification accuracy of 89.37% (SVM)
Thirunavukkarasu et al. (2020a)	Human tongue thermograms	RGB-based colour image segmentation followed by GLCM feature extraction and classification using SVM and CNN	Maximum classification accuracy of 94.28% (CNN)
García-Ordás et al. (2021)	Inputs such as blood pressure, age, number of pregnancies, glucose or insulin levels of a patient	Data and feature augmentation were performed using variational and sparse autoencoders followed by classification using CNN	Final classification accuracy of 92.31%
Lai et al. (2019)	Laboratory information of patients such as blood pressure, age, sex, BMI, blood glucose levels, triglycerides, high- and low-density lipoprotein levels	Classified using gradient boosting machine (GBM) and logistic regression models	GBM exhibited better performance with an AUC and sensitivity of 84.7% and 71.6%, respectively
Sivanandam et al. (2012)	Foot thermograms and palm thermograms	Pearson correlation tests	The use of palm thermograms provided the best results with an accuracy of 73%
Proposed system	Foot thermograms	Segmentation and pre-processing followed by classification using ANN, QSVM, Gaussian naïve Bayes and LDA	Maximum classification accuracy of 93.33% (ANN)

this paper, we analysed the use of different classifiers and determined the best method to diagnose diabetes using foot thermograms. Feature selection illustrated a positive correlation between the average temperature of feet and the occurrence of diabetes. Out of the five classifiers tested, the ANN-based classifier yielded the best results with a classification accuracy, specificity, precision, recall and F1 score of 93.3%, 0.89, 0.96, 0.95 and 0.95, respectively. The hotspot detection algorithm yielded a detection accuracy, specificity, precision, recall and F1 score of 87.1%, 0.85, 0.84, 0.89 and 0.86, respectively. The current study is limited by the small size of the dataset. Consequently, the classifiers and algorithms can be improvised by increasing the size of the dataset. This can be done by manually acquiring thermal images of the feet of diabetic patients and using them to retrain and test the models and algorithms developed. The algorithms can be integrated into an application for easy usage. Thermal imaging provides a low-cost solution and additionally does not cause any side effects; therefore, tests can be repeated multiple times making it particularly useful to detect the spread of ulcerative hotspots. With the help of the proposed methods, diabetes can be diagnosed with greater ease and the detection of ulcer-prone regions would aid in the detection of DFUs, enabling a better prognosis of the disease.

References

Adam M, Ng EYK, Oh SL, Heng ML, Hagiwara Y, Tan JH, Tong JWK, Acharya UR. Automated characterization of diabetic foot using nonlinear features extracted from thermograms. Infrared Phys Technol. 2018;89:325–37. https://doi.org/10.1016/j.infrared.2018.01.022.

Alexiadou K, Doupis J. Management of diabetic foot ulcers. Diabetes Ther. 2012;3(1):4. https://doi.org/10.1007/s13300-012-0004-9.

Ali MAS, Sayed GI, Tarek G, Hassanien AE, Snasel V, Silva LF. Detection of breast abnormalities of thermograms based on a new segmentation method, Proceedings of the Federated Conference on Computer Science and Information Systems. 2015; pp 255–261. https://doi.org/10.15439/2015F318.

American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2009;32(1):S62–7. https://doi.org/10.2337/dc09-S062.

Armstrong DG, Lavery LA. Diabetic foot ulcers: prevention, diagnosis and classification. American family physician. 1998; 57(6):1325–1338. URL: https://ncbi.nlm.nih.gov/pubmed/9531915. Accessed Aug 18, 2020.

Astasio-Picado A, Escamilla Martínez E, Martínez Nova A, Sánchez Rodríguez R, Gómez-Martín B. Thermal map of the diabetic foot using infrared thermography. Infrared Phys Technol. 2018;93:59–62. https://doi.org/10.1016/j.infrared.2018.07.008.

Bennetts CJ, Owings TM, Erdemir A, Botek G, Cavanagh PR. Clustering and classification of regional peak plantar pressures of diabetic feet. J Biomech. 2013;46(1):19–25. https://doi.org/10.1016/j.jbiomech.2012.09.007.

Boulton AJM, Whitehouse RW. The diabetic foot. In: Feingold KR, Anawalt B, Boyce A, et al., eds. Endotext. South Dartmouth

(MA): MDText.com, Inc. 2020; URL: http://www.ncbi.nlm.nih.gov/pubmed/28121117. Accessed Aug 18, 2020.

Cruz-Vega I, Hernandez-Contreras D, Peregrina-Barreto H, Rangel-Magdaleno JdJ, Ramirez-Cortes JM. Deep learning classification for diabetic foot thermograms. Sensors. 2020;20(6):1762. https://doi.org/10.3390/s20061762.

Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal. 2019;6(2):94–8. https://doi.org/10.7861/futurehosp.6-2-94.

García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I. Diabetes detection using deep learning techniques with oversampling and feature augmentation. Comput Methods Programs Biomed. 2021;202:105968. https://doi.org/10.1016/j.cmpb.2021.105968.

Gogoi UR, Majumdar G, Bhowmik MK, Ghosh AK, Bhattacharjee D. Breast abnormality detection through statistical feature analysis using infrared thermograms. 2015 international symposium on advanced computing and communication (ISACC). 2015; 258–265. https://doi.org/10.1109/ISACC.2015.7377351

Hernández-Contreras D, Peregrina-Barreto H, Rangel-Magdaleno J, Renero-Carrillo F. Plantar thermogram database for the study of diabetic foot complications. IEEE Dataport. 2019. https://doi.org/10.21227/tm4t-9n15

Hutting KH, Aan de Stegge WB, Kruse RR, van Baal JG, Bus SA, van Netten JJ. Infrared thermography for monitoring severity and treatment of diabetic foot infections. Vasc Biol. 2020;2(1):1–10. https://doi.org/10.1530/VB-20-0003.

Jahromi AH, Taheri M. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. 2017 artificial intelligence and signal processing conference (AISP). 2017; 209–212. https://doi.org/10.1109/AISP.2017.8324083

Kapoor P, Prasad SVAV. Image processing for early diagnosis of breast cancer using infrared images. The 2nd international conference on computer and automation engineering (ICCAE). 2010; 564–566. https://doi.org/10.1109/ICCAE.2010.5451827

Kumar G, Bhatia PK. A detailed review of feature extraction in image processing systems. 2014 fourth international conference on advanced computing & communication technologies. 2014; 5–12. https://doi.org/10.1109/ACCT.2014.74

Lai H, Huang H, Keshavjee K, Guergachi A, Gao X. Predictive models for diabetes mellitus using machine learning techniques. BMC Endocr Disord. 2019;19:101. https://doi.org/10.1186/s12902-019-0436-6.

Liu C, van Netten JJ, van Baal JG, Bus SA, van der Heijden F. Automatic detection of diabetic foot complications with infrared thermography by asymmetric analysis. J Biomed Opt. 2015;20(2):26003. https://doi.org/10.1117/1.JBO.20.2.026003.

Maheta H, Shroff KP. A comparative study of various feature selection techniques in high-dimensional data set to improve classification accuracy. 2015 international conference on computer communication and informatics (ICCCI). 2015; 1–6. https://doi.org/10.1109/ICCCI.2015.7218098

Maldonado H, Bayareh R, Torres IA, Vera A, Gutiérrez J, Leija L. Automatic detection of risk zones in diabetic foot soles by processing thermographic images taken in an uncontrolled environment. Infrared Phys Technol. 2020;105:103187. https://doi.org/10.1016/j.infrared.2020.103187.

MurtiRawat R, Panchal S, Singh VK, Panchal Y. Breast cancer detection using K-nearest neighbors, logistic regression and ensemble learning. 2020 international conference on electronics and sustainable communication systems (ICESC). 2020; 534–540. https://doi.org/10.1109/ICESC48915.2020.9155783

Quinn S, Saunders C, Cleland I, Nugent C, Constantino MG, Cundell J, Madill G, Morrison G. A thermal imaging solution for early detection of pre-ulcerative diabetic hotspots. Annu Int Conf IEEE Eng

Med Biol Soc. 2019; 1737–1740. https://doi.org/10.1109/EMBC.2019.8856900

Rasyid MBA, Yunidar, Arnia F, Munadi K. Histogram statistics and GLCM features of breast thermograms for early cancer detection. 2018 international ECTI northern section conference on electrical, electronics, computer and telecommunications engineering (ECTI-NCON). 2018; 120–124. https://doi.org/10.1109/ECTI-NCON.2018.8378294

Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R. IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019; 157:107843. https://doi.org/10.1016/j.diabres.2019.107843

Sivanandam S, Anburajan M, Venkatraman B, Menaka M, Sharath D. Medical thermography: a diagnostic approach for type 2 diabetes based on non-contact infrared thermal imaging. Endocrine. 2012;42(2):343–51. https://doi.org/10.1007/s12020-012-9645-8.

Sivayamini L, Venkatesh C, Fahimuddin S, Thanusha N, Shaheer S, Sree PS. A novel optimization for detection of foot ulcers on infrared images. 2017 international conference on recent trends in electrical, electronics and computing technologies (ICRTEECT). 2017; 41–43. https://doi.org/10.1109/ICRTEECT.2017.36

Skrepnek GH, Mills JL, Lavery LA, David G. Health care service and outcomes among an estimated 6.7 million ambulatory care diabetic foot cases in the U.S. diabetes care. 2017; 40(7):936–942. https://doi.org/10.2337/dc16-2189

Spencer R, Thabtah F, Abdelhamid N, Thompson M. Exploring feature selection and classification methods for predicting heart disease. Digit Health. 2020. https://doi.org/10.1177/2055207620914777.

Thirunavukkarasu U, Umapathy S, Janardhanan K, Thirunavukkarasu R. A computer aided diagnostic method for the evaluation of type II diabetes mellitus in facial thermograms. Phys Eng Sci Med. 2020a;43(3):871–88. https://doi.org/10.1007/s13246-020-00886-z.

Thirunavukkarasu U, Umapathy S, Krishnan PT, Janardanan K. Human tongue thermography could be a prognostic tool for prescreening the type II diabetes mellitus. Evidence-Based Complementary and Alternative Medicine. 2020b;2020:3186208. https://doi.org/10.1155/2020/3186208

Usamentiaga R, Venegas P, Guerediaga J, Vega L, Molleda J, Bulnes FG. Infrared thermography for temperature measurement and non-destructive testing. Sensors (Basel, Switzerland).2014; 14(7):12305–12348. https://doi.org/10.3390/s140712305

Van Netten JJ, van Baal JG, Liu C, van der Heijden F, Bus SA. Infrared thermal imaging for automated detection of diabetic foot complications. J Diabetes Sci Technol. 2013;7(5):1122–9. https://doi.org/10.1177/193229681300700504.

Vardasca R, Magalhaes C, Seixas A, Carvalho R, Mendes J. Diabetic foot monitoring using dynamic thermography and AI classifiers. In Proc. QIRT Asia, Tokyo, Japan. 2019; 1- 5. https://doi.org/10.21611/qirt.2019.027

Vilcahuaman L. et al. Automatic analysis of plantar foot thermal images in at-risk type II diabetes by using an infrared camera. IFMBE proceedings, Springer, Cham. 2015; 51:228–231. https://doi.org/10.1007/978-3-319-19387-8_55

Zhao H, Wang Z, Nie F. A new formulation of linear discriminant analysis for robust dimensionality reduction. IEEE Trans Knowl Data Eng. 2019;31(4):629–40. https://doi.org/10.1109/TKDE.2018.2842023.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.