

Original Article

Enhanced Diabetes Detection from Foot Plantar Thermographs Using an Attention-Infused InceptionV3 Residual Block

Nisanth Krishnan¹, V. Balamurugan²

^{1,2}Department of ECE, Sathyabama Institute of Science and Technology, Tamilnadu, India.

¹Corresponding Author : Nisanth.Krishnan11@outlook.com

Received: 23 September 2024

Revised: 22 October 2024

Accepted: 20 November 2024

Published: 03 December 2024

Abstract - A chronic illness, Diabetes Mellitus (DM), occurs due to the inability of the pancreas to produce insulin or to utilize the insulin it produces effectively. People with diabetes have increased risks of developing various life-threatening conditions, resulting in reduced quality of life and increased mortality. Diabetes causes long-term impairment and degradation of many body parts. Early intervention and treatment of diabetes can prevent extreme outcomes such as amputation. Thermography is a non-invasive technique commonly used to detect variations in temperature distribution in the foot region. So, in this study, a hybrid Deep Learning (DL) model incorporating a pretrained inception V3 with custom layers of attention and residual blocks is proposed to detect diabetes from foot plantar thermographic images efficiently. Thermography images of the foot from the foot plantar dataset are utilized in this study with various preprocessing and data augmentation techniques. The proposed model exhibits superior performance when compared to state-of-the-art methods with 95.71% accuracy, 97.85% precision, 93.83% recall, and 95.80 % F1score. In addition to standard evaluation metrics, the performance of the hybrid DL models is measured with Cohen's kappa and the Area under the Curve (AUC). The outcomes indicate the model's potential in real-time clinical application, resulting in more effective diabetic detection and management.

Keywords - Diabetes, Thermography, Deep Learning, InceptionV3, Attention mechanism, Residual block.

1. Introduction

Diabetes is a serious threat to global health that transcends neither socioeconomic rank nor nationwide boundaries [1]. Worldwide, diabetes is one of the top ten reasons for death in adults and was assessed to have attributed to 6.2 million deaths due to diabetes and associated complications in 2023. DM is a chronic disease characterized by hyperglycemia. American Diabetes Association states that prolonged hyperglycemia accompanies long-term impairment, dysfunction, and deterioration of different body parts, specifically the eyes, kidneys, blood vessels, motor nerves, heart, and foot [2]. DM is a persistent illness that develops due to the raised levels of glucose in the bloodstream since the human body cannot generate any or very little amount of the insulin hormone or utilize insulin efficiently. Insulin deficit, if left unchecked over a period of time, could harm many of the body's organs, resulting in disabling and life-threatening health problems. The major forms of diabetes are type 1, type 2 and gestational diabetes mellitus [3].

In type 1 diabetes, the body produces very little or no insulin due to the autoimmune response of the body that strikes the insulin-secreting beta (β) cells of the pancreas. Hence, people with type 1 diabetes must administer insulin

injections daily to maintain a glucose level in the appropriate range [4].

Globally, type 2 diabetes is the most prevalent form of diabetes, accounting for around 90% of overall diabetes. Primarily, hyperglycaemia is an outcome of the incapability of the body's cells to react entirely to insulin, a condition called 'insulin resistance'. Throughout this situation, the hormone is ineffective and eventually causes an increase in insulin secretion. When unrecognized for an extended period of time, problems such as diabetic retinopathy or Diabetic Foot Ulcers (DFU) that remain unhealed may be present at analysis [5]. Gestational Diabetes Mellitus (GDM) or Diabetes in Pregnancy (DIP) is detected first during pregnancy and habitually vanishes after giving birth. GDM can arise anytime during pregnancy [6].

Diabetic foot is a devastating complication, and it comprises lesions in the deeper-lying tissues related to neurological illnesses and Peripheral Arterial Disease (PAD) in the lower limbs in patients with diabetes. The loss of nerve function, Diabetic Neuropathy (DN) and loss of circulation cause serious foot problems like DFU, gangrene, charcot arthropathy and amputation [7]. Figure 1 illustrates the healthy and diabetic foot.



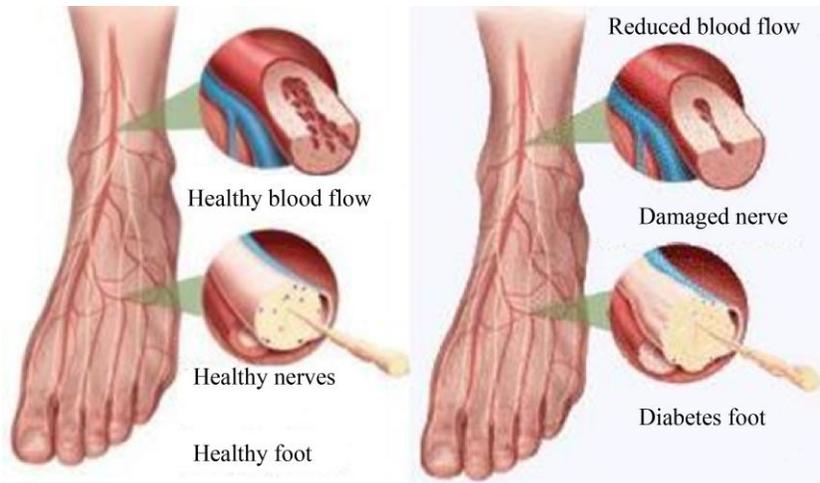


Fig. 1 Illustration of healthy and diabetes foot

Using foot plantar thermography, diabetes can be detected through the temperature disparities in the plantar surface. Thermography is the most common non-invasive method providing cost-effective insights into foot-related issues and circulation problems. By integrating artificial intelligence, abnormal thermal images indicating diabetic complications can be identified, leading to early interventions and reducing the risk of amputation [8]. So, in this study, a hybrid model is suggested for effectively detecting diabetes using foot plantar thermograph images. The major contributions of the study are given below:

- To develop a hybrid model for the efficient detection of diabetes using foot plantar thermograph images.
- To integrate the inception V3 model with custom attention and residual block layers.
- To compare the effectiveness of the suggested model with existing models.
- To evaluate the efficiency through evaluation metrics like accuracy, precision, recall, and F1-score.

The remaining portion of the paper is structured as follows: Section 2 provides a comprehensive literature review emphasizing the need for the current research. Section 3 details the methodology with a hybrid model architecture for effectively detecting diabetes. Section 4 presents the results and discussion, highlighting the potential of the suggested model. Section 5 concludes the paper by summarizing the key contributions.

2. Related Works

Sulayman et al. [9] developed a Convolutional Neural Network (CNN) to detect and classify the healthy and diabetic from foot plantar thermal images. The authors incorporated the Transfer Learning (TL) model to compensate for the small dataset problem and achieved enhanced training speed. Different transfer learning models, such as ResNet-50, VGG16, and EfficientNetB3, were used with Adam and the SGD optimizer for comparison. The model provided an

accuracy of 90.5 % when VGG16 was coupled with the Adam optimizer. The study faced challenges due to the cost of thermal cameras and a lack of programming skills to create a website based on the algorithm. Verma [10] proposed a DL model for detecting DFU using thermal images. Thermal images from open source were utilized for the study, and these images were preprocessed to generate a new dataset by employing watershed segmentation and canny edge detection. ResNet50 and EfficientNetB0 models were employed on both datasets, and EfficientNetB0 achieved the highest accuracy, surpassing ResNet50. The study's main limitation is the model's applicability in low-resource healthcare settings with limited processing power.

Khosa et al. [11] suggested a custom CNN model along with ResNet50 and DenseNet121. Various machine learning models are evaluated using a publicly available thermographic dataset to detect DFU from plantar temperature changes. Among the machine learning models, SVM outperformed others. The proposed custom CNN model achieved greater efficiency with an accuracy of 93% when combining both image and patch-level data. However, comparing the suggested model with conventional models regarding thermogram data was challenging. Torres et al. [12] focused on developing a technological system for detecting diabetes from foot issues by utilizing tools that measure temperature variation, electrical impedance and macule identification for healthy and diabetic patients. The outcomes highlighted that the diabetes patients indicate a temperature difference of 2°C and impedance in the range of 5 kHz to 22 kHz. Temperature measurement under an uncontrolled environment using smartphones included limitations regarding the standardization of the equipment and consistent sensor calibration.

Arteaga-Marrero et al. [13] employed an extended and more generalizable dataset by incorporating a publicly available INAOE dataset and a recently released STANDUP.

Features were extracted using Random Forest (RF), lasso and two variational deep-learning techniques. SVM was used to classify healthy and diabetic subjects. The optimized SVM with the lasso method excelled with the most relevant feature associated with the Medial Calcaneal Artery (MCA) and Lateral Plantar Artery (LPA). However, the study was limited by the consequences of avoiding uniform preprocessing. Cao et al. [14] proposed a Deep Neural Network (DNN), the Plantar Foot Segmentation Network (PFSNet), for extracting the foot contours effectively. The model combined a feature extraction and a CNN module with an attention mechanism incorporating a feature fusion module. The PFSNet outperformed state-of-the-art methods for single-channel thermal images with an accuracy of 95.4% as measured by the Dice Similarity Coefficient (DSC). The study possessed challenges due to the complications associated with the cold immersion process in thermography.

Yogapriya et al. [15] suggested a Diabetic Foot Infection Network (DFINET), a CNN model with 22 layers. The model incorporated parallel convolution filters to extract important features. The DFINET demonstrated a higher accuracy of 91.98% in classifying the healthy and DFU images. The study noted challenges with nonstandard images, dataset imbalance and varying skin textures. Cruz-Vega et al. [16] classified different thermal patterns of patients with DM. The study compared different machine learning models with deep learning structures and introduced the Diabetic Foot Thermograms Network (DFTNet) to detect diabetic foot effectively. The model outperformed GoogleNet with an accuracy of 94.53%.

Khandakar et al. [17] designed a labelled dataset utilizing k-mean clustering to cluster the severity of the diabetic foot. Feature engineering was done using VGG19 with image enhancement techniques like Adaptive Histogram Equalization (AHE), which showed an accuracy of 95.08%. A stacking classifier is proposed combining Gradient Boost (GB), XGBoost, and RF classifiers with an accuracy of 94.47%. The feasibility of image enhancement techniques in real-

world applications limited the study. Anaya-Isaza and Zequera-Diaz [18] developed the Thermal Change Index (TCI), a classification index to increase accuracy. The study introduced ResNet50v2 deep CNN, which was improved by applying twelve data augmentation techniques, including dimensionality reduction methods. The model provided an accuracy of 82.39% with TL from the thermographic database. However, the new coefficient stratification, which corresponded to the severity of the pathology, was not verified.

A notable research gap acknowledged from the above studies is the lack of a comprehensive framework that integrates image enhancement techniques, data augmentation, and multi-modal data (such as thermal images, electrical impedance, and temperature variation) for the severity classification of DFUs. While several studies have explored different aspects, such as applying transfer learning, using various deep learning models, or introducing new indices like the TCI, none have systematically combined these approaches to enhance the robustness and generalizability of the models across diverse clinical settings and heterogeneous datasets. Furthermore, there is a need for validation of the proposed models and indices in real-world clinical environments, particularly in low-resource settings where processing power and access to standardized equipment may be limited.

3. Materials and Methods

Efficient detection of diabetes from foot plantar thermograph images is crucial for early diagnosis and prevention of severe complications like diabetic foot ulcers and amputations. So, in this study, a hybrid model incorporating a modified inceptionV3 model enhanced with attention and residual block is proposed. Initially, the thermographic image data is preprocessed and data augmented to input into the hybrid deep learning model. The model is trained to efficiently and accurately classify diabetes and healthy cases through a sigmoid output layer. The schematic block of the suggested model is illustrated in Figure 2.

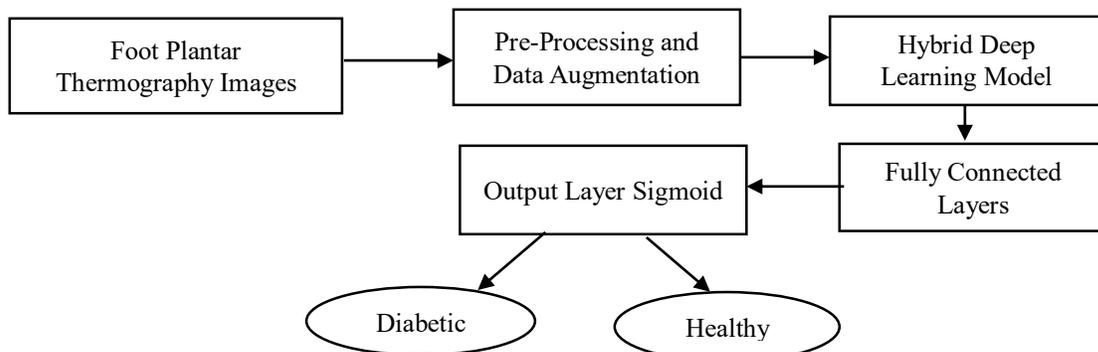


Fig. 2 Schematic block of the suggested model

3.1. Dataset

Foot plantar thermography, a non-invasive thermographic technique, is used for this study since it is crucial for the early detection of diabetes and related complications. Foot plantar thermographic images are selected for this study because of their capability to detect abnormalities in tissue health and blood flow by capturing the temperature distribution across the foot. In the case of diabetic patients, the temperature distribution across the plantar region changes due to the impairment in the tissues and nerves, and thermography can easily detect it even before the symptoms occur. The images have been collected from the Kaggle Repository [19] and are stored in the Foot Plantar Dataset for the early intervention of diabetes. The dataset comprises 1866 thermal images, where 890 images of ‘Healthy’ stands for healthy controls, labelled as ‘0’ and 976 images of ‘Diabetic’ stands for diabetic foot, labelled as ‘1’. For training and testing, 1586 images and 280 images were used. Sample images from the foot plantar thermograph dataset are illustrated in Figure 3. The distribution of diabetic/ healthy classes is represented by Figure 4, indicating that the class distribution of diabetes and healthy images is relatively balanced.

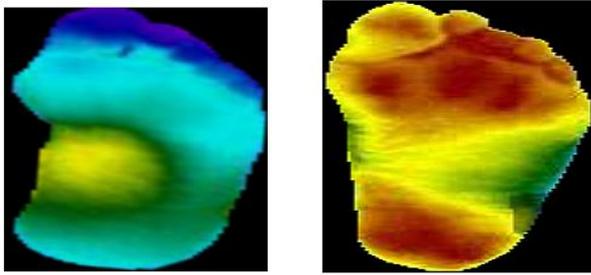


Fig. 3 Sample images from foot plantar thermograph dataset: healthy and diabetic

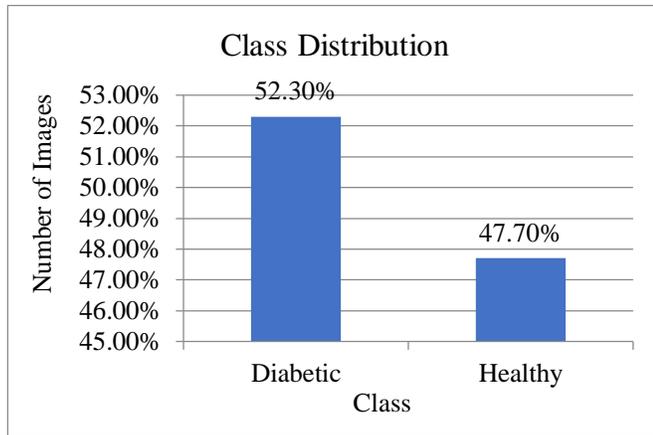


Fig. 4 Class distribution

3.2. Preprocessing and Data Augmentation

The images experienced a series of preprocessing and augmentation steps, following the dataset collection phase, to prepare them for training. Resizing is done to ensure uniformity across the images in the dataset. All the

thermographic images are resized into 224 x 224 pixels. Color channel adjustment is implemented for converting the BGR color format to RGB color format, ensuring the efficient interpretation of the model for accurate feature extraction. The unwanted background information in the foot plantar thermography images introduces noise into the training process.

To avoid this, the Region of Interest (ROI) is recognized and cropped from each image based on the extreme points detected from the contour of the image. Initially, the image is converted to grayscale and is then blurred using a Gaussian filter to reduce noise and smoothen the image. The Gaussian function is expressed by Equation (1).

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Where σ denotes the degree of blurring. Thresholding is performed after blurring to create the binary image as represented by Equation (2).

$$T(x, y) = \begin{cases} 255 & \text{if } I(x, y) > \text{threshold} \\ 0 & \text{if } I(x, y) \leq \text{threshold} \end{cases} \quad (2)$$

Where, $I(x, y)$ is the pixel intensity at (x, y) coordinates. The binary image highlights the ROI; erosion and dilation operations were applied for further refinements. The erosion process removes small white regions with noises, while dilation enlarges the remaining white noise regions, making the ROI more distinct. Now, the contour is selected by joining the points with the same colour or intensity along the boundary. Extreme points were determined by analyzing the contour coordinates, and the ROI was cropped from the original image based on these extreme points. As shown in Figure 5, this cropped image is then resized to ensure uniformity.

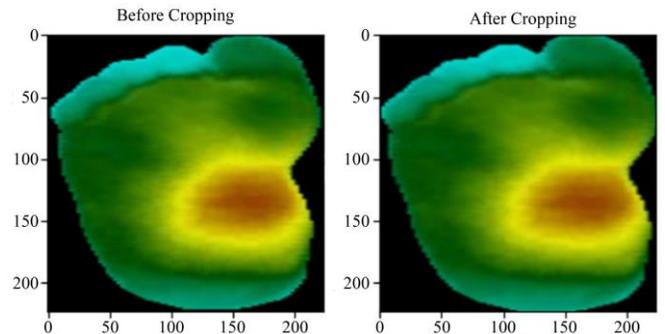


Fig. 5 Foot plantar thermography image before and after cropping

The image pixel values are scaled to a range of [0,1] by the normalization process as given by Equation (3).

$$I_{norm}(x, y) = \frac{I(x, y)}{255} \quad (3)$$

Where $I_{norm}(x, y)$ is the normalized pixel value. Data augmentation is also vital since it enlarges the dataset by applying various transformations.

In this study, transformations such as rotating, shifting, shearing, adjusting brightness, and flipping horizontally and vertically are implemented to get the augmented data for increasing the dataset’s diversity and generalizability and avoiding the risk of overfitting.

3.3. Proposed Hybrid Model

Figure 6 represents the block diagram of the hybrid model architecture for efficiently detecting diabetes from foot plantar thermography images. The input layer consisting of preprocessed and data augmented images of dimension 224 x 224 x 3 is given to the inceptionV3 model, the backbone of the hybrid model.

Then, an attention mechanism attains an attention score for the important regions in the feature map and is provided to the residual block, which helps refine important extracted features. Then, the model is connected to a fully connected layer and a sigmoid output layer to classify healthy/ diabetic cases from the thermography images.

3.3.1. InceptionV3

A CNN architecture, InceptionV3 is recognized for its efficient design and high performance in image classification tasks [20]. The InceptionV3 network uses several inception modules, each comprising multiple convolutional layers with different kernel sizes running in parallel, as shown in Figure 7. This lets the network capture features at multiple scales. The pretrained InceptionV3 model, a transfer learning framework that is trained on ImageNet datasets, is employed and has good performance on a small dataset. An inception module is mathematically represented by Equation (4).

$$OutputConcat(Conv_{1 \times 1}(X), Conv_{3 \times 3}(X), Conv_{5 \times 5}(X), MaxPool_{3 \times 3}(X)) \tag{4}$$

Where *Concat* is the concatenation of feature maps, $Conv_{k \times k}$ denotes a convolution operation with a $k \times k$ kernel and $MaxPool_{3 \times 3}$ is a max-pooling operation with a 3×3 window. The input X is processed in parallel through these operations, concatenating their outputs. To minimize computational complexity, inceptionV3 factorizes larger convolutions into smaller ones. This factorization decreases the computational cost and the number of parameters while maintaining the ability to capture complex patterns.

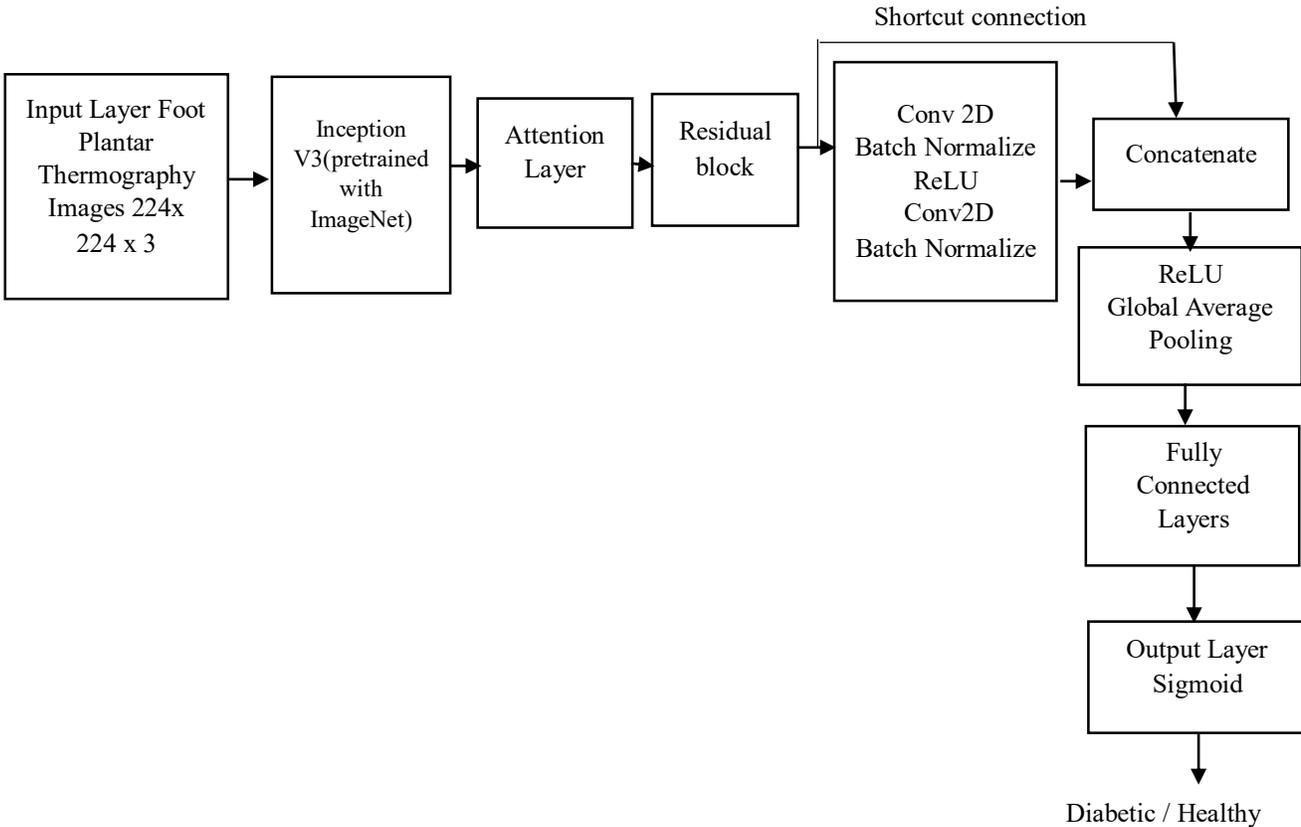


Fig. 6 Block diagram of the proposed hybrid architecture

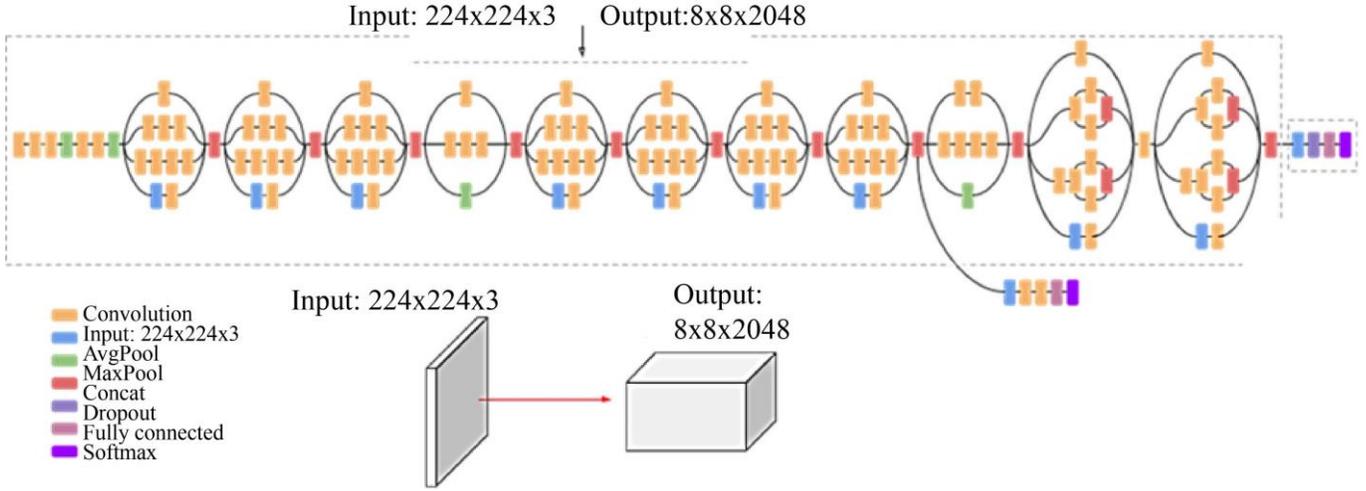


Fig. 7 Basic architecture of inception V3

Batch normalization is used to regulate and accelerate the training process. It normalizes each input layer to have a mean of zero and a variance of one. For an input X with mini-batch mean μ_B and variance σ_B^2 ,

$$\hat{X} = \frac{X - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (5)$$

Where ϵ is a small constant to avoid division by zero, and the learnable parameters are given by Equation (6),

$$Y = \gamma \widehat{X} + \beta \quad (6)$$

Where γ and β are learnable parameters that allow the model to scale and shift the normalized output. InceptionV3 includes auxiliary classifiers to provide additional gradient signals and regularization during training. The mathematical expression of auxiliary classifier loss is given by Equation (7),

$$L = L_{main} + \alpha L_{aux} \quad (7)$$

Where total loss L is a weighted sum of the main loss L_{main} and the auxiliary loss L_{aux} and α is a hyperparameter that determines the contribution of the auxiliary loss. The

optimizer used by inception V3, which adjusts the learning rate for each parameter according to the magnitude of recent gradients, helps to stabilize the training process. For a parameter θ with a gradient g_t at time step t , the optimizer update equation is given by Equation (8).

$$E[g^2]_t = \rho E[g^2]_{t-1} + (1 - \rho)g_t^2$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t \quad (8)$$

Where ρ is the decay rate, η is the learning rate, and ϵ is a small constant.

3.3.2. Convolutional Neural Network

A common class of DNN used to analyze visual imagery is the CNN model, as shown in Figure 8. Unlike conventional methods that rely on matrix multiplication, CNN employs convolution, which involves a mathematical operation on two functions to create a third function that represents how one's shape is altered by the other [21]. The convolved feature is obtained by applying a kernel /filter (3x3 matrix). This feature is given as an input to the next layer. Figure 9 represents the convolution operation.

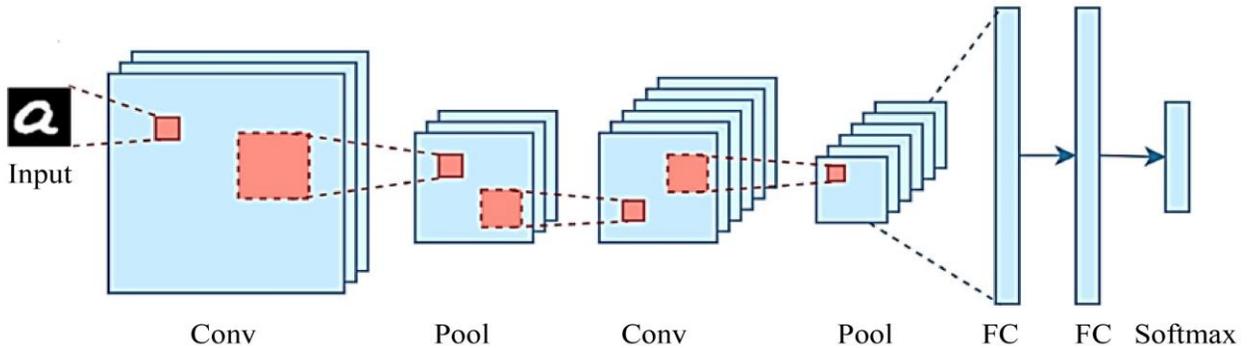


Fig. 8 Basic CNN architecture

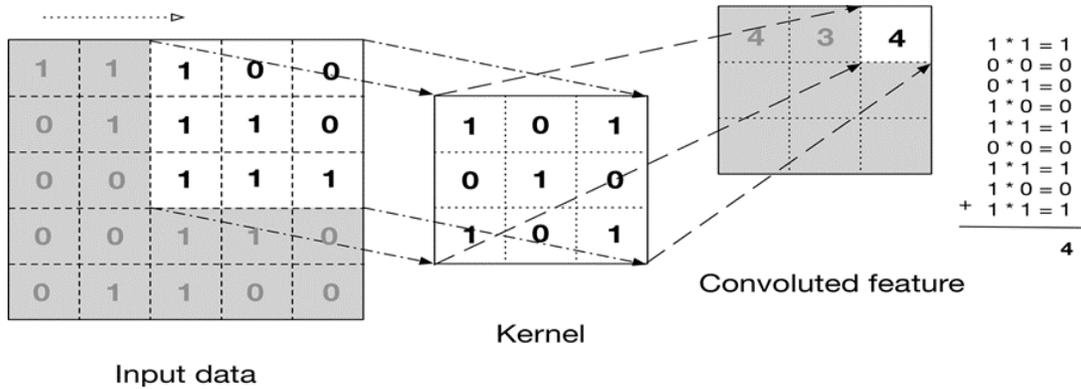


Fig. 9 Basic convolution operation

The average pooling operation is done as a down-sampling method by maintaining important information to reduce the spatial dimensions of the feature maps. The pooling operation is done by applying a filter over the feature map and taking the average of the elements within each patch covered by the filter. This average is used to represent the entire patch. The average pooling is visualized in Figure 10.

Each input node is connected to the preceding output layer to form a fully connected network, as shown in Figure 11. Each neuron in this layer performs a linear transformation on the input vector using a weights matrix, ensuring that each input influences the corresponding output.

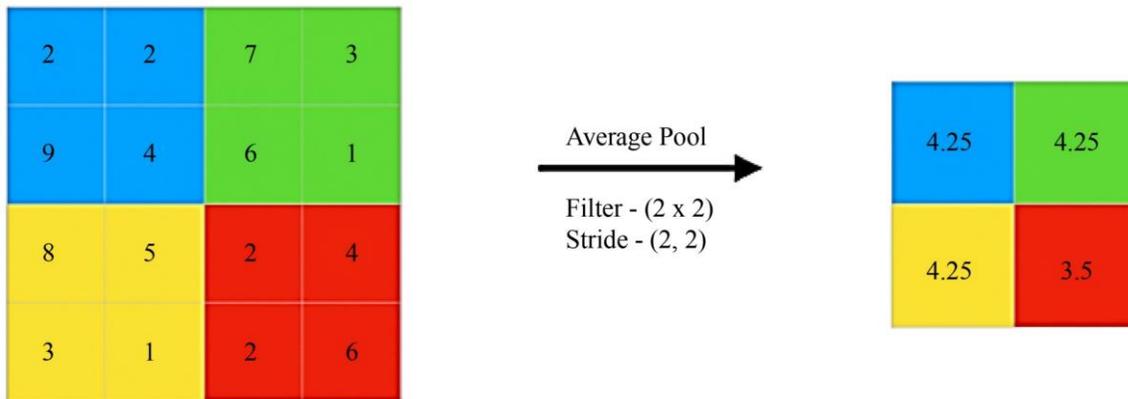


Fig. 10 Visualization of average pooling operation

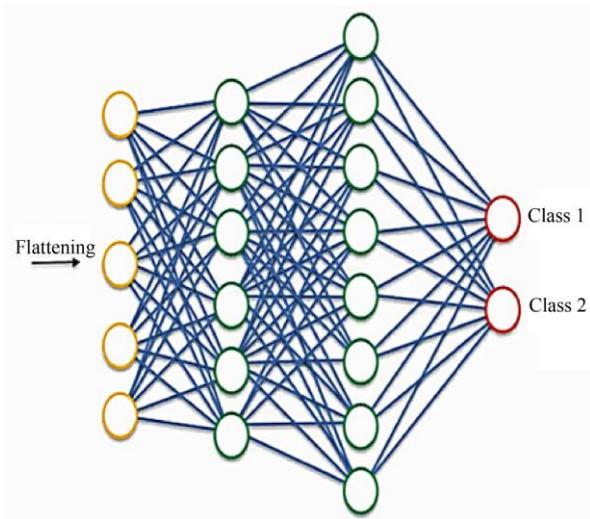


Fig. 11 Fully connected layer

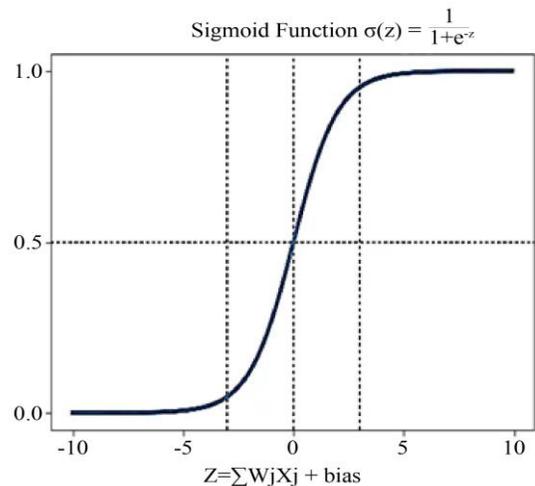


Fig. 12 Activation curve of sigmoid function

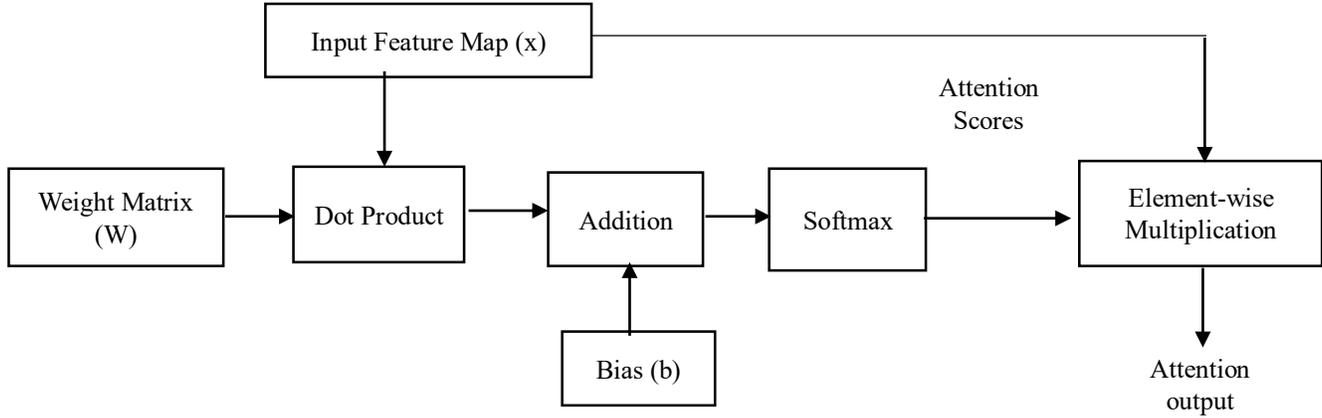


Fig. 13 Process flow of suggested attention mechanism

As shown in Figure 12, the sigmoid function is used as the output layer, producing input values in a range between 0 and 1, making it suitable for classification tasks. At $z = 0$, when the curve crosses 0.5, the activation function's rules are established. The function outputs 1 if the value is equal to or greater than 0.5 and 0 otherwise, providing a probabilistic interpretation of the outputs.

3.4. Attention Mechanism

The attention mechanism allows the model to emphasize the pertinent information of the input for achieving better accuracy in detecting diabetes from foot plantar thermography images [22]. The process flow of the suggested attention mechanism is given in Figure 13.

X is the input feature map of shape (batch size, height, width, channels) from the previous layer. The weight of each feature is calculated using the weight matrix of the shape, W and bias, b . Attention scores are calculated by performing a linear transformation on X . For each position (i, j) in the feature map, the transformation is given by taking the dot product of W and X , then adding the bias function,

$$A_{ij} = X_{ij} \cdot W + b \quad (9)$$

Finally, the attention score is calculated by applying the softmax operation to the transformed function A_{ij} as in Equation (10).

$$S_{ij} = \frac{\exp(A_{ij})}{\sum_{ij} \exp(A_{ij})} \quad (10)$$

Now, the weighted feature map is computed by the element-wise multiplication of the attention score S with the X to obtain the attention output Y , as illustrated by Equation (11), which has the same shape as X .

$$Y_{ij} = S_{ij} \cdot X_{ij} \quad (11)$$

3.5. Residual Block

Residual block addresses the key challenges, such as vanishing gradient and degradation, for effectively training the suggested model [23]. By learning the residuals, the block refines the features extracted for the enhanced detection of healthy/ diabetic feet. The workflow of the proposed residual block is given in Figure 14.

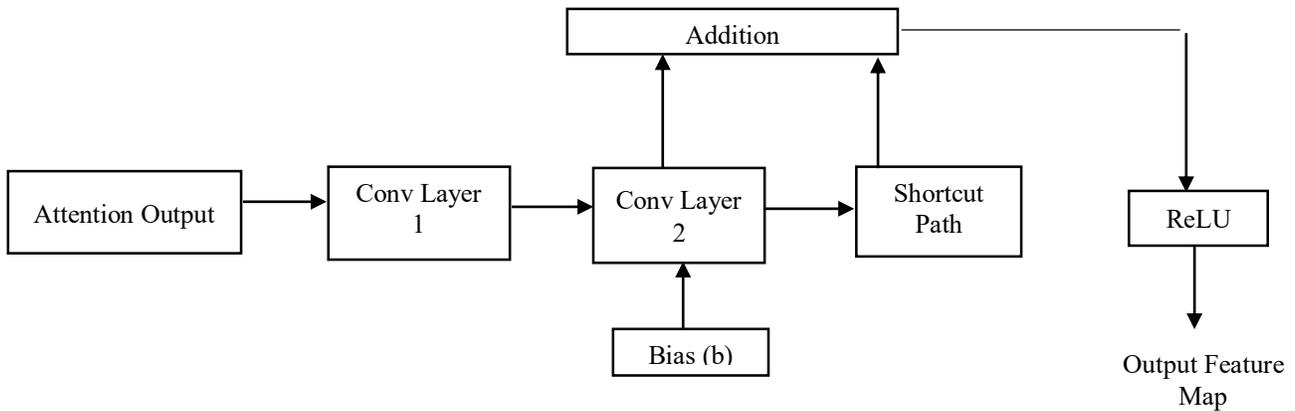


Fig. 14 Process flow of proposed residual block

The input, X , to the residual block is the output from the attention block. The block consists of two convolutional layers.

Initially, the first convolution layer incorporates a 2D convolution operation with a filter size of 3×3 with specified stride and padding to preserve the spatial dimensions. The first convolutional layer is represented as Equation (12).

$$Z_1 = Conv2D(X, W_1) \quad (12)$$

The output is normalized to obtain Equation (13).

$$\widehat{Z}_1 = \frac{Z_1 - \mu_{batch}}{\sqrt{\sigma_{batch}^2 + \epsilon}} \quad (13)$$

This normalized output is then passed to a ReLU activation function as Equation (14).

$$A_1 = ReLU(\widehat{Z}_1) \quad (14)$$

Now, the second convolutional layer with the same filter dimension as the first layer is applied to the ReLU output with bias, as in Equation (15).

$$Z_2 = Conv2D(A_1, W_2) + b_1 \quad (15)$$

The output Z_2 is normalized as,

$$\widehat{Z}_2 = \frac{Z_2 - \mu_{batch}}{\sqrt{\sigma_{batch}^2 + \epsilon}} \quad (16)$$

If the stride is not 1, to match the dimensions of the input, the shortcut path is initialized with a convolutional layer of 1×1 as represented by Equation (17) and the normalization $\widehat{X}_{shortcut}$ is applied if convolution initializes.

$$X_{shortcut} = Conv2D(X, W_s) + b_s \quad (17)$$

The output from the second convolutional layer is added to the shortcut path, where the output will be the original input X or the transformed input if convolution is applied.

$$Y = \widehat{Z}_2 + \widehat{X}_{shortcut} \quad (18)$$

ReLU activation is applied to the summed output and passed into the next layer. Global Average Pooling (GAP) is applied to minimize the dimension of the feature map by reducing the number of parameters. The mathematical representation of GAP is given by Equation (19).

$$GAP(X) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{i,j} \quad (19)$$

The model includes fully connected layers, starting with a dense layer of 1024 units and ReLU activation, followed by a dropout layer to prevent overfitting.

Finally, the output layer is a single neuron with a sigmoid activation function, producing a probability for classifying the given foot plantar thermography images to correspond to a healthy or diabetic foot. Algorithm 1 represents data augmentation, and Algorithm 2 represents the proposed model given below.

Algorithm 1: Data Augmentation for Foot Plantar Thermography Images

Input: Foot plantar thermography images.

Result: Augmented dataset with reliable healthy and diabetic class images

for each image i in N images of healthy class do

$P[i] \leftarrow P[i]$ Rotate (probability = 0.3, max_left_rotation = 10, max_right_rotation = 10)

$P[i] \leftarrow P[i]$ Brightness (probability = 0.3, min_factor = 0.3, max_factor = 1.0)

$P[i] \leftarrow P[i]$ Zoom (probability = 0.4, min_factor = 1.0, max_factor = 1.3)

$P[i] \leftarrow P[i]$ Width_Shift (probability= 0.1)

$P[i] \leftarrow P[i]$ Height_Shift (probability= 0.1)

$P[i] \leftarrow P[i]$ Flip_Horizontal (probability = 0.5)

$P[i] \leftarrow P[i]$ Flip_Vertical (probability = 0.5)

$P[i] \leftarrow P[i]$ Random_Shear (probability = 0.3, max_shear_left = 10, max_shear_right = 10)

end for

3.6. Hardware and Software Setup

The study utilized a comprehensive setup consisting of an NVIDIA GeForce GTX 1080Ti GPU, an Intel Core i7 processor, 32GB of RAM, and the Python-based Keras library integrated with the TensorFlow framework. The user-friendly interface of Keras, integrated with Google Colab's vast computational ability, ensures enhanced model development with GPU acceleration. The data splitting is done in the 75: 15: 10 ratio for training, testing and validation respectively. Table 1 illustrates the hyperparameters used in the training, which is critical for optimizing the model's accuracy and efficiency. These parameters are set before the model execution to govern the training process, ensuring efficient handling of the diabetic foot plantar thermography dataset and facilitating faster convergence.

Algorithm 2: Proposed Hybrid Model

Input: Foot plantar thermography images

Output: Efficient diabetic detection

Begin:

Load, preprocess and augment data:

1. *Collect dataset: $D = \{(X_i, y_i)\}$, where X_i is the foot plantar image in the Foot Plantar Thermo Dataset, and y_i is its corresponding label as '1' for diabetic and '0' for healthy.*
2. *Preprocessing and data augmentation*
 - *Resize: $X_i \rightarrow X'_i \in \mathbb{R}^{224 \times 224}$*
 - *Normalize: $X'_i \rightarrow \frac{X'_i - \mu}{\sigma}$*
 - *Cropping the ROI out of the image*
 - *Data Augmentation: $X'_i \rightarrow \{X''_i\}$ (Shear, Rotation, Brightness, Shift, Zoom, Flipp)*

Split the dataset into training, testing and validation in the ratio of 75:15:10.

Load Models: Hybrid model architecture combining InceptionV3 and custom layers with attention mechanism and residual blocks.

1. InceptionV3:

- *Apply inception modules for multi-scale feature extraction: Concatenate outputs of different convolutional layers:*

$$\text{Output} = \text{Concat}(\text{Conv}_{1 \times 1}(X), \text{Conv}_{3 \times 3}(X), \text{Conv}_{5 \times 5}(X), \text{MaxPool}_{3 \times 3}(X))$$
- *Implement batch normalization and Auxiliary classifiers for regularization.*

2. Attention Mechanism:

- *Calculate attention scores and apply them to the feature maps:*

$$\text{Attention Output, } Y_{ij} = S_{ij} \cdot X_{ij}$$

3. Residual Block:

- *Implement residual connections with batch normalization and ReLU activation:*

$$\text{Residual block output, } Y = \widehat{Z}_2 + \widehat{X}_{\text{shortcut}}$$

4. Global Average Pooling: GlobalAvgPooling2D ()

5. Fully Connected Layers:

- *Dropout (0.5) (x)*
- *Dense (1024, activation='relu') (x)*

6. Output layer

- *predictions = Dense (1, activation='sigmoid') (x)*

Model Compilation and Training:

1. *Compile each model M:*

$$\text{loss} = \text{binary_crossentropy}$$

$$\text{optimizer} = \text{Adam} ()$$

$$\text{metrics} = [\text{accuracy}]$$
2. *Train: $M.\text{fit}(X_{\text{train}}, y_{\text{train}}, \text{validation_data} = (X_{\text{val}}, y_{\text{val}}))$*

Model Evaluation and Comparison:

1. *Evaluate:*

$$\text{metrics} = M.\text{evaluate}(X_{\text{test}}, y_{\text{test}})$$
, where metrics include accuracy, precision, and recall.
Save the Model:

End

Table 1. Hyperparameter specifications

Hyperparameters	Values
Loss Function	Binary Cross Entropy
Epochs	30
Optimizer	Adam
Dropout	0.5
Activation Function	ReLU
Batch Size	64
Learning Rate	0.0001

4. Results and Discussion

4.1. Performance Evaluation

The model’s performance is mainly evaluated on four parameters: accuracy, precision, recall, and F1-score. These measures, which are based on the concepts of False Positive (FP), False Negative (FN), True Negative (TN), and True Positive (TP), are essential for assessing the model’s performance in detecting diabetes from foot plantar thermography images.

Accuracy calculation involves dividing the total number of predictions by the number of right predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{20}$$

The exactness of a prediction is measured by its precision or the number of true positives. Instead, recall quantifies completeness, or the number of real positives that were anticipated as positives.

$$Precision = \frac{TP}{TP+FP} \tag{21}$$

$$Recall = \frac{TP}{TP+FN} \tag{22}$$

$$F1 - Score = 2 * \left(\frac{Precision*Recall}{Precision+Recall} \right) \tag{23}$$

The classification report of the suggested method is given in Table 2, and Figure 15 illustrates its graphical representation.

Table 2. Classification report of the proposed hybrid model

Performance metrics	Values (%)
Accuracy	95.7143
Precision	97.8571
Recall	93.8356
F1 Score	95.8042
Cohen’s kappa	91.4286
Area under curve	95.7984
Specificity	97.7611

The model’s performance metrics indicate a strong and well-balanced classifier, particularly in the context of diabetic and healthy foot plantar thermography image classification. From Table 2, it is clear that the model correctly classified the diabetes cases with an accuracy of 95.71 %. 97.85 % precision indicates effective detection of diabetic cases to avoid unnecessary interventions. 93.83% recall value represents that the model successfully identified the true diabetic cases. The F1 score indicates a balance between precision and recall by avoiding false positives and capturing true positives with a value of 95.80%. The reliability of the model is underscored by the Cohens kappa value of 91.42%. The high specificity value means the model correctly identified 97.76 % of healthy cases by minimizing FP rates.

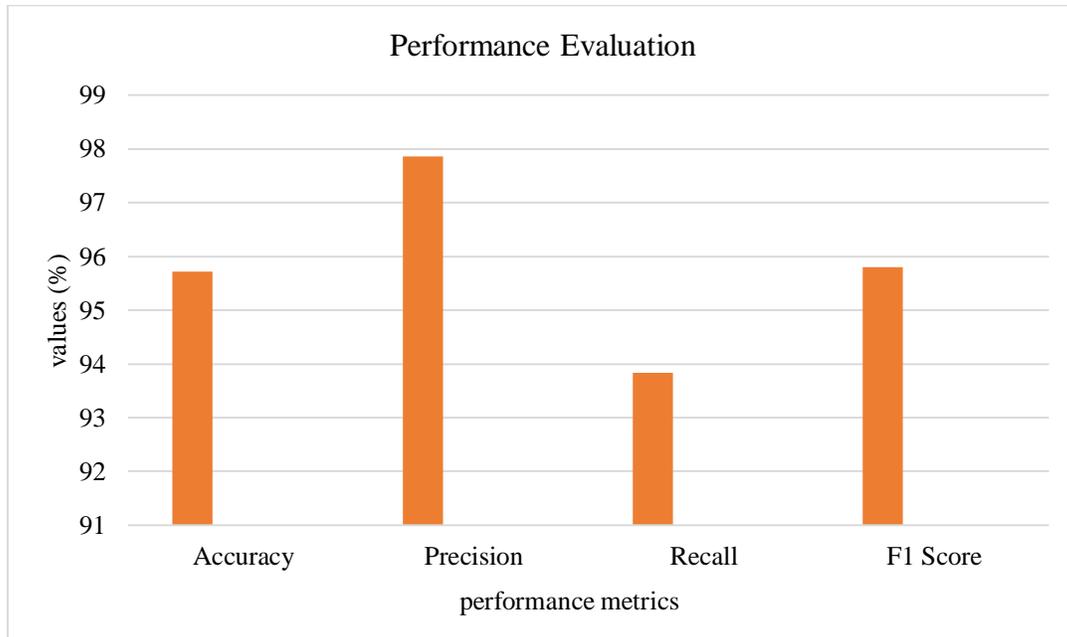


Fig. 15 Graphical representation with evaluation metrics

The accuracy and loss plot evaluate the model performance throughout the training process. The accuracy plot indicates how well the model fits the training data for the correct diabetes prediction. The accuracy plot illustrates the training and validation accuracy over 30 epochs for the proposed model, as shown in Figure 16. The training accuracy starts at around 0.6 and rapidly increases, reaching near-perfect accuracy (1.0) by the 10th epoch, after which it stabilizes. The validation accuracy also improves steadily, beginning at around 0.6 and reaching a peak of around 0.95 by the 20th epoch before showing minor fluctuations and stabilizing. The gap between training and validation accuracy suggests that the model fits the training data very well but may experience slight overfitting, as the validation accuracy does not quite reach the same level as the training accuracy.

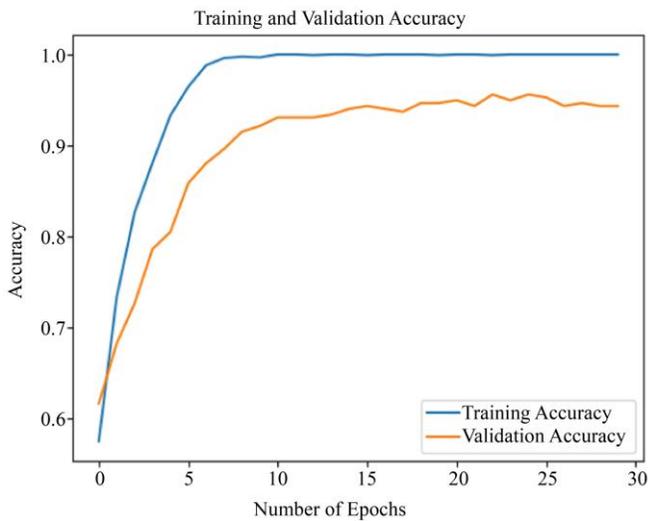


Fig. 16 Accuracy plot of the suggested model

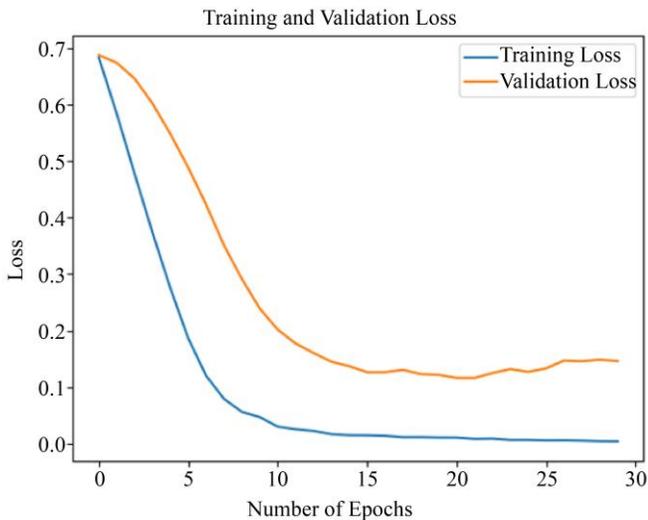


Fig. 17 Loss plot of the suggested model

The loss plot demonstrates a consistent decrease in both training and validation loss, as depicted in Figure 17,

providing insights into the performance of the suggested hybrid model by measuring the dissimilarity between the true output and predicted output. These plots exhibit no overfitting since no significant divergence exists between training and validation.

Initially, both the training and validation losses start high, around 0.7, indicating a substantial error in predictions at the beginning of training. The training loss drops sharply within the first 5 epochs, reaching close to 0.05 and gradually decreasing, eventually nearing zero by the 30th epoch. This indicates that the model is learning efficiently and fitting the training data well.

The validation loss also decreases steadily but at a slower pace compared to the training loss. It reaches a minimum of around 0.1 by the 20th epoch, with slight fluctuations afterwards. The gap between the training and validation losses after the initial epochs suggests some level of overfitting, as the model achieves near-zero training loss but maintains a slightly higher validation loss. Nonetheless, the low validation loss demonstrates that the model performs well in generalization, maintaining effective accuracy for unseen data making it robust for diabetes detection applications.

Figure 18 presents the confusion matrix of the suggested hybrid model, indicating the classification performance across healthy and diabetes cases by comparing the predicted values to actual values. It reveals the number of correctly and incorrectly classified data to identify the most often misplaced cases. In this case, 137 cases of diabetes were correctly classified as a diabetic class, and 131 healthy cases were correctly classified under the healthy class.

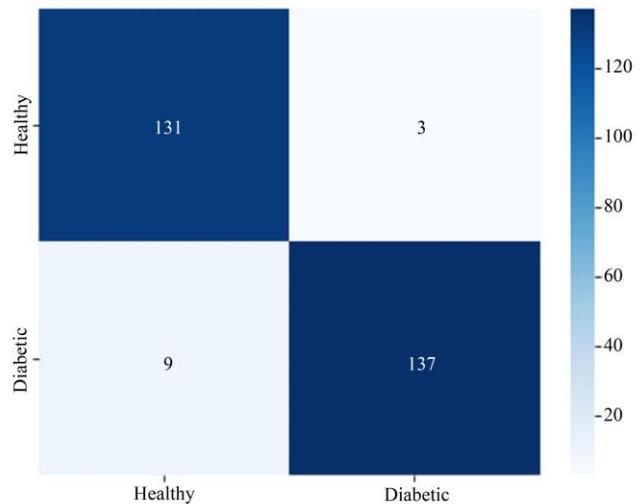


Fig. 18 Confusion matrix of the suggested model

The Receiver Operating Characteristics (ROC) curve illustrated in Figure 19 represents the model's performance across various thresholds. The ROC curve describes the trade-

off between sensitivity and specificity regarding probability. AUC represents the model performance across all threshold levels, with a value closer to 1 representing good performance. A perfect model shows a threshold having a TP rate of '1' and an FP rate of '0'. The ROC-AUC obtained from the proposed model is 0.9579, close to 1, indicating effective classification. Figure 20 illustrates the predicted output, indicating each diabetic case and healthy case were correctly predicted using the proposed model from the foot plantar thermography images.

4.2. Performance Comparison

Table 3 illustrates the performance comparison of state-of-the-art models with the proposed attention-based Inception V3 model with residual block, showing the ability of the proposed model to detect diabetes from foot plantar thermography images. Figure 21 represents the graphical representation of the performance comparison of the suggested model with existing models.

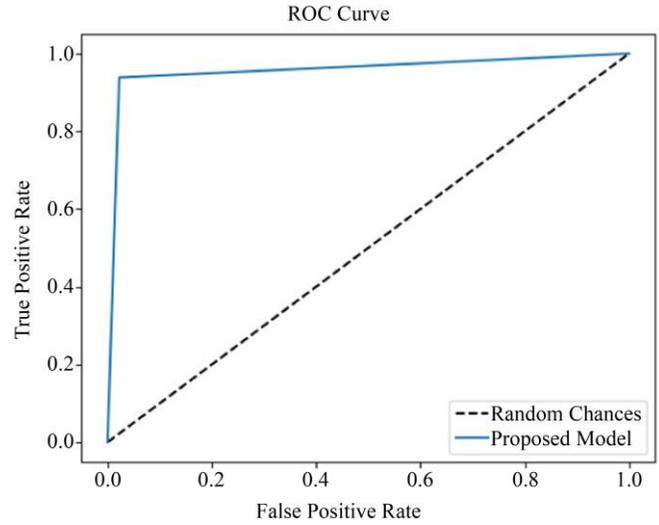


Fig. 19 ROC Curve

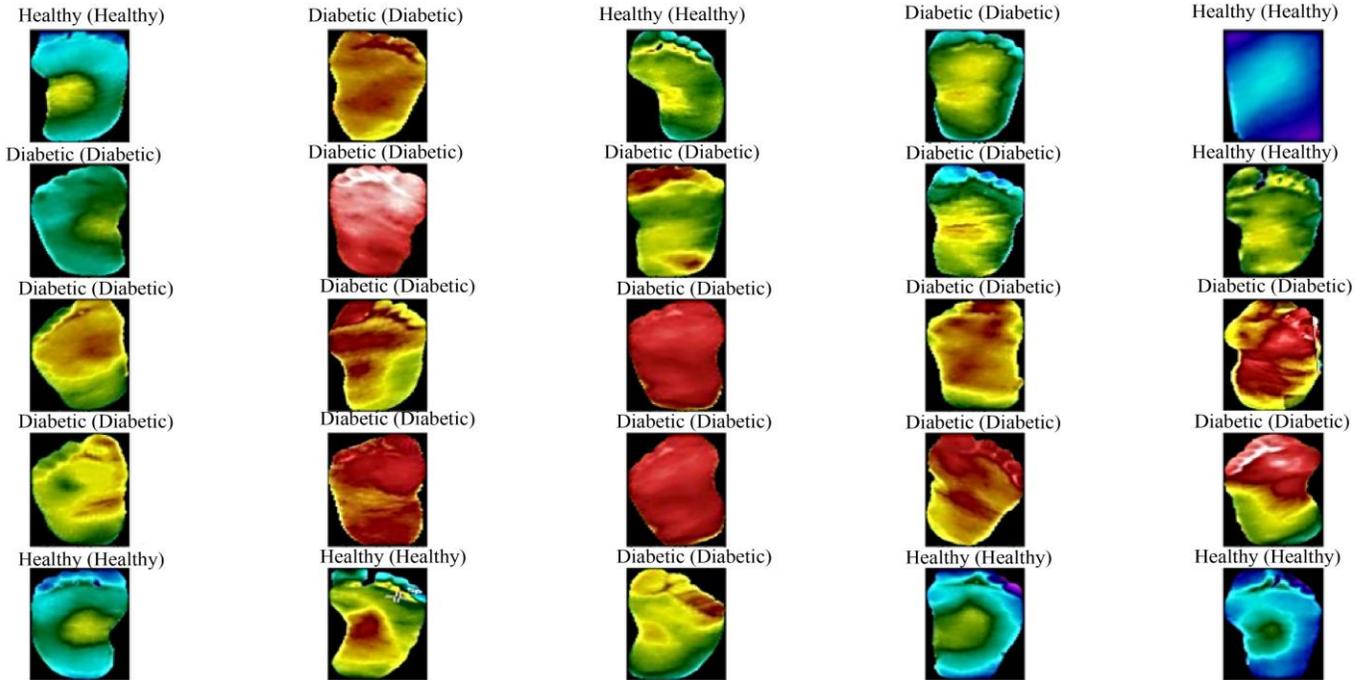


Fig. 20 Predicted output

Table 3. Comparison of the suggested hybrid model with existing methods

Methodology	Accuracy	Precision	Recall	F1 Score
VGG19	95.08	95.08	95.09	95.08
DFTNET	94.53	94.01	93.75	94.57
ResNet 50V2	82.39	80.32	78.65	79.07
DFINET	91.98	93.72	93.49	92.12
Proposed Hybrid Model	95.71	97.85	93.83	95.80

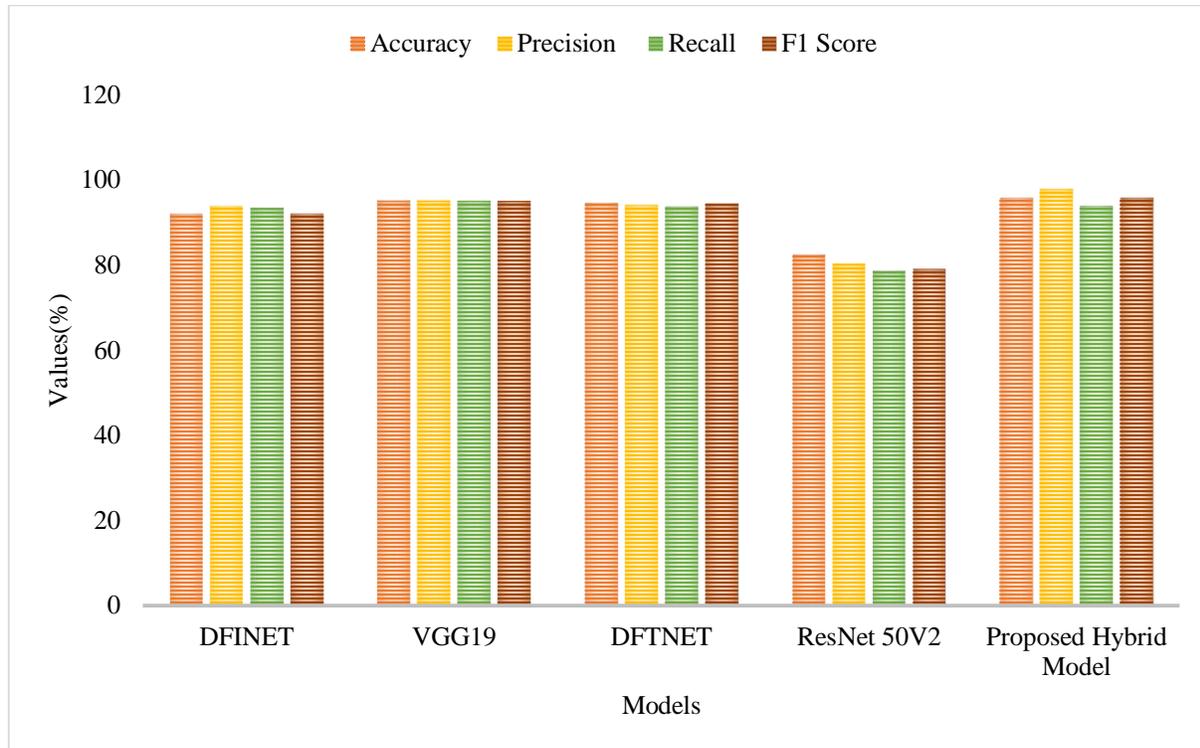


Fig. 21 Graphical representation of performance comparison

Table 3 illustrates that the proposed model demonstrated superior performance over conventional methods under all evaluation metrics in detecting diabetes from foot plantar thermography images. With 95.71 % accuracy, the proposed model surpassed DFINET, DFTNET, VGG19 and ResNet 50V2 models. The high values of precision and recall for the proposed model indicate a very low FP rate. Additionally, the 95.80 % F1 Score, the highest among all, reflects the highest balance between precision and recall. Thus, the proposed model combines the strengths of InceptionV3 and custom layers with attention mechanisms and residual blocks to improve classification performance, making it highly suitable for clinical applications.

5. Conclusion

Early intervention and suitable treatment of diabetes can prevent traumatic outcomes like lower extremity amputation. The circulatory deviations are vital in the pathogenesis of the diabetic foot. So, thermographic foot images are utilized for

the effective detection of diabetic cases since patients with diabetes show higher temperatures in foot regions than non-diabetic patients. The study proposed a hybrid deep learning model incorporating a pretrained inceptionV3, custom attention, and residual block layers. The results indicated the superior performance of the model with 95.71% accuracy, 97.85% precision, 93.83% recall and 95.80 % F1score. The reliability of the classification task is evaluated by comparing it with existing models, such as VGG19, DFINET, DFTNET, and ResNet50V2, across multiple performance metrics. The proposed model enhances real-time implementation in health care for better management of diabetic conditions and improves diagnostic practices in clinical settings.

Acknowledgements

The author expresses profound appreciation to the supervisor for providing guidance and unwavering support throughout the course of this study.

References

- [1] Joseph R. Kraft, "Detection of Diabetes Mellitus in Situ (Occult Diabetes)," *Laboratory Medicine*, vol. 6, no. 2, pp. 10-22, 1975. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Ali Mohebbi et al., "A Deep Learning Approach to Adherence Detection for Type 2 Diabetics," *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Jeju, Korea (South), pp. 2896-2899, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Aftab Alam et al., "Unveiling Diabetes: Categories, Genetics, Diagnostics, Treatments, and Future Horizons," *Current Diabetes Reviews*, vol. 20, no. 4, pp. 10-22, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [4] Åke Lernmark, "Type 1 Diabetes," *Clinical Chemistry*, vol. 45, no. 8, pp. 1331-1338, 1999. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Rodolfo Valdez, "Detecting Undiagnosed Type 2 Diabetes: Family History as a Risk Factor and Screening Tool," *Journal of Diabetes Science and Technology*, vol. 3, no. 4, pp. 722-726, 2009. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Darcy Barry Carr, and Steven Gabbe, "Gestational Diabetes: Detection, Management, and Implications," *Clinical Diabetes*, vol. 16, no. 1, pp. 1-4, 1998. [[Google Scholar](#)] [[Publisher Link](#)]
- [7] William J. Jeffcoate, and Keith G. Harding, "Diabetic Foot Ulcers," *The Lancet*, vol. 361, no. 9368, pp. 1545-1551, 2003. [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Maria Teresa Verrone Quilici et al., "Risk Factors for Foot Amputation in Patients Hospitalized for Diabetic Foot Infection," *Journal of Diabetes Research*, vol. 2016, no. 1, pp. 1-8, 2016. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Sara Ali Sulayman, Asma Rajab Qishqish, and Gofran Fathalla Dayhoom, "Early Detection of Diabetic Foot Using Thermal Imaging and Deep Learning Techniques," *African Journal of Advanced Pure and Applied Sciences*, vol. 3, no. 3, pp. 564-567, 2024. [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Garima Verma, "Leveraging Smart Image Processing Techniques for Early Detection of Foot Ulcers Using a Deep Learning Network," *Polish Journal of Radiology*, vol. 89, pp. 368-377, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Ikramullah Khosa et al., "Automatic Diabetic Foot Ulcer Recognition Using Multi-Level Thermographic Image Data," *Diagnostics*, vol. 13, no. 16, pp. 1-19, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Ilse Anahi Torres et al., "Proposal of a Non-Invasive Measurement of Physical Properties of Tissues in Patients with Diabetic Foot: Measurement Experiences in Diagnosed Patients," *Applied Sciences*, vol. 13, no. 4, pp. 1-16, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Natalia Arteaga-Marrero et al., "State-of-the-Art Features for Early-Stage Detection of Diabetic Foot Ulcers Based on Thermograms," *Biomedicine*, vol. 11, no. 12, pp. 1-16, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Zhenjie Cao et al., "Diabetic Plantar Foot Segmentation in Active Thermography Using a Two-Stage Adaptive Gamma Transform and a Deep Neural Network," *Sensors*, vol. 23, no. 20, pp. 1-17, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] J. Yogapriya et al., "Automated Detection of Infection in Diabetic Foot Ulcer Images Using Convolutional Neural Network," *Journal of Healthcare Engineering*, vol. 2022, no. 1, pp. 1-12, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Israel Cruz-Vega et al., "Deep Learning Classification for Diabetic Foot Thermograms," *Sensors*, vol. 20, no. 6, pp. 1-22, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Amith Khandakar et al., "A Novel Machine Learning Approach for Severity Classification of Diabetic Foot Complications Using Thermogram Images," *Sensors*, vol. 22, no. 11, pp. 1-23, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Andrés Anaya-Isaza, and Matha Zequera-Diaz, "Detection of Diabetes Mellitus with Deep Learning and Data Augmentation Techniques on Foot Thermography," *IEEE Access*, vol. 10, pp. 59564-59591, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Vuppala Adithya Sairam, Thermography Images of Diabetic Foot, Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/vuppalaadithyasairam/thermography-images-of-diabetic-foot>
- [20] Cheng Wang et al., "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," *IEEE Access*, vol. 7, pp. 146533-146541, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Paolo Arena et al., "Image Processing for Medical Diagnosis Using CNN," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 497, no. 1, pp. 174-178, 2003. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Xiang Li et al., "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," *International Journal of Network Dynamics and Intelligence*, vol. 2, no. 1, pp. 1-24, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Zahra Tabatabaei et al., "Residual Block Convolutional Auto Encoder in Content-Based Medical Image Retrieval," *IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop*, Nafplio, Greece, pp. 1-5, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]