IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 June 2024, accepted 9 July 2024, date of publication 17 July 2024, date of current version 2 August 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3430194*

RESEARCH ARTICLE

# Improving Automated PSN Assessment in Type 2 Diabetes: A Study on Plantar Lesion Recognition and Probe Avoidance Techniques

**HAFEEZ UR REHMAN SIDDIQUI[1], (Student Member, IEEE), RICCARDO RUSSO[2], ADIL ALI SALEEM[1], SANDRA DUDLEY[3], (Member, IEEE), AND FURQAN RUSTAM[4]**

[1]Institute of Computer Science, Khawaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan
[2]Department of Brain and Behavioural Sciences, University of Pavia, 27100 Pavia, Italy
[3]School of Engineering, London South Bank University, SE1 0AA London, U.K.
[4]School of Computer Science, University College Dublin, Belfield Campus, Dublin 4, D04 V1W8 Ireland

Corresponding author: Furqan Rustam (furqan.rustam@ucdconnect.ie)

This work was supported by University College Dublin, Ireland.

**ABSTRACT** Peripheral Sensory Neuropathy (PSN) affects a large proportion of individuals suffering from type 2 diabetes. To avoid ulceration and other damage to the patient's feet, regular PSN testing, and assessment must be undertaken. Currently, the Semmes-Weinstein Monofilament Examination (SWME) is one of the most widely accepted techniques for PSN assessment. This process is time-consuming, requires special training, and is prone to errors. The number of type 2 diabetes sufferers globally is growing at alarming rates with healthcare workers under enormous pressure to continue to provide one-to-one regular care. In order to reduce the burden on existing services whilst providing the necessary care to patients, automated approaches for PSN detection provide many advantages. Importantly, with respect to an automated SWME method, there will be areas on the plantar surface where the SWM probe should not be applied i.e., areas with lesions or suspect regions. The research presented in this manuscript conducted a comprehensive analysis of different feature sets and classifiers for the task of lesion classification. Three distinct feature sets Local Binary Pattern (LBP), Mel Frequency Cepstral Coefficients (MFCC), and Scale-Invariant Feature Transform (SIFT) were evaluated across various classifiers, including Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Random Forest (RF), Naïve Bayes (NB), and XGBoost. The results revealed nuanced performances across the combinations of feature sets and classifiers. While each feature set demonstrated strengths, the NB classifier applied to the LBP feature set emerged as the most notable performer with an accuracy score of 100%. This combination achieved perfect accuracy, precision, recall, and F1-score metrics, showcasing its robustness in accurately classifying lesion instances. The 5-fold cross-validation results underscored the stability of NB on the LBP feature set, with a negligible standard deviation, affirming its consistent performance across different data subsets. Additionally, the computational time complexity of 0.91 seconds highlighted its efficiency, making NB on the LBP feature set a practical and reliable choice for real-world applications. Statistical analysis using the one-way ANOVA test revealed significant differences in classifier performance across feature sets, with MFCC resulting in significantly lower accuracy compared to LBP and SIFT, which showed similar performance. The Tukey HSD post-hoc test confirmed these findings, highlighting the crucial role of feature set selection in classifier effectiveness.

**INDEX TERMS** Plantar sensory neuropathy, diabetes, Semmes–Weinstein monofilament, image processing, plantar surface, lesion, LBP, MFCC, genetic algorithm, SVM, MLP.

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

## I. INTRODUCTION

Diabetes Mellitus is a persistent medical condition that presently affects over 11.3% of the developed world, and its

prevalence is on the rise in developing nations like China and India [1]. Diabetic peripheral neuropathy affects nearly 50% of adults with diabetes and can lead to foot ulcers and amputations, requiring aggressive screening and management [2]. In Europe, the number of registered individuals with diabetes is approximately 59.8 million, with a majority of over 80% presenting Type 2 diabetes [1]. Approximately 50% of the aforementioned population is comprised of individuals aged 60 years and above [1]. According to sources, the effective handling of diabetes and its associated complications necessitated nearly 10% of the complete budget allocated to the National Health Service (NHS) in the United Kingdom (UK) [3]. As per the 2017 report published by the International Diabetes Federation (IDF), the global population of adults afflicted with diabetes amounts to 463 million individuals [1]. According to a reliable source [1], it is projected that the number of individuals impacted by this phenomenon will reach 700 million by the year 2045, representing a 48% surge. According to recent data, the total amount of health expenditure in England during the 2018-19 fiscal year amounted to £129 billion, while Scotland and Wales reported health expenditures of approximately £13 billion and £7 billion, respectively [4]. The expenses related to medication or general practitioner and community care are not incorporated. The study did not take into account indirect consequences, such as sick leave and the requirement for unpaid family care [4].

The charity provided additional approximations regarding the expenses linked to managing diabetes. The estimated cost of diabetes in the UK, as reported by Diabetes UK, is £10 billion annually. However, it should be noted that this figure has not been revised in recent years. According to Diabetes.co.uk, a prominent information and support group, an estimated sum of £14 billion was calculated for England and Wales in 2012. That comprises an assumption of £1 billion for yearly pharmaceutical expenses and a projection for expenses related to social care. As of now, there has been no computation conducted regarding the escalation rate of the yearly National Health Service expenditures [4]. Peripheral sensory neuropathy (PSN) is a prevalent complication of diabetes mellitus. It has been observed that approximately 60% of individuals diagnosed with Type 2 diabetes will develop this condition within a decade of initial diagnosis. The ailment typically impacts the distal appendages of the anatomy, frequently commencing with paraesthesia of the digits and subsequently plantar surface of the lower limbs. The somatosensory expression of peripheral sensory neuropathy (PSN) is widely recognized as a significant contributor to the pathogenesis of foot ulcers. If not subjected to vigilant monitoring and appropriate medical intervention, lower limb amputation is frequently recommended [5]. Individuals afflicted with this condition experience a decline in their physical mobility, leading to a significant impact on their overall well-being and quality of life. The maintenance of healthy feet is a crucial aspect for individuals with type 2 diabetes, as it facilitates an active and healthy lifestyle and mitigates the risk of falls, particularly among the elderly population.

This research, recognizing the imperative to address diabetes-related complications, particularly those affecting the feet, delves into the realm of automated diagnostics using machine learning and image processing. The motivation arises from the potential to revolutionize podiatric diagnostics, presenting an innovative approach to decode the subtleties hidden beneath the soles of individuals with diabetes. As we navigate through this interdisciplinary landscape, a captivating narrative unfolds – an algorithmic symphony deciphering the ballet of pixels, distinguishing between normalcy and abnormalities. It's not merely a technical pursuit; it's a transformative quest where pixels metamorphose into diagnosticians, marking a leap towards a healthcare future intricately connected with the very foundations we walk upon. This research makes several significant contributions to the field which are:

- One of the key contributions of this research lies in the ethical sourcing of lesion images from the internet. The collaboration with a podiatrist clinician ensured that all selected plantar lesion images were validated by a medical professional, adding credibility and reliability to the dataset. This meticulous process ensures that the research aligns with ethical standards and guarantees the accuracy and authenticity of the lesion images used in the study.
- The study systematically evaluated three distinct feature sets—Local Binary Pattern (LBP), Mel Frequency Cepstral Coefficients (MFCC), and Scale-Invariant Feature Transform (SIFT)—in the context of lesion classification. This analysis provides insights into the effectiveness of different feature representations for capturing relevant information.
- The research comprehensively compared the performance of five classifiers—Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Random Forest (RF), Naïve Bayes (NB), and XGBoost—across multiple feature sets. This comparative analysis enhances the understanding of how different classifiers interact with specific feature sets and their suitability for lesion classification.
- The research considered not only classification performance but also computational time complexity, addressing the practical viability of the proposed models in real-world applications. This dual-focus approach contributes to the development of efficient and deployable systems for lesion classification, where both accuracy and computational efficiency are crucial considerations.

This paper is further divided into several subsections: Section II contains the literature review on the problem statement, Section III presents the methodology for the proposed approach. Section IV provides information on data models, and Section V describes the experimental setup. The results and discussion on experiments are presented in

Section VI, and finally, Section VII presents the conclusion about the study.

## II. LITERATURE REVIEW

According to research findings, the presence of neuropathy, which is indicative of an elevated risk of ulceration, is suggested by the inability to perceive a 10g force (equivalent to 98mN) applied to key weight-bearing points. At present, the Semmes-Weinstein Monofilament (SWM), an extruded homopolymer, is utilized by a skilled clinician to administer pressure at selected points on the patient's foot for testing purposes. The monofilament has been engineered to exhibit a deflection of 10mm upon the application of a force of 10g (98mN). At this juncture, the patient is expected to acknowledge their ability or inability to perceive the presence of the probe in the relevant area. The procedure is iterated for a predetermined set of locations on each foot that have been mutually established. The SWME manual technique, despite its widespread usage, is regarded as being arduous, demanding in terms of labor, challenging to maintain consistency, and susceptible to experimenter bias [6]. The authors have previously documented their development of an automated device designed to replicate the SWME method. This information can be found in their published work [7]. According to reference [8], the system has the potential to eliminate the requirement of a skilled professional's physical presence, thereby enabling individuals to conduct frequent self-assessments.

In order for an automated SWME testing technique to gain widespread adoption, it is imperative to implement measures that enable the system to automatically identify and aggregate regions of interest, such as lesions, blisters, and open wounds, where automated probing is contraindicated. This paper centers on the extraction of embedded software and the utilization of machine learning techniques to prevent the occurrence of automated probe applications on pressure points that may coincide with a lesion. The detection of plantar surface lesions through automated means poses a significant challenge due to the inherent inconsistencies in the appearance of such lesions. The process of extracting an object from an image involves utilizing the visual features of the image to differentiate the object based on various characteristics such as color, shape, facial features, and spatial arrangement. The morphology of wounds can exhibit significant diversity, characterized by indistinct contrast between the lesion and the adjacent skin, irregular or blurred margins, varied pigmentation within the lesion, and the presence of artifacts such as skin lines, hairs, black borders, and blood vessels [9], [10]. The presence of slough and coagulated blood in and around the lesion, along with the potential impact of certain dressing materials, contribute additional intricacies to the wound's presentation, potentially altering its overall coloration [9].

Numerous algorithms have been created and implemented to address the task of computerized recognition. However, none have been specifically designed to combat the widely utilized SWME. The algorithm recognition techniques presented are limited to lesion-specific orientation and may not be applicable in facilitating the identification and recognition of diverse plantar lesion types or detecting their different stages. The categorization of these techniques is typically based on three main approaches, namely thresholding, edge-based detection, and region-based methods, as documented in reference [10]. The efficacy of thresholding is positively correlated with the quality of contrast between the skin and lesion but is negatively impacted when the two entities overlap, as per sources [10], [11]. Edge-based detection exhibits suboptimal performance in scenarios where boundaries are not clearly defined, while region-based detection encounters challenges when the lesion or skin region exhibits texture or contains multiple colors, resulting in excessive segmentation [11]. Several algorithms that have been introduced rely on color space segmentation and histogram equalization, utilizing specialized equipment like Nevoscope, and are designed to address particular lesions, such as melanoma [12]. In a study conducted by [12], the Nevoscope transillumination light microscopy technique was utilized to capture the vascular architecture of skin lesions and its performance was compared to that of the epiluminescence light microscopy imaging method in order to assess its efficacy for skin-lesion characterization. The present observation highlights that the malignant lesions derived from transillumination exhibit an elevated proportion of transillumination image to epiluminescence image of the corresponding lesion, which can be attributed to augmented blood circulation. The dataset utilized in reference [12] encompasses a collective sum of 40 images, which encompasses 13 benign lesions, 18 dysplastic lesions, and 9 malignant melanomas. There is a lack of information pertaining to the precision of the methodology. The proposed algorithm for identifying the foot region at risk of developing ulcers is based on the observation that the temperature of the foot ulcer region is lower than that of the healthy foot region due to impaired blood flow. This algorithm is referred to as the ''snakes'' algorithm. Thermography is utilized to leverage the variance in temperature. The utilization of a fuzzy c-means algorithm has been employed to discern the most thermally active areas of the plantar surface. The study referenced in [13] utilized a dataset consisting of 59 images and achieved a classification accuracy of 91%. Moreover, the significant computational expenses and the necessity for real-time operations render it unsuitable for deployment on processing boards with limited resources, such as the Raspberry Pi (RPi). Researchers in [14] have used edge-based segmentation, here again, detection is confined to a specific lesion, in the case of [14] malignant melanoma of dermoscopy images.

The study by [15] introduces a robust deep learning model for Diabetic Foot Ulcer (DFU) detection, leveraging a meticulously curated dataset of 1775 images. Employing Faster R-CNN with InceptionV2 and two-tier transfer learning in a five-fold cross-validation, the model achieves an impressive

mean average precision of 91.8%. Notably, it demonstrates real-time applicability with a swift 48 milli seconds inference speed per image and a compact 57.2 MB model size. The study [16] presents a cost-effective ''Foot Selfie'' system designed for smartphone use, empowering patients to independently capture and transmit images of their plantar feet to a remote server. In a pilot study involving limb-salvage clinic patients, 15 participants (10 male) successfully used the system with minimal 5-minute training. They uploaded images on a median of 76% of eligible days over six months. The system effectively captured high-quality images of the entire plantar surface, enabling remote clinical decisions. Monitoring 12 active wounds and 39 pre-ulcerative lesions revealed significant improvements, including the healing of seven wounds and the reversal of 20 pre-ulcerative lesions. DFUC2020 presented a dataset of 2,000 training and 2,000 testing images, serving as a benchmark for evaluating various deep learning algorithms. This paper by [17] compares winning teams' proposals, including Faster R-CNN (and its variants), YOLOv3, YOLOv5, EfficientDet, and a new Cascade Attention Network. Each algorithm's architecture, training parameters, and additional stages (pre-processing, data augmentation, and post-processing) are detailed. All methods incorporated data augmentation to augment the training image pool and post-processing to eliminate false positives. Notably, Deformable Convolution, a Faster R-CNN variant, outperformed others with a mean average precision (mAP) of 0.6940 and an F1-Score of 0.7434.

The study [18] aims to develop an efficient algorithm utilizing infrared thermal images for the early detection of diabetic foot issues through asymmetry analysis. The method involves segmenting left and right foot regions using a region-growing technique. Normal plantar thermograms exhibit symmetric temperature distributions, while diabetic foot cases show asymmetry between ipsilateral and contralateral foot regions. Texture and temperature features are extracted from 11 regions of interest, and asymmetry analysis is conducted on these features. Support Vector Machine (SVM) is employed for classifying regions into normal and ulcer categories. The proposed algorithm achieves impressive results with a maximum accuracy of 95.61%, sensitivity of 96.5%, and specificity of 92.41%. The article [19] presents the application of advanced deep learning techniques to aid in DFU treatment, specifically focusing on ulcer detection from patient foot photos. The study proposes enhancements to the original Faster R-CNN, incorporating data augmentation techniques and adjusting parameter settings. The training dataset comprises 2000 annotated DFU images, validated through the Monte Carlo cross-validation method. The proposed model yields notable results, achieving a mean average precision of 91.4% and an impressive F1 score of 94.8%. Additionally, the average detection speed of 332ms surpasses the performance of conventional detector implementations.

The present study concerns the automated acquisition of images and the maintenance of a uniform background on a perforated sheet utilized as a platform for the placement of a patient's foot. It is noteworthy that the perforations on the sheet bear resemblance to lesions. Furthermore, the present study suggests that the images acquired in this manuscript exhibit suboptimal sharpness, which can be attributed to the presence of a perforation sheet measuring 1 cm in thickness that was interposed between the foot and scanner. The operational control of the scanner is governed by a Raspberry Pi (RPi) which is subject to limitations with respect to its processing capacity, memory allocation, and power consumption. The present manuscript is limited in its focus on the normal and abnormal skin of two distinct classes. The skin that is considered abnormal exhibits a diverse array of irregularities. This encompasses lacerations, contusions, and any observable irregularities that may resemble a pathological tissue alteration. Moreover, the proposed system incorporates automatic image acquisition and extraction of dimensional patches from specific areas. The manuscript's system additionally examined the identification of the normal skin ratio in conjunction with the lesion.

## III. METHODOLOGY

A combination of machine learning and textural recognition techniques can be used to efficiently discriminate normal and abnormal skin based on appearance using image processing, importantly creating an automated non-invasive approach. The scenario presented here greatly differentiates from those used in other published work as the intended decision algorithm controlling the automated probe must be intelligent enough to avoid abnormal plantar skin patterns i.e. Plantar lesions or discolorations on the plantar surface under test. The main objective is to discriminate the location of normal and abnormal skin on the plantar surface. The focus of this manuscript is to develop a model for discriminating normal and abnormal skin on the plantar surface using various machine learning techniques. The following models with their respective features have been exploited:

1) Support Vector Machine (SVM) with statistical features of RGB, HSV color space, and Local Binary Patterns (LBP) histogram.
2) Multilayer Perceptron (MLP) optimized using Genetic Algorithm.
3) SVM with Scale-Invariant Feature Transform (SIFT) features.
4) MLP with Mel Frequency Cepstral Coefficients (MFCC).

The comparison of the evaluation of the aforementioned models is given in the result section. The technical setup shown in Figures 1a and 1b, has been described in previous publications [8] and is an in-house system used to validate the developed lesion recognition and detection method proposed in this paper. A simple, ION COPYCAT handheld document scanner is modified to integrate with the system that produces RGB images. The main concern of this paper is to automate the accepted SWME and ensure that the probe doesn't

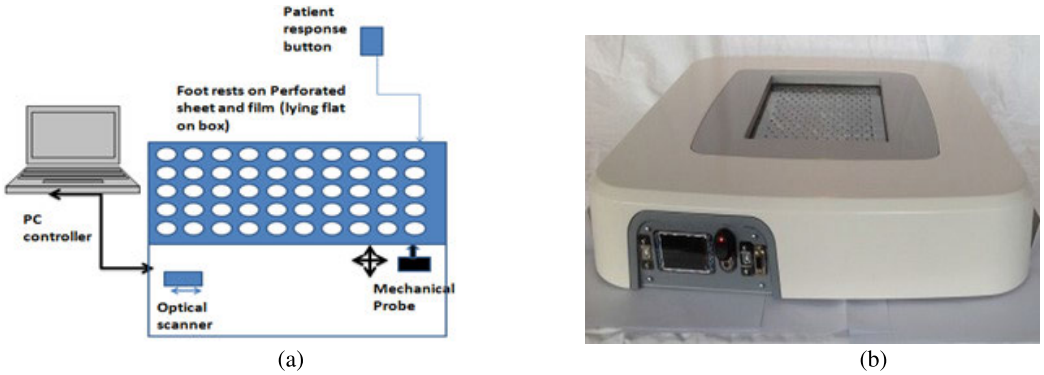



**FIGURE 1.** (a) Schematic of in-house automated SWME replicator set-up [7]. (b) Imaging and probe system for pattern recognition validation.

apply to a lesion. The scope is based on human visual lesion perception detection and then avoiding probing the area. The setup is also used to examine various force applications for automated probes aimed at replicating the SWME. To inform the reader and explain the placement of the pattern recognition methods described in this work, some of the pivotal steps carried out within the device are explained here. The user rests their foot on the perforated sheet as shown in Figure 1b. A photographic image of the plantar surface is then taken. When a photographic image of a plantar surface is obtained, the embedded image processing algorithm extracts five pressure points to be tested by an embedded mechanically driven probe (10mg/98mN of force) [8].

In practice, the trained podiatrist visually avoids probing where lesions overlap with the chosen pressure points and selects adjacent healthy plantar surface points as well as noting the lesions. The embedded algorithms and mechanical probe must apply the same rules as the practitioner but in an automated, unsupervised manner. If the extracted plantar surface pressure points and a lesion overlap, that area must be avoided; its location marked with photographic evidence for health practitioner information, and an adjacent non-overlapped suitable pressure test point must be selected. In this paper, an automated method is described that extracts the image of the pressure point area and categorizes it into either healthy skin that can be probed, or a lesion to be avoided while subsequently deciding on the next best pressure test point nearest the lesion to be selected for the probe.

### A. PATTERNS AND FEATURES GENERATION

Texture contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. As in our case the normal skin patch contains the perforation of the sheet in centre. This is a particular format as can be seen in 2b, this textural pattern disturbs when there exists a wound. Each foot scanned produces a plantar surface image. Each image is subsequently broken down and categorized into one of four patch segments. Their description is provided below along with their appearance shown in Figure 2.

- Blank Patch: This is the patch of the input image that resides outside of the plantar surface as shown in Figure 2-left.
- Foot Edge Patch: This is the patch that lies at the edge of the plantar surface in the image as shown in Figure 2-right.
- Foot Pressure Area Patch: This is important patch as pressure point lies here. This is the patch of the input image that lies inside a detected pressure area zone of the plantar surface, shown in Figure 2b-left. Since the discrimination process is a post pressure point selection process, so the Classification of normal and abnormal skin will always consider this patch.
- Foot Non-Pressure Area Patch: This is the patch that lies on the plantar surface non-pressure area e.g. middle arch and shown in the Figure 2b-right.

The choice of four patch categories above covers the entire foot area, i.e. these patches cover all eventualities ensuring optimized lesion detection and probe application to healthy skin. As the system extracts pressure point and those points lie on pressure areas of plantar surface, so only those patches will be examined that belong to pressure areas, an example is shown in Figure 2b-left. If a pressure point is found to have a lesion, the areas where the next ‘‘best’’ hole lies will be in one of these categories, so they must all be included for each test.

In this study, texture features are used to train machine learning approach and the following features are used to extract the textures of the images:

#### 1) LOCAL BINARY PATTERN

LBP is one of the most popular simple straight, robust and powerful mathematical approaches to extract features based on texture [20]. LBP can achieve effective description ability with appearance invariance and adaptability of patch matching based methods. LBP only thresholds the differential values between neighbourhood pixels. It is computationally

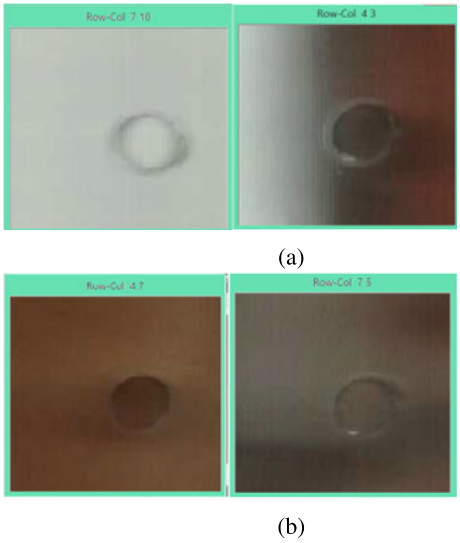



**FIGURE 2.** Patch patterns through perforated sheet. (a-left) Blank hole (no foot) (a-right) foot edge (b-left) Plantar pressure area (d-right) Plantar non-pressure area.

efficient and invariant to monotonic grey-scale changes [21]. LBP is used with combination of SVM to classify into normal and abnormal plantar surface patch. LBP extracts a pattern by thresholding neighbouring pixels in a 3×3 block around the central pixel. It then obtains an LBP code for the central pixel [22], [23], [24]. As HSV has good capability of representing the colours of human perception and simplicity of computation. HSV space hue channel is used in LBP (P, R) histogram. A histogram is produced showing the distribution of the pattern and this histogram is subsequently used one of the features in SVM model. The scale and rotation invariant LBP (P, R) histogram of the dataset was calculated. P is the total number of neighbours of the central pixel in the circle, R is the radius of the circle. Two different LBP scales (8, 1) and (16, 2) were used to ensure greatest accuracy for different size lesions. The resulting pattern information is transformed into an LBP code histogram. The statistical data (mean ± SD) of the constituent colour (channel) intensities of HSV and RGB colour spaces of data set were first obtained.

### 2) SCALE-INVARIANT FEATURE TRANSFORM (SIFT)

Scale Invariant Feature Transform (SIFT) is one of the most widely used feature detection methods. It is an algorithm to perform feature extraction using key point in an image (local features). It is invariance to illumination, scale, and rotation.

### 3) IMAGE TO SOUND CONVERSION

The computational complexity of 1D array (audio) is significantly lower than 2D array (images). The 2D images were converted into 1D acoustic data. As an example, an 8 × 8, 5 grey-tone images are shown in Figure 3. The mapping translates, for each pixel, vertical position into frequency, horizontal position into time, and brightness into sinusoidal signal amplitude in the audible frequency range (500Hz-5000Hz). For a given column, every pixel in this

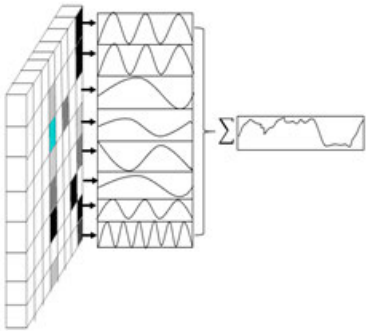



**FIGURE 3.** Image to sound conversion.

column is used to excite an associated sinusoidal oscillator in the audible frequency range. In this manuscript, The RGB image patch of size 181×181×3 is converted and resized into 2048×2048, 64 grey-tone image patch. Frequency range 500Hz-5000Hz on Mel scale is used.

### 4) MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCC)

Mel frequency cepstral coefficients (MFCC) often suggested for identifying monosyllabic words in continuously spoken sentences. MFCC computation is a replication of the human hearing system intending to artificially implement the ear's working principle with the assumption that the human ear is a reliable speaker recognizer [25]. The Mel frequency scale has linear frequency spacing below 1000 Hz and logarithmic spacing above 1000 Hz. MFCC is based on signal disintegration with the help of a filter bank. The MFCC gives a discrete cosine transform (DCT) of a real logarithm of the short-term energy displayed on the Mel frequency scale [26]. The formula used to calculate the mels for any frequency is:

$$mel(f) = 2595 \ * \ log10(1 + \frac{f}{700}) \tag{1}$$

where 2595 is a scaling factor used to convert Hertz to Mel scale in a way that aligns with human auditory perception, mel(f) is the frequency (mels) and f is the frequency (Hz). MFCC feature extraction consists of 6-steps.

1) Audio signal breaks up into a frame size of 2048 with hop size 512 with hamming window.
2) Discrete Fourier Transform calculation: As frequencies in a signal change over time, so instead of taking Discrete Fourier transform (DFT) of entire signal, DFT is carried out at frame level. Extracting the Short time Fourier Transform (STFT) of the frame with 512 number of FFT points i.e. NFTT = 512.
3) Periodogram: Computing the power spectrum to estimate the spectral density of a signal.
4) Filter Banks: The Mel spaced Filter Bank is formally a set of 20×40 triangular filters. To calculate filter bank energies, each filter bank is multiplied with the power spectrum, then add up the coefficients. This ends up with 40 numbers indicating energy in each filter bank.
5) Log filter bank energies: Taking log of spectrogram values to get the log filter bank energies.

6) Discrete cosine transform (DCT): Filter bank coefficient obtained are highly correlated and need to be correlate. DCT is applied.

MFCC with 30 coefficient features were extracted from the acoustic data of plantar patch images to be used in MLP and CNN for classification of normal and abnormal plantar skin patch. The whole feature extraction process has been implemented in Python with the support of the librosa library [27].

#### 5) ARTIFICIAL NEURAL NETWORKS OPTIMIZATION USING GENETIC ALGORITHM

GAs was first proposed by McCall [28] as a tool to find best solutions to problems that were otherwise computationally intractable. GAs is a heuristic optimisation technique that mimics the Darwinian principle of evolution through (genetic) selection and uses a highly abstract version of evolutionary processes to evolve solutions to given problems [29]. GA is used to solve optimization problems by replicating the Darwinian evolutionary behaviour of species. Starting from initial random population of solutions, it evolves by set of activities that includes selection, mutation and crossover operators, inspired in natural evolution. Following aforementioned activities in an order, the population goes through an iterative process transforms to different states, each state is called generation. As a result of this process, the population is expected to reach a generation in which it contains a good solution to the problem [30]. Starting with a randomly generated chromosomes population, a GA performs a process of fitness-based selection and recombination to produce a successor population, the next generation [31]. During recombination, parent chromosomes are selected, and their genetic material is recombined to produce child chromosomes. These are then passed into the successor population. Over successive generations, the population ''evolves'' toward an optimal solution [31].

The image patch shown in Figure 2 is fed to the model, HSV histogram based on hue is calculated for each patch. The histogram is stored in 2D matrix of size $5880 \times 360$, where 5880 are number of image patches and 360 is the bin number. GANN structure parameters are listed below:

- Initial population of neural networks = 8
- Number of neurons in input layers = 5880
- Number hidden layers = 2
- Number of neurons in first hidden layers = 150
- Number of neurons in second hidden layers = 50
- Hidden layer activation function = Rectified Linear Unit (ReLU)
- Number of neurons in output layers = 2
- Output activation function = softmax

The comparison of the feature extractions techniques used in the manuscript are presented in Table 1.

**TABLE 1. Comparison of feature extraction techniques.**

| Technique | Key Features | Advantages |
|---|---|---|
| Local Binary Pattern (LBP) | Thresholds differential values between neighboring pixels. | Invariant to monotonic gray-scale changes, Uses HSV space hue channel for histogram generation, Histogram of LBP codes used in SVM. Computationally efficient, Invariant to scale and rotation when combined with appropriate histograms, Effective description ability with appearance invariance, Adaptability for patch matching methods |
| Scale-Invariant Feature Transform (SIFT) | Detects key points in images (local features), Scale and rotation invariant feature extraction | Invariant to illumination, scale, and rotation, Robust for various image transformations |
| Image to Sound Conversion | Converts 2D images to 1D acoustic data, Maps pixel positions to frequency, time, and amplitude | Reduced computational complexity compared to image data, Facilitates feature extraction using audio signal processing techniques |
| Mel Frequency Cepstral Coefficients (MFCC) | Uses filter banks and Discrete Cosine Transform (DCT), Converts audio signal into 30 coefficient features | Mimics human auditory perception, enhancing feature extraction from audio data, Effective for identifying patterns in audio data derived from image patches |
| Artificial Neural Networks Optimization using Genetic Algorithm (GA) | Optimizes neural network parameters through evolutionary processes, Iterative process with selection, mutation, and crossover, applied to neural networks with specified parameters for input, hidden, and output layers | Finds optimal solutions to complex problems by mimicking natural selection, Enhances the performance of neural networks by optimizing structure and weights, Adaptive learning with improved classification accuracy |

## IV. DATA MODEL

The lesion images used in this study were sourced from the internet, ensuring that no identifiable features such as faces were included, thereby protecting the privacy and anonymity of the subjects. Consequently, ethical approval was not required for their use. Ninety-two plantar lesion images were divided into twenty-three sets, with each set containing four lesions. These sets were then superimposed on foot images of seventy participants, which included both left and right feet, in random order and at random locations on the plantar surface. To accommodate all 92 lesions, twenty-three copies of each foot image were used. The distribution of these lesions is illustrated in Figure 4.The lesion images were categorized by severity into three groups: mild, moderate, and severe. Mild lesions constituted 40% of the images, totaling 37 images. Moderate lesions made up 35% of the images, amounting to 32 images. Severe lesions accounted for 25% of the images, with 23 images in this category.

Each image underwent a rigorous validation process by a certified podiatrist to ensure the accuracy and relevance of the depicted lesions. The criteria for selecting these images included clarity, resolution, and distinct visibility of

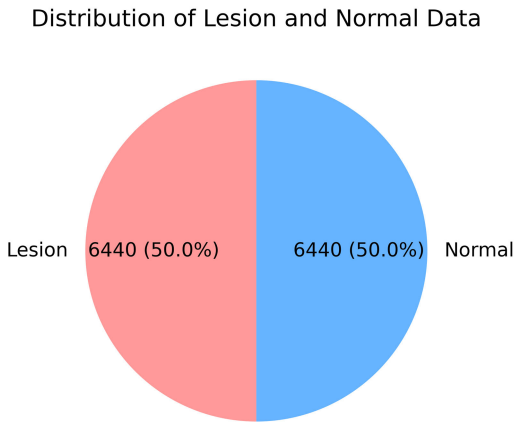



**FIGURE 4.** Distribution of the dataset in lesion and normal class.

**FIGURE 5.** Scanned plantar surface image example through perforated sheet.

the lesions to facilitate precise analysis and diagnosis. The podiatrist assessed each image for clinical validity, ensuring that the lesions met specific diagnostic criteria and could be reliably used in this study.

- Number of feet = 70
- Number of lesioned patches = 70* 23*4 = 6440
- Number of normal pressure area patches = 6440

## V. EXPERIMENT

The data set is divided into two sets training and testing. The models are trained on the training set. Once the models have been trained with the respective features, these models are then applied on the test set to classify into lesion or non-lesion areas based on the training. The surrounding area of the extracted pressure point is examined by the lesion detection code in the manner given below. The full-foot input image an example shown in Figure 5 is divided into equally sized windows such that each window contains a hole on the perforated footrest roughly at its centre.

The holes within the perforated sheet are laid out in 11 columns and 16 rows i.e. in total there are 176 holes. Consequently, the input image is divided into 176 windows. When the algorithm extracts a pressure point, the corresponding window (now called a patch) is sent to the classifier to determine the said patch as either a lesion or a non-lesion patch. If the patch is classified as a non-lesion or in other words 'normal skin', by the classifier, the extracted pressure point is added into the pressure point stream that will be finally sent to the probe application microcontroller. If the patch classifies as a lesion, the corresponding hole nearest that lesion is encircled in red as shown in Figure 6. Figure 6 shows a number of lesions of various colours imposed across a foot surface. In this case the next best neighbouring hole away from the lesion is selected. This classification process is repeated until a healthy pressure skin area is found in the respective pressure region.

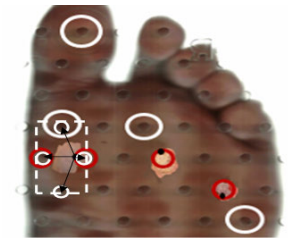


**FIGURE 6.** Best hole selection for lesion avoidance in pressure area.

Let the extracted test point be P(x, y) shown as a black tiny dot in Figure 6 (only dot in Fig). The surrounding "best" and closest holes to the extracted pressure point are determined and shown in Figure 6 with white circles (12, 3, 6, 9pm positions). If the best hole overlaps a lesion, encircled in red (3pm and 9pm position), the next best and closest hole must be chosen, thus the process is repeated to realize whether it overlaps with the lesion or not. The 2-dimensional location of the corresponding best hole is stored in the stream of points to be sent to the microcontroller. The best non-lesion hole is shown as a large white circle. To summarize, the lesion detection process is divided into three main steps: Step 1. Here the incoming scanned patch is examined by the classifier. If the patch is considered normal the coordinates of the relevant hole will be stored for delivery the microcontroller for subsequent pressure execution once steps 1 and 2 for the whole foot is finished. In step 2, if the classifier discriminates the area enclosed in the patch as a lesion, the corresponding sheet hole in the input image is encircled with red and the next best patch closest to the extracted pressure point patch is selected. Steps 1 and 2 are repeated until a normal pressure area patch is found.

## VI. RESULTS AND DISCUSSIONS

This section presents a comprehensive analysis and discussion of the experimental results obtained during the course of this research. The objective is to provide a meticulous examination of the findings, elucidating their significance within the specific context of this study. Furthermore, it involves an extensive discussion exploring the implications and importance of these results, thereby contributing to a deeper understanding of the broader academic and practical consequences arising from the research endeavour.

### A. EXPERIMENTAL SETUP

The HP EliteBook x360 1040 G6 served as the primary computing platform for the experimental analyses. Powered

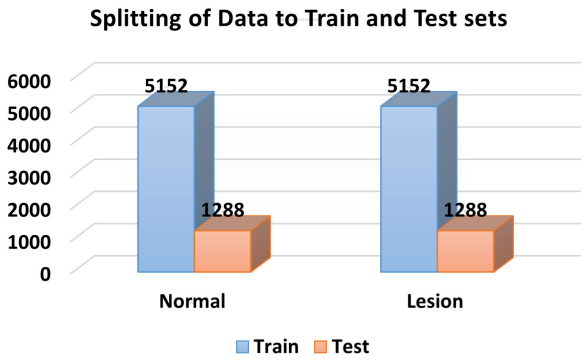



**FIGURE 7.** Distribution of data to train and test sets.

by an Intel(R) CORE (TM) i5-8365U processor running at a base speed of 1.60GHz and capable of reaching a peak speed of 1.90GHz, this system demonstrated impressive computational capabilities. With 16.0 GB of RAM supplementing the CPU, the overall performance was significantly enhanced, particularly in terms of multitasking and data management. Operating on a 64-bit architecture with Windows 11 Pro, the system showcased the seamless integration of cutting-edge hardware and software components. This technical setup underscored the utilization of advanced capabilities throughout the experimentation phase, ensuring a stable and adaptable computing environment.

### B. DATA SPLITTING

The dataset comprises of a total of 12,880 images, evenly divided between the lesion and normal categories, with 6,440 images in each category. To facilitate the training and evaluation of classifiers, the dataset was partitioned into training and testing sets at an 80-20 ratio as shown in Figure 7. Consequently, 80% of the images were allocated to the training set, employed for the classifier training process, while the remaining 20% formed the test set for subsequent evaluation. During the training phase, the classifiers leveraged the diverse features extracted from the images within the training set to learn patterns and features associated with both lesion and normal categories. This process aimed to enhance the classifiers' ability to accurately categorize new, unseen images. The test set, distinct from the training set, served as a critical component for evaluating the classifiers' performance.

### C. HYPERPARAMETERS TUNING

For the classification task in this research, a suite of diverse algorithms, namely SVM, MLP, NB, RF, and XGBoost, has been chosen. The hyperparameters governing the behaviour of these algorithms are systematically selected through a Grid Search approach, as detailed in Table 2. This method involves exhaustively testing a predefined set of hyperparameter values to identify the combination that optimizes the performance of each algorithm. The rationale behind employing multiple algorithms lies in the goal of comprehensively exploring their respective strengths and weaknesses in handling the classification task at hand. Each algorithm brings unique characteristics and capabilities to the table, and their comparative analysis allows for a more nuanced understanding of their efficacy in different scenarios.

**TABLE 2.** Hyperparameters used in this study.

| Model | Hyperparameters |
|---|---|
| SVM | kernel = 'linear', C = 1 |
| MLP | hidden_layer_sizes = (100,), max_iter = 1000, random_state = 42 |
| RF | n_estimators = 100, random_state = 42 |
| NB | Default |
| XGBoost | Default |

**TABLE 3.** Classification results of the classifiers on LBP features.

| Classifier | Accuracy | Precision | Recall | F1-Score | Kfold Accuracy ± Std | CT |
|---|---|---|---|---|---|---|
| SVM | 0.994 | 0.9907 | 0.9906 | 0.9906 | 0.966± 0.0036 | 42.62 |
| MLP | 0.999 | 0.9907 | 0.9906 | 0.9906 | 0.969± 0.0052 | 23.70 |
| RF | 0.958 | 0.9907 | 0.9906 | 0.9906 | 0.957± 0.0029 | 0.057 |
| NB | 1.00 | 0.9907 | 0.9906 | 0.9906 | 0.995± 0.0006 | 0.91 |
| XGBoost | 1.00 | 0.9907 | 0.9906 | 0.9906 | 0.992± 0.0022 | 47.37 |

### D. RESULTS OF CLASSIFIERS ON LBP FEATURES SET

Following the effective training of classifiers on the designated training set, the evaluation phase involves employing the test set featuring LBP descriptors. The outcome of this evaluation is presented in Table 3. Table 3 not only showcases the performance metrics derived from the test set but also includes results obtained from the cross-validation process, which further validates the robustness and consistency of the classifiers. The utilization of 5 folds in the cross-validation ensures a comprehensive assessment, as the dataset is partitioned into five subsets, and the training and testing cycles are iteratively performed. Table 3 also provides valuable information regarding the computational time complexity (CT) of each model. This insight into the time efficiency of the classifiers enhances the understanding of their practical viability for real-world applications, particularly in scenarios where computational resources are a crucial consideration.

From the results presented in Table 3 it is evident that SVM and MLP both achieved exceptional accuracy scores of 0.994 and 0.999, respectively, showcasing robust performance. Moreover, the precision, recall, and F1-score metrics for these classifiers are consistently high, indicating their ability to accurately classify instances across both lesion and normal categories. The 5-fold cross-validation results further support the stability of SVM and MLP, with small standard deviations (0.0036 and 0.0052, respectively) highlighting their consistent performance across different subsets. While Random Forest demonstrates a slightly lower accuracy of 0.958, it maintains competitive precision, recall, and F1-score metrics. The 5-fold cross-validation accuracy with a small standard deviation of 0.0029 underscores its stability. NB and XGBoost both achieve perfect accuracy, showcasing their proficiency in accurately classifying instances. The 5-fold cross-validation results with minimal standard deviations (0.0006 for NB and 0.0022 for XGBoost) further validate their consistency. Considering the overall performance

**TABLE 4.** Results of the classifiers on the MFCC feature set.

| Classifier | Accuracy | Precision | Recall | F1-Score | Kfold Accuracy ± Std | CT |
|---|---|---|---|---|---|---|
| SVM | 0.668 | 0.702 | 0.668 | 0.653 | 0.724 ±0.030 | 5.13 |
| MLP | 0.702 | 0.704 | 0.702 | 0.702 | 0.707±0.003 | 7.56 |
| RF | 0.732 | 0.732 | 0.732 | 0.731 | 0.748±0.007 | 0.06 |
| Naïve Bayes | 0.803 | 0.805 | 0.803 | 0.803 | 0.808±0.008 | 0.77 |
| XGBoost | 0.812 | 0.817 | 0.812 | 0.811 | 0.807±0.005 | 19.10 |

and computational time complexity, the NB model stands out as the best-performing classifier in this context. It attains perfect accuracy while exhibiting minimal computational time complexity (0.91 seconds). This combination of accuracy and efficiency positions NB as a promising model for real-world applications, where both classification precision and computational speed are crucial considerations. The small standard deviation in the 5-fold cross-validation results add to the confidence in the model's reliability.

### E. RESULTS OF CLASSIFIERS ON MFCC FEATURES SET

After successfully training the classifiers on the designated training set, the evaluation phase involves the use of the test set containing MFCC features. The results of this evaluation are presented comprehensively in Table 4. This table not only highlights the performance metrics obtained from the test set but also incorporates results from a 5-fold cross-validation process, further affirming the robustness and consistency of the classifiers. The employment of 5 folds in cross-validation ensures a thorough assessment by iteratively partitioning the dataset into five subsets, facilitating rigorous training and testing cycles. Table 4 additionally furnishes essential insights into the computational time complexity of each model. Understanding the time efficiency of the classifiers becomes pivotal for gauging their practical viability in real-world applications, especially in scenarios where computational resources are a critical consideration. This multifaceted presentation contributes to a comprehensive evaluation, making Table 4 an invaluable reference for researchers and practitioners seeking to assess both the classification performance and efficiency of these models.

It is evident from the Table 4 that the SVM, while achieving an accuracy of 0.668, exhibits balanced precision and recall scores. MLP slightly improves on accuracy, reaching 0.702, with consistent precision and recall. RF displays better performance with an accuracy of 0.732, maintaining competitive precision, recall, and F1-score metrics. NB surpasses the others with an accuracy of 0.803, along with excellent precision, recall, and F1-score values. XGBoost also performs well with an accuracy of 0.812, demonstrating high precision, recall, and F1-score. The 5-fold cross-validation results, reflected in the mean accuracy and standard deviation, provide insights into the classifiers' consistency across different subsets. NB exhibits the highest mean accuracy (0.808) and a relatively low standard deviation (0.008), indicating robust and stable performance. XGBoost also demonstrates consistent results with a mean accuracy of 0.807 and a small standard deviation of 0.005. Considering both classification performance and computational time complexity, NB emerges as the most promising model for this task. With a solid accuracy of 0.803 and an efficient computational time of 0.77 seconds, NB strikes a balance between accuracy and practical feasibility. The minimal standard deviation in cross-validation further supports the reliability of NB in achieving consistent results. NB stands out as the preferred model for this MFCC- feature set.

**TABLE 5.** Classification results on the SIFT features.

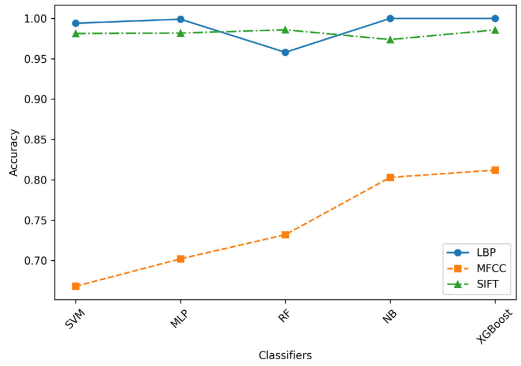
| Classifier | Accuracy | Precision | Recall | F1-Score | Kfold Accuracy ± Std | CT |
|---|---|---|---|---|---|---|
| SVM | 0.981275 | 0.984778 | 0.984396 | 0.984421 | 0.981275± 0.003 | 579.8154 |
| MLP | 0.981832 | 0.985535 | 0.985288 | 0.985306 | 0.981832± 0.003 | 74.31277 |
| RF | 0.985845 | 0.99068 | 0.990638 | 0.990642 | 0.985845± 0.002 | 308.2703 |
| Naïve Bayes | 0.973807 | 0.977508 | 0.976371 | 0.976431 | 0.973807± 0.003 | 7.335622 |
| XGBoost | 0.985733 | 0.991117 | 0.991083 | 0.991088 | 0.985733± 0.002 | 381.2864 |

### F. RESULTS OF CLASSIFIERS ON SIFT FEATURES SET

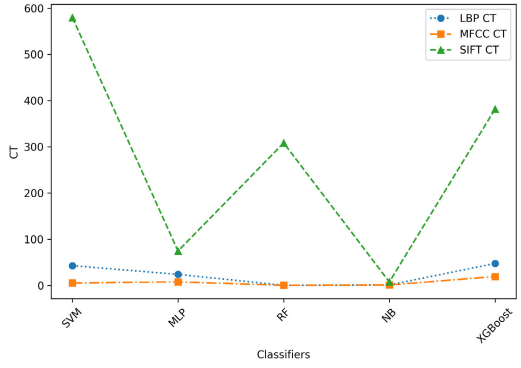
Table 5 provides a comprehensive overview of the classifier performances using the SIFT feature set. In addition to the results achieved on the SIFT features, the table incorporates cross-validation scores, offering a broader perspective on the classifiers' robustness and generalization capabilities. The inclusion of computational complexity metrics further enhances the understanding of the practical implications of each model.

It is evident from Table 5 that SVM and MLP both demonstrate remarkable accuracy, with SVM achieving 0.981275 and MLP slightly higher at 0.981832. These classifiers also maintain impressive precision, recall, and F1-score metrics, showcasing their effectiveness in accurately categorizing images represented by SIFT features. The 5-fold cross-validation results indicate consistent performance, with minimal standard deviations (0.003) affirming the stability of SVM and MLP across different subsets. RF stands out with the highest accuracy of 0.985845, accompanied by competitive precision, recall, and F1-score values. The 5-fold cross-validation results reinforce the reliability of RF, as indicated by a small standard deviation of 0.002. XGBoost closely follows with an accuracy of 0.985733, demonstrating high precision, recall, and F1-score metrics. The 5-fold cross-validation standard deviation of 0.002 further supports the consistency of XGBoost.

NB, while achieving a slightly lower accuracy of 0.973807, maintains commendable precision, recall, and F1-score values. The 5-fold cross-validation results with a standard deviation of 0.003 suggest stable performance across different data subsets. Considering both performance and computational time complexity, RF emerges as the most promising model for this SIFT feature set. With a top-tier accuracy and a moderate computational time complexity of 308.2703 seconds, RF strikes a favourable balance between classification precision and efficiency. The minimal standard deviation in the 5-fold cross-validation results further strengthens the case for RF's reliability and consistency. In the context of SIFT features, RF stands out as the preferred model, offering robust performance and practical suitability for real-world applications.




(a)




(b)

**FIGURE 8.** Machine learning models performance comparison in terms of (a) accuracy (b) CT.

### G. COMPARISON OF RESULTS

Upon a thorough examination of the classification results in various set of characteristics, the NB model applied to the set of characteristics of the LBP emerges as the best performer for the classification of lesion in this study as compared to Figure 8. The LBP feature set, which captures textural information, proved to be suitable for NB, demonstrating an exceptional accuracy of 100% and demonstrating precision, recall, and F1 score metrics consistently above 99%. The model exhibited stability in 5-fold cross-validation, with a negligible standard deviation of 0.0006, affirming its robust and consistent performance across different data subsets. Moreover, NB showcased remarkable computational efficiency with a minimal time complexity of 0.91 seconds, making it a practical choice for real-world applications where both precision and computational speed are crucial considerations. Although other models excelled in specific feature sets, comprehensive assessment designates NB in the LBP feature set as the optimal choice to achieve accurate and efficient lesion classification in this research effort.

### H. STATISTICAL ANALYSIS OF CLASSIFIER PERFORMANCE ACROSS DIFFERENT FEATURE SETS

To validate the performance of classifiers across different feature sets, a one-way ANOVA test was conducted to compare the accuracy scores obtained from the 5-fold cross-validation process as shown in Table 6. This analysis aims to determine whether there are significant differences in the performance of the classifiers when trained on different feature sets (LBP, MFCC, and SIFT). The null hypothesis (H0) states that there is no significant difference in accuracy scores across the different feature sets, while the alternative hypothesis (H1) states that there is a significant difference.

**TABLE 6.** Classifier accuracy across different feature sets.

| Classifier | LBP Accuracy | MFCC Accuracy | SIFT Accuracy |
|---|---|---|---|
| SVM | 0.966 ± 0.0036 | 0.724 ± 0.030 | 0.981275 ± 0.003 |
| MLP | 0.969 ± 0.0052 | 0.707 ± 0.003 | 0.981832 ± 0.003 |
| RF | 0.957 ± 0.0029 | 0.748 ± 0.007 | 0.985845 ± 0.002 |
| Naïve Bayes | 0.995 ± 0.0006 | 0.808 ± 0.008 | 0.973807 ± 0.003 |
| XGBoost | 0.992 ± 0.0022 | 0.807 ± 0.005 | 0.985733 ± 0.002 |

**TABLE 7.** ANOVA test results for classifier accuracy across different feature sets.

| Source of Variation | SS (Sum of Squares) | p-Value |
|---|---|---|
| Between Groups | 0.0483 | 0.0021 |
| Within Groups | 0.0097 | - |
| Total | 0.0580 | - |

**TABLE 8.** Tukey HSD Post-Hoc test results.

| Comparison | Mean Difference | p-Value |
|---|---|---|
| LBP vs. MFCC | 0.205 | 0.002 |
| LBP vs. SIFT | -0.003 | 0.972 |
| MFCC vs. SIFT | -0.208 | 0.002 |

The one-way ANOVA test was conducted on the accuracy scores of the classifiers across the three feature sets and results are presented in Table 7.

From Table 7 it is evident that, the p-value is less than 0.05, indicating significant differences in classifier performance across the different feature sets. This suggests that the choice of feature set significantly impacts the accuracy of the classifiers.

To identify which feature sets contribute to the significant differences, a post-hoc analysis using the Tukey Honestly Significant Difference (HSD) test was performed and results are presented in Table 8.

Table 8 indicates that the mean difference between the LBP and MFCC feature sets, as well as between the MFCC and SIFT feature sets, is significant (p-value $< 0.05$). However, the difference between the LBP and SIFT feature sets is not significant. The results of the one-way ANOVA test and the Tukey HSD post-hoc analysis reveal significant differences in classifier performance across different feature sets. Specifically, the MFCC feature set results in significantly lower accuracy compared to both the LBP and SIFT feature sets. In contrast, the LBP and SIFT feature sets yield similar classifier performance, showing no significant difference in accuracy between them. These findings highlight the crucial role of feature set selection in determining the effectiveness of classifiers. The LBP and SIFT feature sets demonstrate greater effectiveness for the classifiers studied, whereas the MFCC feature set leads to lower accuracy. This information

**TABLE 9.** Comparative analysis of accuracy with previous studies.

| Study | Methodology | Accuracy |
|---|---|---|
| [13] | Thermography, "snakes" algorithm | 91% |
| [15] | Faster R-CNN, InceptionV2 | 91.8% |
| [18] ) | Infrared thermal imaging, SVM | 95.61% |
| [17] | Deformable Convolution R-CNN | 69.4% |
| [19] | Enhanced Faster R-CNN | 91.4% |
| **Proposed Study** | **LBP, Naive Bayes** | **100%** |

is particularly valuable for researchers and practitioners when selecting feature sets for classification tasks, emphasizing the importance of feature selection in achieving optimal classifier performance.

### I. COMPARISON WITH EXISTING STUDIES

The current study demonstrates exceptional accuracy in diabetic foot ulcer detection, achieving a perfect 100%. This marks a significant improvement over previous methodologies. For instance, [13] achieved a 91% accuracy using thermography and the ''snakes'' algorithm, while [15] reported a mean average precision of 91.8% with their Faster R-CNN and InceptionV2 model. Saminathan et al. [18] reached 95.61% accuracy through infrared thermal imaging and SVM analysis. The DFUC2020 [17] benchmark highlighted various deep learning algorithms, with the top performer being a Deformable Convolution variant of Faster R-CNN, achieving a mean average precision of 69.4%. In comparison, [19] enhanced Faster R-CNN for DFU detection, achieving an accuracy of 91.4%.

The presented study's use of the LBP feature set and NB model for classification outperformed all previous studies as shown in Table 9, demonstrating a perfect 100% accuracy, which underscores its superiority in lesion detection. This exceptional performance is indicative of the model's robustness and reliability in real-world applications.

### J. DISCUSSION

The results from this study provide a comprehensive evaluation of the classifiers' performance across different feature sets, highlighting the strengths and limitations of each approach. For the LBP feature set, the Naïve NB classifier emerged as the most effective model, achieving perfect accuracy of 100% along with precision, recall, and F1-score metrics consistently above 99%. This exceptional performance indicates that NB can effectively leverage the textural information captured by the LBP features, leading to highly accurate lesion classification. Additionally, the stability of the NB model was evident from the 5-fold cross-validation, which showed a negligible standard deviation of 0.0006, confirming its robust and consistent performance across different data subsets. The results on the MFCC feature set revealed a different landscape. While the NB classifier again showed strong performance with an accuracy of 0.803, it was the XGBoost classifier that slightly outperformed it with an accuracy of 0.812. Both classifiers demonstrated high precision, recall, and F1-score values, indicating their effectiveness in handling MFCC features. However, the NB model's efficiency and stability, as evidenced by its computational time of 0.77 seconds and a standard deviation of 0.008 in cross-validation, still made it a competitive choice.

For the SIFT feature set, the RF classifier performs best, achieving the highest accuracy of 0.985845. This performance was closely followed by XGBoost and MLP, which also exhibited high accuracy and consistent cross-validation results. Despite NB achieved a slightly lower accuracy of 0.973807, it maintained commendable precision, recall, and F1-score values. The RF model's favorable balance between classification precision and moderate computational time complexity of 308.2703 seconds, coupled with a minimal standard deviation in cross-validation, made it the most promising model for the SIFT feature set. The computational efficiency of the NB classifier, particularly on the LBP feature set, was noteworthy. The simplicity of the NB algorithm, which leverages the Bayes theorem with strong independence assumptions between features, results in lower computational overhead. This efficiency is further enhanced by the LBP feature set, which provides straightforward statistical relationships that the NB model can rapidly process. As a result, the NB classifier exhibited minimal computational time complexity of 0.91 seconds, making it highly suitable for real-world applications where both precision and speed are crucial. Despite these promising results, the study has several limitations that should be acknowledged. The dataset used was evenly divided between lesion and normal categories, which cannot fully represent the variability found in real-world scenarios. Future research will aim to include more diverse datasets with a wider range of lesion types and imaging conditions to enhance the generalizability of the findings. Additionally, while three distinct feature sets were analyzed, there can be other feature extraction techniques or combinations that could yield better performance. Exploring deep learning-based representations could be a valuable direction for future studies. Real-time implementation and testing of these classifiers in operational environments would require further optimization to ensure consistent performance under varying conditions. By addressing these limitations and exploring these future directions, researchers can refine the classifiers and enhance their applicability to real-world lesion classification tasks.

## VII. CONCLUSION

An image processing technique designed to help in the evolution of fully automated SWME methods has been presented. Algorithms have been developed to avoid the application of the probe when the lesion and the pressure point overlap. The input plantar surface image is segmented into 176 windows, known as patches once declared as pressure points. Each patch is then sent to a classifier trained for lesion and non-lesion feature spaces. In the case of lesion detection, the next best patch is selected, and this process is repeated until a normal patch pattern representing normal plantar pressure skin area is found close to the original

patch. This research conducted a comprehensive analysis of different feature sets and classifiers for the task of lesion classification. Three distinct feature sets, LBP, MFCC, and SIFT, were evaluated on various classifiers, including SVM, MLP, RF, NB, and XGBoost. The results revealed nuanced performances across combinations of feature sets and classifiers. While each feature set demonstrated strengths, the NB classifier applied to the LBP feature set emerged as the most notable performer. This combination achieved perfect accuracy, precision, recall, and F1 score metrics, showcasing its robustness in accurately classifying lesion instances. The 5-fold cross-validation results underscored the stability of NB on the LBP feature set, with a negligible standard deviation, affirming its consistent performance across different data subsets. Additionally, the computational time complexity of 0.91 seconds highlighted its efficiency, making NB on the LBP feature set a practical and reliable choice for real-world applications. Statistical analysis using the one-way ANOVA test further validated these findings, revealing significant differences in classifier performance across different feature sets. The results indicated that the MFCC feature set resulted in significantly lower accuracy compared to both the LBP and SIFT feature sets. In contrast, the LBP and SIFT feature sets yielded similar classifier performance, showing no significant difference in accuracy between them. The Tukey HSD post-hoc test confirmed these results, emphasizing that the choice of feature set plays a crucial role in determining the effectiveness of classifiers. The LBP and SIFT feature sets demonstrated greater effectiveness for the classifiers studied, whereas the MFCC feature set led to lower accuracy. These findings highlight the importance of feature selection in achieving optimal classifier performance. The combination of the NB classifier with the LBP feature set stands out as the most effective approach for lesion classification in this study, providing both high accuracy and computational efficiency. This information is valuable for researchers and practitioners when selecting feature sets and classifiers for similar tasks, guiding them towards making informed decisions to enhance classification performance.

## ACKNOWLEDGMENT

The authors would like to thank Michelle Spruce, Honorary Research Fellow at London South Bank University and the Managing Director of Kinetic Healthcare Ltd., for her support as a podiatrist during the research.

## REFERENCES

[1] I. D. Federation. (2019). *IDF Diabetes Atlas 9th Edition 2019*. Accessed: Jan. 17, 2024. [Online]. Available: https://www.diabetesatlas.org/en/

[2] C. W. Hicks and E. Selvin, "Epidemiology of peripheral neuropathy and lower extremity disease in diabetes," *Current Diabetes Rep.*, vol. 19, no. 10, pp. 1–8, Oct. 2019.

[3] U. Diabetes, *NHS Spending on Diabetes 'to Reach£16.9 Billion by 2035*. Stockport, U.K.: Wired Gov Ltd, 2021. [Online]. Available: https://www.wired-gov.net/wg/wg-news-1.nsf/0/8E12C1A0724858F9802579EB004894E8?OpenDocument

[4] B. News. *How Much Does Diabetes Cost the NHS*. Accessed: Jan. 17, 2021. [Online]. Available: https://www.bbc.com/news/health-49758070

[5] S. R. Shah and K. M. Patil, "Processing of foot pressure images and display of an advanced clinical parameter PR in diabetic neuropathy," in *Proc. 9th Int. Conf. Rehabil. Robot. (ICORR)*, Sep. 2005, pp. 56–59.

[6] G. Bove, "Mechanical sensory threshold testing using nylon monofilaments: The pain field's 'Tin standard,'" *Pain*, vol. 124, no. 1, pp. 13–17, 2006.

[7] H.-Ur-R. Siddiqui, S. R. Alty, M. Spruce, and S. E. Dudley, "Automated peripheral neuropathy assessment of diabetic patients using optical imaging and binary processing techniques," in *Proc. IEEE Point Care Healthcare Technol. (PHT)*, Jan. 2013, pp. 200–203.

[8] 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.

[9] S. M. Seyyed Ebrahimi, H. Pourghassem, and M. Ashourian, "Lesion detection in dermoscopy images using sarsa reinforcement algorithm," in *Proc. 17th Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2010, pp. 1–4.

[10] R. Hu, C. M. Queen, and G. Zouridakis, "Lesion border detection in buruli ulcer images," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 5380–5383.

[11] P. Liang, Y. Cong, and M. Guan, "A computer-aided lesion diagnose method based on gastroscopeimage," in *Proc. IEEE Int. Conf. Inf. Autom.*, Jun. 2012, pp. 871–875.

[12] A. P. Dhawan, B. D'Alessandro, S. Patwardhan, and N. Mullani, "Multispectral optical imaging of skin-lesions for detection of malignant melanomas," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 5352–5355.

[13] M. Etehadtavakol, E. Y. K. Ng, and N. Kaabouch, "Automatic segmentation of thermal images of diabetic-at-risk feet using the snakes algorithm," *Infr. Phys. Technol.*, vol. 86, pp. 66–76, Nov. 2017.

[14] A. A. A. Al-Abayechi, R. Logeswaran, X. Guo, and W.-H. Tan, "Lesion border detection in dermoscopy images using bilateral filter," in *Proc. IEEE Int. Conf. Signal Image Process. Appl.*, Oct. 2013, pp. 365–368.

[15] 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.

[16] S. Nagaraju, K. V. Kumar, B. P. Rani, E. L. Lydia, M. K. Ishak, I. Filali, F. K. Karim, and S. M. Mostafa, "Automated diabetic foot ulcer detection and classification using deep learning," *IEEE Access*, vol. 11, pp. 127578–127588, 2023.

[17] M. H. Yap, "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104596.

[18] J. Saminathan, M. Sasikala, V. Narayanamurthy, K. Rajesh, and R. Arvind, "Computer aided detection of diabetic foot ulcer using asymmetry analysis of texture and temperature features," *Infr. Phys. Technol.*, vol. 105, Mar. 2020, Art. no. 103219.

[19] A. Oliveira, A. Britto de Carvalho, and D. Dantas, "Faster R-CNN approach for diabetic foot ulcer detection," in *Proc. 16th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2021, pp. 677–684.

[20] S. Balammal@Geetha, R. Muthukkumar, and V. Seenivasagam, "Enhancing scalability of image retrieval using visual fusion of feature descriptors," *Intell. Autom. Soft Comput.*, vol. 31, no. 3, pp. 1737–1752, 2022.

[21] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

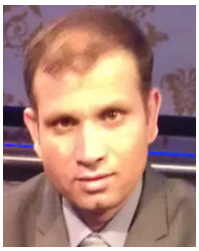
[22] Z. Li, G. Liu, Y. Yang, and J. You, "Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2130–2140, Apr. 2012.

[23] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*, vol. 40. Cham, Switzerland: Springer, 2011.

[24] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation-invariant image and video description with local binary pattern features," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1465–1477, Apr. 2012.

[25] S. E. Agustina and I. Mukhlash, "Implementasi metode scale invariant feature transform (SIFT) dan metode continuously adaptive mean-shift (Camshift) pada penjejakan objek bergerak," *Jurnal Sains dan Seni*, vol. 1, no. 1, pp. 1–6, 2012.

[26] S. Chakroborty, A. Roy, and G. Saha, "Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification," in *Proc. IEEE Int. Conf. Ind. Technol.*, Sep. 2006, pp. 387–390.
[27] K. Ravikumar, B. Reddy, R. Rajagopal, and H. Nagaraj, "Automatic detection of syllable repetition in read speech for objective assessment of stuttered disfluencies," *Proc. World Acad. Sci., Eng. Technol.*, vol. 36, pp. 270–273, Oct. 2008.
[28] J. McCall, "Genetic algorithms for modelling and optimisation," *J. Comput. Appl. Math.*, vol. 184, no. 1, pp. 205–222, Dec. 2005.
[29] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, nos. 14–15, pp. 2627–2636, Aug. 1998.
[30] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network—A deep learning approach," *Proc. Comput. Sci.*, vol. 132, pp. 679–688, Jan. 2018.
[31] M. Valenti, S. Squartini, A. Diment, G. Parascandolo, and T. Virtanen, "A convolutional neural network approach for acoustic scene classification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1547–1554.



**HAFEEZ UR REHMAN SIDDIQUI** (Student Member, IEEE) received the B.Sc. degree in mathematics from Islamia University Bahawalpur (IUB) and the M.Sc. and Ph.D. degrees in electronic engineering from London South Bank University, in 2012 and 2016, respectively. His research interests include biomedical and energy engineering applications, data recognition, image processing, and system-embedded programming IoT-based smart system incorporation with machine learning. He is a Reviewer of IEEE INTERNET OF THINGS JOURNAL.

**RICCARDO RUSSO** is currently with the Department of Brain and Behavioral Sciences, University of Pavia, and the Department of Psychology, University of Essex. His research interests include cognitive psychology, experimental psychology, and brain neuromodulation.

**ADIL ALI SALEEM** received the B.S. degree in computer science from the University of Lahore, Lahore, Pakistan, in 2016, and the M.S. degree in computer science from the Khwaja Fareed University of Engineering and Information and Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2021, where he is currently pursuing the Ph.D. degree with the Institute of Computer Science. His research interests include the IoT-based smart systems embedded with machine learning, text mining, and biomedical engineering.

**SANDRA DUDLEY** (Member, IEEE) was born in Carlow, Ireland. She received the B.Sc. (Hons.) and Ph.D. degrees in physics from the University of Essex, Essex, U.K., in 1998 and 2004, respectively. She spent two years working as a Postdoctoral Researcher with Essex University on a British Telecom Project resulting in world-record broadband-system power consumption for last-mile access. In August 2009, she joined the School of Engineering, London South Bank University, London, U.K., as a Lecturer, and became a Principal Lecturer, in March 2014. Her research interests include physical layer system design, automated health technologies, data recognition, optical-wireless systems, and solutions for the broadband urban-rural divide. She is a reviewer of IET and IEEE journals.

**FURQAN RUSTAM** received the M.C.S. degree from the Department of Computer Science, The Islamia University of Bahawalpur, Pakistan, in October 2017, and the Master of Computer Science degree from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He is currently pursuing the Ph.D. degree in computer science with University College Dublin, Ireland. He was a Research Assistant with the Fareed Computing and Research Center, KFUEIT. His research interests include data mining, machine learning, and artificial intelligence, mainly involved in creative computing and supervised machine learning.