ENSEMBLE DEEP LEARNING BASED CLASSIFICATION OF DIABETIC FOOT ULCER

Dissertation submitted in fulfillment of the requirements for the Degree of

MASTER OF TECHNOLOGY

By

Priyanshi Pandey

विद्या तत्व ज्योतिसमः

Department of Computer Science Engineering and Information Technology

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY
(Declared Deemed to be University U/S 3 of UGC Act)
A-10, SECTOR-62, NOIDA, INDIA

May – 2023

CONTENT

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the M.Tech Dissertation entitled **" Ensemble Deep Learning Based Classification of Diabetic Foot Ulcer"** submitted at **Jaypee Institute of Information Technology, Noida, India,** is an authentic record of my work carried out under the supervision of **Dr. Ankit Vidyarthi.** I have not submitted this work elsewhere for any other degree or diploma. I am fully responsible for the contents of my M.Tech theses.

Signature of the Scholar

Priyanshi Pandey
Department of Computer Science Engineering and Information Technology
Jaypee Institute of Information Technology, Noida, India
May 2023

SUPERVISOR’S CERTIFICATE

This is to certify that the work reported in the M.Tech Dissertation entitled “**Ensemble Deep Learning Based Classification of Diabetic Foot Ulcer”** submitted by **Priyanshi Pandey** at **Jaypee Institute of Information Technology, Noida**, **India,** is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Signature of Supervisor

Dr. Ankit Vidyarthi

Assistant Professor, Jaypee Institute of Information Technology, Noida, India

May 2023

PREFACE AND ACKNOWLEDGEMENT

Writing this thesis has been an incredible journey of discovery, growth, and challenge. The aim is to develop and apply an ensemble model to accurately recognize diabetic foot ulcers from medical images, outperforming commonly used models and improving patient outcomes. I would like to acknowledge all the people who have helped and supported me for the success of this thesis.

First and for most, I would like to express my heartfelt gratitude to my supervisor, **Dr. Ankit Vidyarthi** for hir patience, kind, continuous support, and encouragement throughout my research. Hir guidance and continuous support are priceless to me. It has been a great honor and pleasure for me to do my M.Tech under her supervision.

Furthermore, we would also like to acknowledge with much appreciation the crucial role of the panel for my mid-term evaluation, whose suggestions to improve project outlines and features gave direction to the project and helped in its improvement. A special thank goes to my batch mates, who also gave suggestions to improve further. Last but not least, I would like to thank all the professors who have invested their effort in guiding us throughout the academic curriculum with the technical expertise required in achieving the goal.

Signature of the Student

Priyanshi Pandey

May 2023

ABSTRACT

This dissertation investigates the use of transfer learning models to classify diabetic foot ulcers (DFUs) utilising ensemble learning approach. The study uses assessment criteria including accuracy, precision, recall, and F1 score to examine the performance of ten well-known pre-trained models, including ResNet50, InceptionV3, Xception, VGG16, MobileNetV3Small, ResNet101V2, InceptionResNetV2, MobileNetV2, and DenseNet121.

The results show that, in terms of accuracy, precision, recall, and F1 score, InceptionResNetV2 and EfficientNetB0 display the greatest individual model performance. An ensemble model is developed by combining the strengths of InceptionResNetV2 and EfficientNetB0 in order to further improve classification accuracy.

By outperforming the individual models in terms of accuracy, precision, recall, and F1 score, the ensemble model triumphs. This demonstrates how ensemble learning may enhance DFU categorization. The work offers insightful information on ensemble learning and its potential benefits in the processing of medical images.

The suggested method shows the potential of ensemble learning in improving DFU diagnosis and categorization, which is essential for efficient treatment planning and patient care. The advantages of ensemble learning in the context of DFU classification are highlighted in this study, which makes a contribution to the field of medical image analysis.

LIST OF ACRONYMS AND ABBREVIATIONS

AUC	Area Under the ROC Curve
CNN	Convolutional Neural Network
CGM	Continuous Glucose Monitoring
DFU	Diabetic Foot Ulcer
DL	Deep Learning
FP	False Positive
GAN	Generative Adversarial Networks
GRU	Gated Recurrent Units
LSTM	Long Short-term Memory
ReLu	Rectified Linear Activation Unit
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
RNN	Recurrent Neural Network
TN	True Negative
TP	True Positive
VAE	Variational Auto Encoder
VGG	Visual Geometry Group

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND AND SIGNIFICANCE OF THE STUDY

A significant portion of the diabetic population suffers from the frustrating problem of DFU, a well-known consequence of diabetes mellitus. The timely detection and appropriate management of diabetic foot ulcers represent a paramount necessity to prevent complications and enhance patient outcomes. Within the vast expanse of healthcare, deep learning, a subfield of artificial intelligence, has emerged as an invigorating prospect, offering an unbridled potential in a myriad of healthcare facets, spanning medical imaging, diagnosis, and treatment. To this end, this dissertation endeavors to conceive and appraise a deep learning-based classification model for diabetic foot ulcers, leveraging a sizable corpus of patient records, primed to classify diabetic foot ulcers into disparate categories based on the severity and other clinical characteristics.

The salience of this study is unequivocal, and lies in its potential to advance the accuracy and efficiency of diabetic foot ulcer diagnosis and classification. Harnessing the prodigious capacity of deep learning algorithms, clinicians can expeditiously scrutinize colossal swaths of data, decode the intricate patterns and risk factors, and beget indispensable insights that can guide their treatment decisions. Such a paradigm shift would ostensibly result in the earlier detection of the disease, a more efficacious treatment regimen, and ultimately better patient outcomes. Beyond that, the deep learning model proposed in this study could also serve as a potent screening and monitoring tool for diabetic foot ulcers, leading to reduction of healthcare cost and outcomes of patient health. Furthermore, it could galvanize the innovation of novel treatments and interventions for DFUs.

The application of DL in classification brings several decisive advantages in comparison to traditional machine learning algorithms. Deep learning algorithms can autonomously extract features from raw data, such as medical images, and learn the underlying representations without requiring the intervention of manual feature engineering. This renders deep learning particularly apt for tasks that necessitate the analysis of vast and intricate datasets. In sum, DL based classification model for diabetic foot ulcer is developed which carries a promise of a momentous impact on clinical practice, elevating patient outcomes, and alleviating the burden of diabetic foot ulcer on individuals and healthcare systems alike. Although DL application in the field of healthcare is still in its inchoate stages, its potential to revolutionize healthcare is immense and unassailable.

1.2 RESEARCH QUESTIONS AND OBJECTIVES

The purpose of this dissertation is to comparison of models based on DL algorithms that might be used to detect diabetic foot ulcers [1]. Diabetes frequently causes diabetic foot ulcers, which can have disastrous effects if left untreated and can also be life threatening. One of the fast evolving field in artificial intelligence is DL, that possess the potential to transform the detection, treatment, and help in prevention of DFU. The effectiveness of ulcer diagnosis and prevention techniques on the basis of DL will be evaluated, and also the classification will be performed to detect the likelihood of diabetic foot ulcers.

The objective can be achieved, we will analyze and compare various deep learning models that can help to classify diabetic foot ulcers and assess the accuracy and reliability of various approaches with our proposed model. We will also investigate and assess how the model performs when compared to other DL models in predicting the probability of a diabetic foot ulcer developing and evaluate the efficiency.

Traditional wound assessment and care techniques depend on human assessment, which can be subjective and time-consuming, delaying treatment and harming patient results. The management of DFU, a common complication that is faced by diabetes patients, presents significent challenges for healthcare providers as it can result in serious complications.

DL has the potential to alter the treatment of DFU by delivering efficient and effective as well as more precise methods for treating wounds. This study will improve our understanding of the topic by addressing research concerns and objectives linked to implementation of DL in the early detection and treatment of diabetic foot ulcers. The overarching objectives of this project are enhancing the efficiency and precision of DFU diagnosis, anticipating and preventing the emergence of DFUs, and providing a reliable and efficient plan for wound assessment and management. This work has the potential to lessen the burden of this diabetes complication and enhance the chances of better treatment of anyone that is suffering from DFU by utilizing the power of deep learning. The outcomes of this study will significantly affect the DFUs treatment and the body of knowledge on the application of deep learning in healthcare.

1.3 SCOPE AND LIMITATIONS OF THE STUDY

Around 25% of people with diabetes have the possibility to develop diabetic foot ulcers at one time in thier lives[2], making them a frequent complication of the disease. These ulcers can drastically lower a person's quality of life, pose a threat to their life, and have negative side effects including infections and amputations. Traditional wound assessment and management techniques mainly rely on arbitrary, time-consuming, and prone to error human evaluation. Deep learning models have shown a lot of promise recently for improving medical image processing and diagnosis. By leveraging large datasets and potent algorithms, these models can learn to spot patterns and features in medical images that are difficult for human professionals to discover.

Effectiveness and efficiency of categorization and severity assessment in the context of DFUs can be enhanced using deep learning algorithms, potentially improving patient treatment outcomes. The study's main purpose is to examine the potential of DL techniques for treating diabetic foot ulcers, notably in the classification and severity assessment domains. The project will compare the accuracy of deep learning models to other models and traditional approaches for evaluating and controlling wounds. The goal of the study is to add to the expanding body of knowledge regarding deep learning algorithms applications for detection of DFUs at an early stage.

The results obtained from the model proposed have the potential to make a significant impect on the field of classification of diabetic foot ulcers by improving the accuracy and speed of diagnosis, anticipating and preventing the development of diabetic foot ulcers, and providing a successful and reliable wound assessment and management strategy. Thus, those who have this diabetes complication may experience better patient outcomes and a higher quality of life as a result.

Another limitation is that this study solely takes into account the implementation using DL models for categorization and severity assessment. Additional pertinent factors that might have an impact in the process of development and management of DFU, such as patient demographics, comorbidities, and wound care practices, were not considered in this study. It is critical to keep in mind that the use of DL models can be considared in conjunction with other factors and features that may affect how diabetic foot ulcers are managed and treated generally.

Additionally, the study did not consider any potential ethical repercussions of using deep learning models for medical diagnosis and treatment. Deep learning models may improve the precision of diagnoses and the effectiveness of treatments, but it's crucial to consider any biases and unforeseen consequences that may arise from their use. Future studies should investigate the ethical implications of utilizing deep learning models in the healthcare industry and develop the required regulations and safety measures to ensure their responsible and secure use.

1.4 OVERVIEW OF THE DISSERTATION STRUCTURE

The dissertation objective is to provide a complete analysis of deep learning potential in the field of categorization as well as severity evaluation of diabetic foot ulcers. It does this by using multi-class classification and comparison with existing deep learning models. The study will be rationally and cogently constructed, starting with a general overview about the topic of DFUs, their prevalence, and the current state of diagnosis and treatment. The motivation behind the study will also be discussed in the introduction, which highlights the need for better and more accurate diagnostic techniques for diabetic foot ulcers.

The deep learning approaches implemented to categorize as well as detect the level of the severity of diabetic foot ulcers will be thoroughly examined in the literature review section. The many DL models that have already been implemented will be covered in this part, along with a review of relevant publications. The assessment will also identify any gaps that the current inquiry aims to fill and criticize any faults of past investigations.

The methodology section of this study will include a description of various existing algorithms and the approach taken, along with information about the dataset used, the deep learning models applied, and the evaluation metrics compared to determine how well the models performed. This part will go over the pre-processing techniques used to enhance the performance by working on the dataset for analysis, such as augmentation of data and methodologies used for normalisation, to make sure that the dataset is representative of actual situations. Then data optimization techniques are used to decrease time complexity. Finally, discuss the proposed methodology.

The study findings are reported in the results section along with the performance evaluation for each deep learning model that was used. The findings will also be compared to current deep learning models for categorization and severity assessment of diabetic foot ulcers in this section. Relevant tables and figures will be added to the results area to help with data display and comprehension. A critical evaluation of the findings and their implications for the field of treating and detecting diabetic foot ulcers will be provided in this section. This section will go over the study's advantages and disadvantages as well as the shortcomings of the most recent research. The section will go over possible uses of deep learning for the detection at an early stage and treatment of DFUs and will offer suggestions for further study.

In the conclusion section, which also highlights the study's contributions to the field of classification of DFUs and treatment, the primary findings from the research will be discussed. The implications of the findings for clinical practises and patient care will also be discussed in this section. The section will conclude by discussing the study's limitations and presenting suggested actions for future research, emphasizing the areas that may require additional research.

The primary objective is to conduct a thorough analysis of DL competence to enhance detection at early stage and management of DFUs. The objective is to present insightful information concerning classification and grading of diabetic foot ulcers employing deep learning techniques, which may ultimately enhance the accuracy and efficacy of diagnostic methods in this critically important area of healthcare.

CHAPTER 2
LITERATURE REVIEW

2.1 DEFINITION AND PREVALENCE OF DIABETIC FOOT ULCER

A prevalent hazardous diabetic side effect is the development of DFUs, impairing circulation and sensation and happens when glucose levels in blood to rise and harm the nerves in the feet and also the blood vessels. It is a chronic, non-healing wound that can appear on the toes, heels, or soles of the foot and may result in amputation in addition to other negative health consequences. Diabetes-related foot ulcers can also raise the threat of infection, which may retard healing and raise the probability of amputation.

With an estimated 15% of diabetics experiencing a foot ulcer at a particular point in their lives[3], diabetic foot ulcers are extremely prevalent. The amount of time of diabetes, poor glucose control, and the existence of additional factors that can be cause further risk such peripheral neuropathy and vascular disease, and foot abnormalities all strengthen the probability of developing a foot ulcer. The risk of developing DFU increases even more by the fact that they smoke, have high blood pressure, have a history of foot ulcers or amputations, or any of the aforementioned factors. With the global prevalence of diabetes increasing, with it will be expected that the occurrence of diabetic foot ulcers will increase significantly. Different populations and geographical regions have distinct incidences of diabetic foot ulcers.

Diabetes-related foot ulcer prevalence in India has been predicted to be around 4 and 10% [4]. This is in accordance with the global estimate that a diabetic foot ulcer impacts 15% among individuals with diabetes. Given the high prevalence of diabetes in the nation, India's high rate of diabetic foot ulcers is concerning. After China, India possesses the second-highest global diabetes prevalence rate. In India, there were approximately 77 million persons diagnosed with diabetes in 2019. This means that a substantial portion of the population in India is vulnerable to acquiring diabetic foot ulcers and the complications that accompany it. The critical importance of addressing diabetic foot ulcers in this population as soon as they are identified in order to prevent complications like infection and amputation is emphasized through this.

The estimated 62 million people with diabetic foot ulcers worldwide are alarming since it puts a significant burden on the healthcare system and those who are affected by this condition. Due to the complexity of the problem, a multidisciplinary approach encompassing wound care, glycemic control, and comorbidities management is required for treating DFU. Detection at an early stage and starting treatment at early stage of diabetic foot ulcers are crucial for preventing complications and optimizing outcomes for diabetics. The high incidence of DFU in India and throughout the world underlines the importance of early detection, efficient therapy, and prevention of this condition to decrease the burden it takes on people and healthcare systems.

2.2 RISK FACTORS AND CAUSES OF DIABETIC FOOT ULCER

Foot ulcers are a prevalent consequence of diabetes which, left untreated, can have serious consequences for one's health. Numerous risk factors and causes can be attributed in the formation of DFU.

Major risk factors is peripheral neuropathy, which is nerve damage brought on by diabetes. It might be difficult for patients to detect sores or foot ulcers due to peripheral neuropathy, which can cause a decreased or complete lack of sensation in the feet. This could lead to complications like infections and amputations. Peripheral artery dysfunction is another danger sign for diabetic foot ulcers. It is possible that this illness, which is brought on by blood vessel damage, will result in poor circulation in the feet. Poor circulation can hinder both the ability to heal wounds and the danger of infection.

A significant risk factor for diabetic foot ulcers is high blood glucose levels. Uncontrolled diabetes increases chances of formation of foot ulcers by damaging the nerves and blood vessels in the feet. The chance may be further increased by foot deformities such Charcot foot, hammer toes, and bunions. These elements contribute to increased pressure on particular foot areas, an increased risk of skin deterioration, and the development of ulcers.

Foot injuries and ulcers might also be more common while wearing improperly fitting footwear or walking barefoot. Smoking can harm blood vessels and impair circulation, which raises the risk of foot ulcers and consequences. Poor foot hygiene is another risk and a primary cause of diabetic foot ulcers. Neglecting foot cleanliness can lead to various foot issues as well as an increase in the risk of infections. Obesity, poor nutrition, and someone who already had foot ulcers or amputation are at more risk for development of dfu.

Although there are a variety of potential underlying causes for diabetic foot ulcers, numerous elements, such as poor circulation, nerve damage, and pressure on the feet, frequently contribute to their occurrence. Foot ulcers may develop as a result of foot trauma from wounds like burns, bruises, or scrapes. When socks, shoes, or other things rub against the foot repeatedly, ulcers may form.

A foot ulcer may contract bacteria or fungi, which may make healing more challenging and raise the possibility of complications. Diabetes can hinder the body's ability to heal injuries, which can result in the development of recurrent foot ulcers. In addition to the weakening and collapse of the foot's bones, Charcot foot can result in deformities, ulcers, and other conditions. Maintaining their feet sanitary, wearing shoes that fit properly, and periodically checking their feet for cuts or ulcers are all necessary for this.

2.3 CLASSIFICATION AND STAGING SYSTEMS FOR DIABETIC FOOT ULCER

Healthcare experts classify diabetic foot ulcers according to their severity and characteristics using categorization and staging systems in order to manage and treat them effectively. The subsequent section includes some of the most well-liked classification and stage systems for diabetic foot ulcers:

1. The Wagner Classification method: Diabetic foot ulcers are categorized by employing the Wagner classification method, which has been extensively employed[5]. It offers a straightforward approach for classifying the severity of the ulcer according to the depth and scope of tissue involvement. The grading scale goes from grade 0, which denotes a lesion that is just beginning to ulcerate, to grade 5, which denotes the need for gangrene or an amputation. Based on the severity of the ulcer, this system aids clinicians in selecting a course of action and forecasting results. The Wagner classification system is helpful for standardizing the reporting and documenting of diabetic foot ulcers in clinical practice and research.

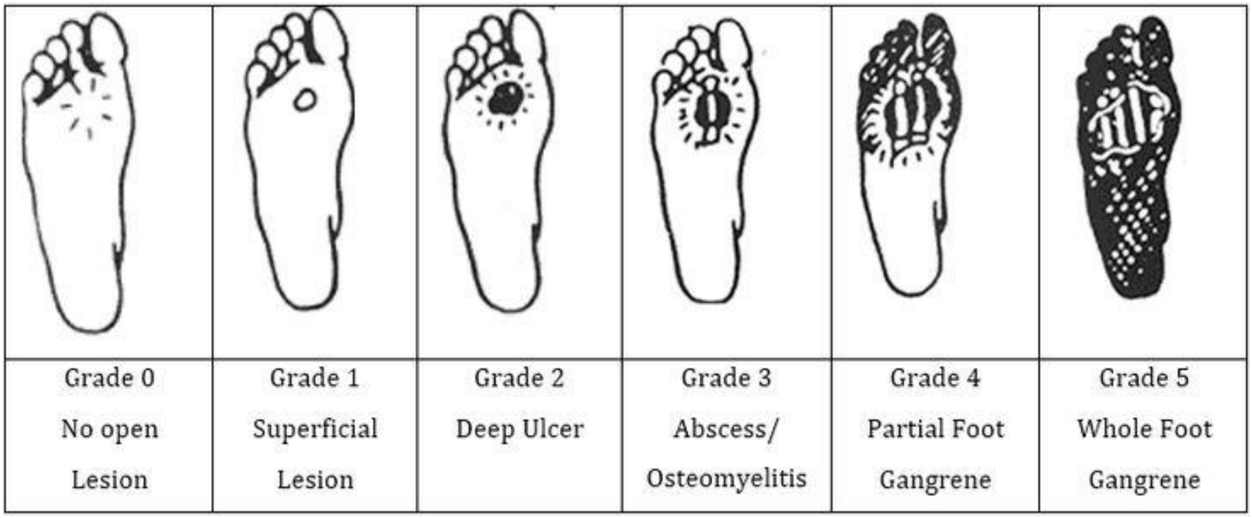

Fig 2.1: Wagner Classification System [5]

2. University of Texas Diabetic Wound Classification System: This system [6] classifies diabetic wounds according to the depth of the ulcer and whether an infection is present. Table 2.2 depicts all the four stages it assigns to diabetic foot ulcers. This approach is extensively used to categorize diabetic foot ulcers according to their severity and inform treatment choices in medical practice and research. It is helpful to provide a thorough examination of dfu using this classification model, that additionally aids medical professionals in making educated decisions regarding treatment.

	0	**1**	**2**	**3**
A	Pre or postulcerative lesion completely epithelialized	Superficial wound, not involving tendon, capsule, or bone	Wound penetrating to tendon or capsule	Wound penetrating to bone or joint
B	with infection	with infection	with infection	with infection
C	with ischemia	with ischemia	with ischemia	with ischemia
D	with infection and ischemia	with infection and ischemia	with infection and ischemia	with infection and ischemia

Table 2.1 : University of Texas Diabetic Wound Classification System [6]

3. PEDIS Classification System: Diabetic foot ulcers can be categorized using the PEDIS Classification System, which is extensively used [7]. Perfusion, size of ulcer, loss, the depth of

ulcer and level of infection, and feeling are the five variables it takes into account. This approach aids clinicians in determining the extent of diabetic foot ulcers and informing treatment choices. The PEDIS classification system is helpful for standardizing the reporting and documenting of diabetic foot ulcers in clinical practice and research. It offers a thorough analysis of diabetic foot ulcers, taking into account a number of variables that impact their severity and potential treatments.

Grade	**Perfusion**	**Extent**	**Depth**	**Infection**	**Sensation**	**Score**
1	No PAD	Sin intact	Skin intact	None	No loss	0
2	PAD, No CLI	<1 cm^2	Superficial	Surface	Loss	1
3	CLI	1-3 cm^2	Fascia, muscle, tendon	Abscess, fasciitis, septic arthritis		2
4		>3cm^2	Bone or joint	SIRS		3

PAD, peripheral disease; CLI, critical limb ischemia

Table 2.2 : PEDIS Classification System [7]

These classification and staging systems have helped healthcare providers to improve the assessment of the severity and scope of DFU , which enables more focused and efficient treatment as well as early treatment

2.4 MEDICAL IMAGE ANALYSIS

Medical image analysis is useful for accurately diagnosing and treating diabetic foot ulcers (DFU). Medical image analysis tools can help identify and diagnose DFU by analyzing photographs of the foot and identifying regions of tissue damage or ulceration. Digital photography is one of the most widely used imaging techniques for DFU as it results in image with high-resolution of the foot that can be examined using computer vision algorithms. Deep learning techniques, such as semantic segmentation methods using convolutional neural networks (CNN), and speedier R-CNN methods, have been used in a number of studies for DFU segmentation and identification. In order to solve the lack of tagged medical image data for DFU, additional deep learning techniques, such as transfer learning.

Class knowledge banks and mapped binary patterns have also been used to completely identify DFUs. The promise of cloud-based deep learning frameworks allows even patients with outdated mobile technology to access the advancements made recently in automated.

The use of deep learning for medical image analysis has increased recently because accessibility of large datasets of medical images and the exceptional performance of these models in a range of applications. Deep learning approaches are classified into two further main categories: image categorization and image segmentation.

2.4.1 Image Classification

Labeling or classifying a whole image based on its properties is a deep learning challenge known as image classification in context of analysis of medical image. Image classification is basically implemented to differentiate several kind of medical images into categories like normal vs aberrant or malignant vs benign this is an essential function in medical imaging since it may help automate the procedures of making diagnoses and formulating treatment plans.

CNNs are mostly implemented deep learning technique for classification of images tasks in medical image analysis since cnn have been developed to interact with images they can automatically learn features at numerous degrees of abstraction from the input image they are therefore especially appropriate for image classification tasks due to the fact they are capable of identifying patterns that are not evident immediately as well as the features to the human eye.

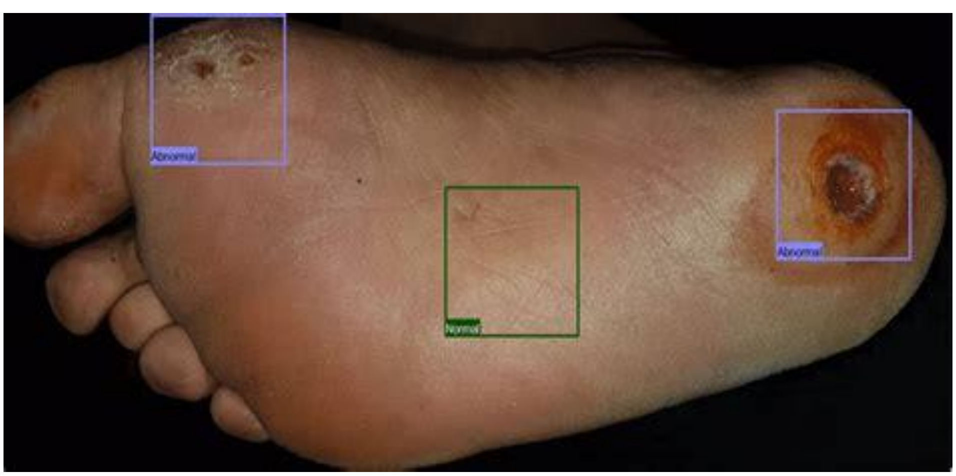

Fig 2.2: Image classification on DFU [8]

In medical image classification the first step is typically to preprocess the image data to standardize the format and reduce noise the preprocessed images are then fed for training purpose into a cnn model on a labeled dataset during training phase identification of features in the images is learned by CNN that are associated with some certain class or category or label it and firther implemented on the test dataset when the model is trained.

Evaluation metrics can be used for measurement of the parformance for image classification model with help of these evaluation criteria assessment of how well the classification is done by the model and classify them according to their classes a high performing image classification model can have important clinical implications as it can enable faster and more accurate diagnosis and treatment planning for patients

2.4.2 Image Segmentation

Fragmentation of an image apart into various segments or ROI that basically corresponds to different sections of image or objects present is known as Image Segmentation. This method enables a more in-depth analysis of medical images and can be used to detect and track abnormalities or pathologies in medical imaging. For example, image segmentation can be used to locate tumors, gauge the size of organs, identify particular structures in the brain, and more.

In order to collect contextual data on several scales and produce precise segmentation results, it employs an encoder decoder structure. The decoder portion of the network uses up sampling and concatenation operations to restore the spatial resolution of the original image and produce a segmentation mask. The encoder portion of the network is made up of convolutional and pooling layers that down sample the image and capture features at various scales.

Mask R-CNN is another deep learning approach that combines object detection and image segmentation to identify and segment objects of interest within an image. the network first generates object proposals using a region proposal network (RPN), which identifies regions that has the most probability of having object in the image the proposals are then refined using a ROI pooling layer and fed through a convolutional network to generate class predictions and bounding box regressions. Finally, a mask prediction branch is added to the network which generates a binary mask for each object proposal indicating which pixels belong to the object.

DeepLabv3+ is a variant of the deep lab architecture that uses atrous convolutional layers to improve the performance of semantic segmentation tasks. Atrous convolution layers allow the network to capture multi scale contextual information and make sure that the spacial resolution is preserved of the input image the network also capture different scale features using a feature pyramid network and a decoder module that refines the segmentation results image segmentation using deep learning has the potential to revolutionize medical imaging by enabling automated and accurate detection, quantification, and tracking of abnormalities or pathologies, which can improve patient diagnosis and treatment.

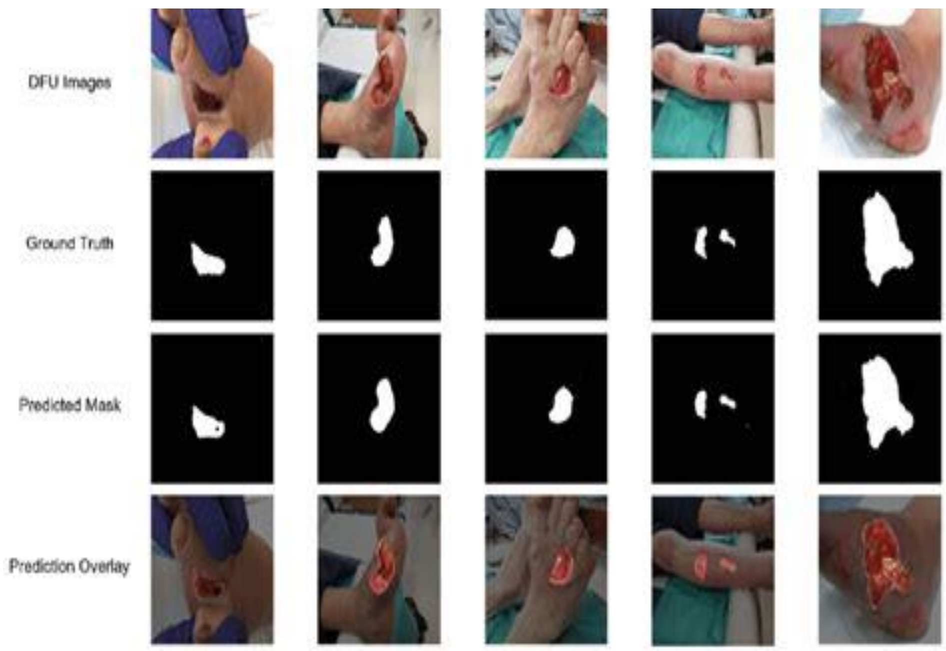

Fig 2.3 : Image Segmentation in DFU images [9]

Medical image analysis in general plays an essential role for the early detection and controlling of DFU. Deep learning and transfer learning, together with classification and segmentation techniques, have the potential to significantly improve patient outcomes by greatly increasing the accuracy and efficacy of DFU diagnosis and management.

2.5 RELATED STUDIES ON USING DEEP LEARNING FOR DIABETIC FOOT ULCER CLASSIFICATION

A chronic complication of diabetes is a lesion or sore on the foot known as a diabetic foot ulcer (DFU), which either doesn't heal or becomes worse over time. Patients with DFU suffer significantly, which negatively impacts their overall standard of life and elevates their probability of morbidity and mortality. Early detection and the appropriate treatment are crucial to preventing more serious problems like infections or lower extremity amputations.

Deep learning (DL), an instance of machine learning, uses artificial neural networks for assessing complex data and generating classifications or forecasts based on patterns identified in the data. Medical imaging research has an abundance of potential for DL, particularly when it relates to recognizing and categorizing DFU. Studies have demonstrated the significance of DL in the detection and classification of DFU.

Qu et al. [12] proposed a deep learning model for segmentation as well as classification automatically of DFU images. The model is designed to address the challenges of DFU image analysis, such as large inter- and intra-class variations, different lighting conditions, and different image resolutions. The two major components of the proposed model, an image segmentation network based on U-Net architecture CNN based algorithm for the task for image classification. The U-Net architecture is a popular and efficient image segmentation method that uses an encoder-decoder to capture various contextual information and achieve accurate segmentation results. A CNN-based image classification network was then used to classify the segmented ROIs into three categories: non-ulcerated, superficially ulcerated, and deeply ulcerated. The classification network was trained on a subset of the same DFU dataset and optimized to achieve high accuracy in detecting different types of ulcers. The final model achieved 91.7 percent accuracy in segmenting DFU images and 94.3 percent accuracy in classifying the three classes of DFU images.

Hassanie et al. [13] proposed a model based on deep learning algorithms to detect and classify DFU based on their severity. The model combined two popular neural network architectures, CNN and LSTM to form a hybrid LSTM-CNN. CNN was used to extract spatial features from DFU images to enhance the performance and accuracy, while LSTM was used to capture the temporal dependence of the sequential input of DFU images and classify them. The algorithm used in model used a dataset of 450 DFU images for training and testing collected from patients with DFUs of varying severity. The model achieved 87.3% accuracy in identifying DFU images, meaning that it correctly identified 87.3% of the DFU images in the dataset. The model also achieved 80% accuracy in classifying DFU images into three severity levels, i.e., mild, moderate and severe. The model was able to accurately classify 80% of the DFU image into three different classes according to their respective severity levels.

He et al. [14] developed a deep learning (DL) model for automatic recognition of DFU images using transfer learning. The model used the InceptionV3 architecture, which was pre-trained on a large dataset of nature images. Transfer learning allows the model to use the knowledge learned during pre-training on images for detection of the features of DFU images and further classify. The model achieved an impressive 98.5% accuracy in identifying DFU images, outperforming several other methods that already exists. The efficiency was increased of DL models for DFU analysis using transfer learning is demonstrated in this study.

Hossain et al. [15] proposed a DL model for automatic detection and classification of DFU images into three classes—no ulceration, shallow ulceration, and deep ulceration. The model was based on a CNN architecture and trained on a large dataset of DFU images. The CNN architecture is well suited for image classification tasks and has proven effective in many computer vision applications. In identifying DFU images the model achieved 92.5 percent accuracy and in classifying the three classes of DFU images the model achieved 89.2 percent accuracy. This study highlights the potential of DL models in identifying and classifying different types of DFU based on their severity.

Zhang, J. et al. [16] proposes a deep ensemble learning framework that combines transfer learning and traditional ensemble techniques for image classification. They compare various transfer learning models and ensemble methods, achieving superior performance. The ensemble approach improves classification accuracy and robustness by leveraging the diverse knowledge of transfer learning models.

Liu, Y. et al. [17] study investigates ensemble deep learning models that incorporate transfer learning for image classification. Multiple transfer learning models, such as VGG16, ResNet50, and InceptionV3, are combined through voting, resulting in improved accuracy and robustness. The ensemble approach effectively captures complementary information from different models, enhancing classification performance.

Nguyen, T. D. et al. [18] proposes an ensemble deep learning model that integrates transfer learning for image classification. They explore ensemble techniques like bagging and boosting with transfer learning models such as AlexNet and GoogLeNet. The ensemble model achieves higher accuracy and reduces overfitting. Combining transfer learning models in an ensemble framework improves classification performance by providing a diverse representation of the data.

The study of DFU photos has generally yielded good findings, and numerous studies have demonstrated excellent accuracy in the detection and classification of DFU images. To corroborate these findings in larger datasets as well as develop more efficient DL models for precise and efficient comprehension of DFU images, more research is required.

2.6 COMPARITIVE STUDY OF RESEARCH PAPERS

Study Title	**Proposed Work**	**Limitations**	**Advantages**
Qu et al. [12]	- Deep learning model for DFU image segmentation and classification using U-Net architecture.	- Limited evaluation on real-world DFU images.	- Accurate segmentation and classification of DFU images using U-Net architecture.
Hassanie et al. [13]	- Hybrid LSTM-CNN model for DFU detection and severity classification.	- Relatively small dataset used for training and testing.	- Effective utilization of both spatial and temporal information for DFU detection and severity classification.
He et al. [14]	- Recognition of DFU images using transfer learning with InceptionV3 architecture.	- Dependency on pre-training dataset for transfer learning.	- High accuracy in identifying DFU images through transfer learning with InceptionV3 architecture.

Hossain et al. [15]	- CNN-based model for automatic detection and classification of DFU images into three categories.	- Limited exploration of other features besides CNN.	- Successful detection and classification of DFU images into three categories.
Zhang, J. et al. [16]	- Deep ensemble learning framework combining transfer learning and traditional ensemble techniques for image classification.	- Computational complexity due to ensemble framework.	- Improved classification accuracy and robustness through deep ensemble learning.
Liu, Y. et al. [17]	- Ensemble deep learning models incorporating transfer learning for image classification.	- Potential model complexity and increased computational requirements.	- Enhanced accuracy and robustness through ensemble approach and transfer learning.
Nguyen, T. D. et al. [18]	- Ensemble deep learning model integrating transfer learning with bagging and boosting techniques for image classification.	- Ensemble approach may require more computational resources.	- Higher accuracy and reduced overfitting through ensemble model with transfer learning.

Table 2.3: Comparative Study Of Research Papers

CHAPTER 3
METHODOLOGY

3.1 DATASET

The DFU classification dataset obtained from Kaggle consists of a total of 1055 images[19][20][21]. These images are categorized into two classes: "Healthy Skin" and "Ulcer." The dataset is intended for the task of classifying diabetic foot ulcer (DFU) images into these two categories. Each image in the dataset represents a specific area of a foot, captured using various imaging techniques.

The images in the "Healthy Skin" class depict normal, non-ulcerated skin areas. These images serve as the reference for identifying and distinguishing healthy foot regions from those affected by ulcers. On the other hand, the images in the "Ulcer" class represent foot regions with visible signs of ulcers, such as open wounds, lesions, or other ulcer-related characteristics.

Fig 3.1: Labeled images from Training dataset

In addition to the main DFU classification dataset, a separate test dataset is available, which includes unlabeled images. These unlabeled images are provided for evaluation and testing purposes, allowing researchers and practitioners to assess the performance of classification models on unseen data.

The test dataset consists of a collection of DFU images that have not been assigned any class labels. These images are representative of real-world scenarios where new and unseen images need to be classified. By using this test dataset, researchers can assess the generalization capabilities and performance of their DFU classification models on previously unseen cases.

	ULCER	HEALTHY SKIN
TRAINING SET	425	420
VALIDATION SET	110	100
AUGMENTED TRAINING SET	2125	2100
AUGMENTED VALIDATION SET	550	500

Table 3.1: Images in Dataset

3.2 DATA PREPROCESSING

Data preprocessing is an essential stage in the evaluation of DFUs using deep learning. The following are some typical data processing techniques that can be applied with DFU datasets:

3.2.1. Data Cleaning

Data preprocessing is a critical step in deep learning, where raw data is converted into a format suitable for analysis. The major task of data processing is missing data handling, which occurs when data points are missing some observations or variables.

Incorrect handling of missing data can result in the final model having biassed or incorrect interpretations. Imputation, which involves substituting missing values with approximated values, is one of the frequently used techniques for dealing with missing data. There are a few ways to impute data that is missing, two of them being mean imputation and regression imputation.

Replacement of missing values to the mean of the available data for the variable is basically mean imputation. If, for example, the age of some individuals is missing from the data, missing values is replaced with the mean age of those individuals whose ages are known and eradicate all the missing values.

Regression imputation uses a regression model based on the values of other variables in the data set to predict missing values to get rid of all the missing values in the dataset and increase efficiency. For example, if some people's salary is missing from the data set, regression imputation would use a regression model to predict the missing salary values based on variables such as education level, experience, and job title.

Data preprocessing might include dealing with missing data in addition to handling outliers, standardizing or scale data, categorical variables are encoded, split the dataset into training set and test set for model building and evaluation. It is possible to increase the accuracy as well as dependability of deep learning models with proper data preprocessing.

3.2.2. Feature Selection

Feature selection is the process of selecting some of the most usefully significant features (or variables) from a dataset for analysis. In machine learning and deep learning, feature selection is an important step in data processing because it helps improve the accuracy and efficiency of models by reducing the number of inputs. There are various methods of selecting features, including:

- Filter methods: These approaches pick features based on statistical factors including information sharing between characteristics or correlation with the target variable. The Chi-Square test, Pearson correlation, and Information Gain are typical methods for filtering data.

- Wrapper methods: By training a model on a selected group of characteristics and reviewing its performance, these techniques choose features based on the capacity they have to make predictions. Frequent wrapping techniques are Forward Selection and Recursive Feature Elimination (RFE) .

- Embedded methods: These techniques identify characteristics for the model during the time it is being trained. Regularisation (such as L1 or L2 regularization), Decision Trees, and Neural Networks are often used embedded approaches.

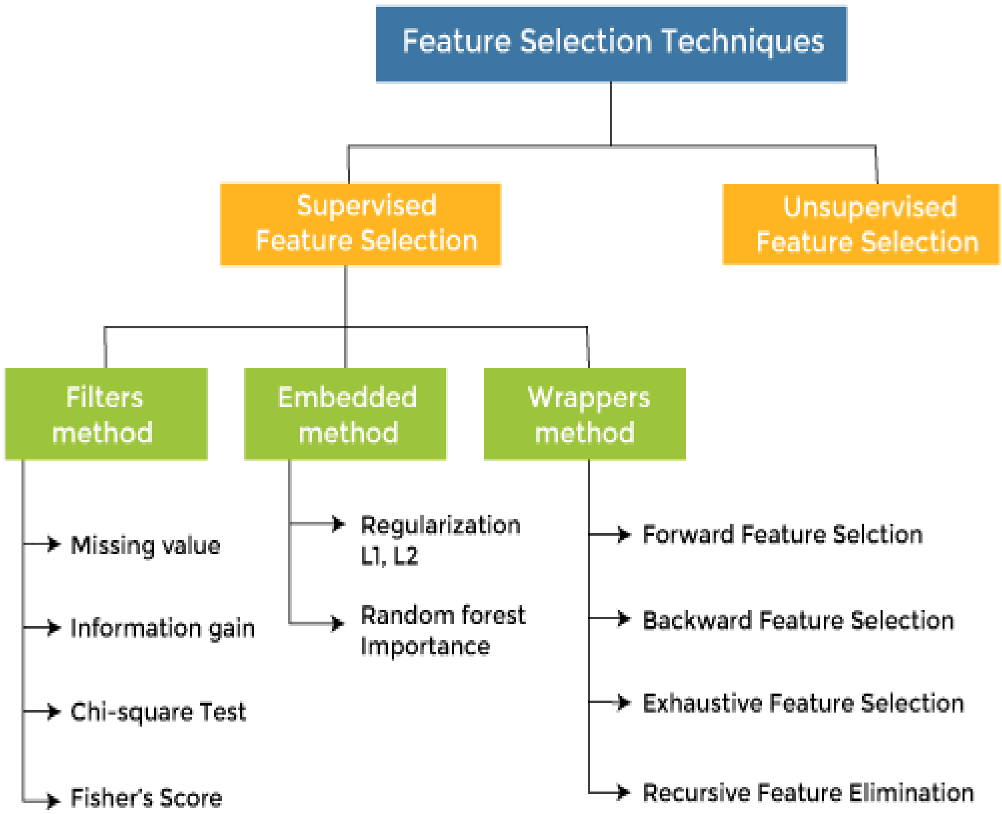

Fig 3.2 : Feature Selection for Data Pre-processing[22]

The strategy of feature selection that will be employed depends on the type of data and the aims of the research. The degree to which features are related to the goal variable, how they correlate with other traits, and how redundant they are should all be considered while analyzing features. The benefits of feature selection include minimizing the models' computational and memory needs, preventing overfitting by removing pointless or redundant features, and improving the models' interpretability by focusing on the most important features.

It is essential to keep in mind that feature selection can lead to information loss and that the features chosen may not necessarily represent the most suitable subset for the research's goals. As a result, the most crucial thing is to carefully evaluate the degree to which the algorithms perform while implementing various feature subsets that are then selected those that are most effective based on the assessment criteria.

3.2.3. Data Splitting

An essential phase for development of models is data splitting, which comprises splitting the dataset into two parts for training and validation. This method is essential for evaluating diabetic foot ulcers (DFUs), since it ensures that the model that results is reliable, precise, and able to be generalised to new data.

While train the model training dataset is used on the information that is accessible, the validation set is employed to execute hyperparameter modifications and evaluate the development of the model throughout the duration of training. The test set is implemented for assessing the extent to which the model will eventually function. The major purpose of the test set is to present an objective evaluation of the model's performance on new, untested data; as such, it should only be used rarely and subsequently in the model building process.

In order to guarantee whether the method is completed effectively, it is crucial to separate the data randomly and intentionally to prevent bias. In order to obtain results that are precise and trustworthy, the dataset should be evaluated for its significance and accuracy ahead of splitting. The appropriate ethical considerations should be made while interacting with information pertaining to patients, and plagiarism should be avoided through employing reliable sources.

Stratified techniques of sampling can be used to ensure that the number of instances of classes in each set is indicative of the full data set. It's also crucial to strike a balance between the size of each collection and the model's complexity. This increases the likelihood that the final model will effectively generalise to new data and not be biassed towards a certain type of data or subset of data.

3.2.4. Data Normalisation

Data scaling, more commonly referred to as data normalisation, is a deep learning technique used for DFU analysis to scale the numerical data in the dataset to a common range. This prevents the model from preferring features with more substantial magnitudes and guaranteeing that all features are assigned equal importance in the model. Data normalisation can be accomplished in numerous different methods, but some of the most common are the following:

Min-Max Scaling	By eliminating the lowest possible value from each data point and splitting by the range (i.e., the variance between both the highest and lowest values), the min-max scaling method does scaling between zero to one.	x_norm = (x - min(x)) / (max(x) – min(x))
Z-Score Normalization	By calculating the average value from each data point and dividing by the standard deviation of each point, this technique modifies the data to have a mean of zero along with one variance	x_norm = (x - mean(x)) / std(x)
Scaling to Unit Length	The Euclidean norm of the feature vector is divided by each data point in this method to scale the data to have unit length	x_norm = x / ||x||

Table 3.2 : Data Normalisation

Although features with large magnitudes can dominate loss of function during training, slow convergence or inadequate results may follow, data normalization is important. In the case of DFU analysis, features including wound area, wound depth, and ulceration length may have radically different scales, which may impact how well the model performs. By normalizing the data, we are able to make sure that each characteristic contributes equally to the model while further enhancing its performance. To be clearly stated only training data should undergo data normalization before it gets transmitted to the validation and test sets. By performing this, it can be guaranteed that every data point utilizes the same normalization parameters and that the model is not prejudiced towards any particular data collection.

3.2.5. Data Augmentation

Data augmentation is an approach employed in deep learning algorithms for diabetic foot ulcer (DFU) analysis to generate new examples by altering the dataset's current instances. This type of method tends to be used when the dataset is inadequate and insufficient to build an effective model. Numerous methods for augmenting data can be employed in DFU analysis, a few of them involve the following:

Rotation	This technique involves rotating the image by a certain angle (e.g., 90 degrees, 180 degrees, or 270 degrees) to create new images.
Flip	This technique involves flipping the image horizontally or vertically to create new images.
Translation	This technique involves shifting the image horizontally or vertically by a certain distance to create new images.
Zoom	This technique involves zooming in or out of the image to create new images.
Noise injection	This technique involves adding random noise to the image to create new images.
Brightness adjustment	This technique involves adjusting the brightness of the image to create new images.

Table 3.3 :Data Augmentation in DFUs

Data augmentation provides the potential to minimize overfitting, enhance the performance of models, and enhance model generalization. The model gets more robust and becomes more equipped for dealing with completely novel, not tested data through offering additional instances. It is additionally crucial for making certain the data augmentation techniques employed are suitable for the task at concern. For instance, some augmentation techniques, such as flipping and rotation, might be appropriate for DFU images, while others, such as brightness and contrast adjustments may not be useful.

Making sure that the augmented data effectively represents the original data is additionally important. This indicates that the improved data should retain all aspects of the original dataset's characteristics and not create any biases. In order to guarantee this, it is recommended to use controlled and random data augmentation during training. For DFU analysis, data augmentation is a fundamental deep learning method that can help prevent overfitting, which enhances model functionality, and broadens the model's range of applications. Implementing the appropriate data augmentation techniques is crucial, especially to make sure that the modified information accurately represents the original data.

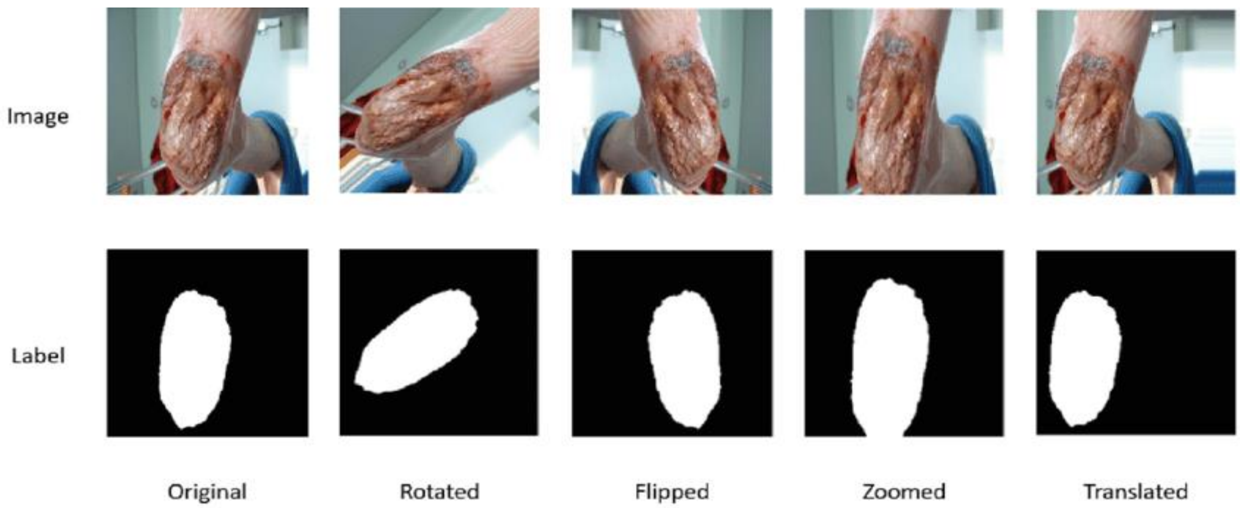

Fig 3.3: An Example of data augmentation on DFU image[23]

3.3 DEEP LEARNING TECHNIQUES FOR DFU ANALYSIS

Diabetic foot ulcers (DFU), which develop as a result of damage to the nerves and inadequate blood supply to the feet, represent some of the most serious adverse effects of diabetes. DFU is the most frequent reason for non-traumatic amputation of the lower limb globally, impacting 15–25% of the people with diabetes at some stage in their life span. Deep learning, an aspect of machine learning, has the potential to completely transform the manner in which DFU is controlled by assisting doctors recognise DFU as early and accurately as possible, predicting the probability of ulceration, and facilitating appropriate treatment. Several deep learning applications in DFU include the following:

- Early detection of DFU: After being trained on an immense number of images of healthy and sick foot, deep learning algorithms may effectively recognise early indications of DFU, such as discoloration of the skin, edoema, or calluses. Amputation risk can be minimized and DFU development can be slowed with detection at an early stage.

- Image analysis: After analyzing images of DFU, deep learning applications may generate automated evaluations concerning the size, depth, and extent of the wound. This can be employed by doctors in order to monitor the development of the ulcer and evaluate the extent to which the treatment is working.
- Risk prediction: Incorporating several kinds of clinical characteristics, including age, sex, the amount of time of diabetes, blood sugar levels, and foot sensation, deep learning models can forecast the probability of DFU development and the possibility of healing.
- Treatment optimization: Large datasets of clinical and therapeutic data can be analysed using deep learning algorithms to find trends and improve DFU therapies. For instance, deep learning models may determine which therapies are most successful for a specific DFU or patient demographics.

3.3.1 Convolutional Neural Networks

For evaluating DFU and other medical images, deep learning algorithms have proven valuable tools. CNN, an established deep learning architecture, is used for classification of diabetic foot ulcers (DFU) [24][25]. Because they are capable of efficiently identify and continue to retain particular characteristics and trends in an image, CNNs have been especially appropriate for image classification tasks such as identifying DFUs in medical images. CNNs have been employed in several kinds of medical image analysis applications because they have been developed to examine image data. These networks employ convolutional layers to extract from input images elements like colour, texture, and shape. Convolutional layers are followed by fully integrated layers, which are categorised according to the features collected. For a number of reasons, including their capacity to extract complex properties from massive amounts of data, CNNs are the ideal solution for DFU analyses that primarily use image data. In order to recognise and classify the different stages of wound healing, CNNs were used in the DFU study.

CNNs can also be used to extract features that reveal the biological processes behind the development and repair of DFUs. A CNN was used in one study, for instance, to assess DFU photos and identify traits that would indicate angiogenesis and inflammation, two critical processes involved in DFU healing. The ability of CNNs to predict the likelihood of wound healing using these features was subsequently proven. CNNs are capable of categorising DFUs as well as providing insights into the biological mechanics of DFU healing.

Integrating CNNs with DFU analysis may enhance patient outcomes through improving the precision and effectiveness of both diagnosis and therapy. For a number of reasons, including the capacity they have to gather information from enormous volumes of data and extract characteristics that provide insights into the fundamental biological mechanisms of DFU formation and healing, CNNs are a desirable tool in the field of DFU study.

In CNNs, different types of layers perform certain functions. The input layer sends the image to the convolutional layer, which processes it. A convolutional neural network (CNN) must have a convolution layer in order to recognise specific characteristics in an image. This layer moves a tiny matrix or kernel throughout the input image and performs element-wise multiplication between the two separate matrices employing a kernel-based approach. The results of each of these multiplications are incorporated together to generate the final output, which is used for the feature map for the convolutional layer that follows.

The CNN efficiently detects features utilising variable parameters. The kernel size influences the height and width in pixels that constitute the kernel, which impacts the capacity of the network to differentiate between specific features in an image. A 3x3 filter size is frequently employed in CNNs. The stride parameter determines the extent to which the kernel slides over the input data following each computation, enabling the kernel to move forward pixels at a time and down-sample the input data in order to generate smaller feature maps. Padding is a crucial option that allows for the incorporation of extra zeros at the edge of input data, ensuring that the feature map size remains unchanged even when using larger kernels for convolution.

The filter count is another factor that influences the number of kernels used in the convolutional layer. By applying extra layer kernels, the network may be enabled to differentiate between a wide variety of different properties. Because every single kernel is able to concentrate on emphasising a distinct feature, feature identification becomes increasingly accurate and specific. Filters are placed on input image in the first layer of a CNN with the objective of extraction of vital features like edges, corners, and other visual patterns. This layer has four hyperparameters:

- K basically denotes number of filters, will determine how many feature maps are generated throughout the convolution process. Each filter learns to recognise a certain feature in the input image, and the number of filters used affects how many different features are learned. However, when K increases, the network's parameter count does as well, which could lead to overfitting.
- The size of the filters F determines the area of the input image that each filter is going to concentrate on, or its field of view. more expansive filters add more parameters to the network despite the fact they have a broader receptive area and can capture more complex properties.
- Step S determines how much the filter is moved over the input image. A larger step results in a smaller output size and reduces the computational cost of the convolution operation. However, a smaller step can lead to more accurate feature detection, especially for small or detailed features.
- The zero-padding P surrounds the input image with a border of zeros prior to performing the convolution method. This is helpful for maintaining the general structure of the input image because spatiel dimensions of the image are preserved. Without padding, the size of output of the convolution operation will be smaller than the input size, and subsequent convolutional layers will cause the output size to continuously decrease, potentially resulting in the loss of important information. However, if used excessively, padding can lead to overfitting and increase the computational cost of the convolution process.

For an input images where width, height, and depth are W, H, and D respectily the pooling layer will result in the matrix of dimension shown below:

$$Wc = (W-F+2P/S) + 1 \tag{1}$$

$$Hc = (H-F+2P/S) + 1 \tag{2}$$

$$Dc = K \tag{3}$$

The activation layer receives the output of the convolutional layer and utilises it to apply a nonlinear function. By making the network non-linear with the aid of the activation function, the network can learn complex correlations between the input and output. The most commonly used activation function are namely Sigmoid, ReLU, and TanH. Spatial downsampling is a process performed by the pooling layer employing the resultant data of the activation layer for reduction in the size of the feature maps. Reduction of the input's spatial dimensionality, increasing the efficiency of computing while reducing the risk of overfitting is done by pooling layer.

Execution is done after the result of pooling layer is passed to fully connected layer, which executes classification. Classification task are performed by output neurons. The output neurons calculate the probability that each class in the data being processed occurs using a Softmax function. To extract significant data from the input image and categorise it using those features, CNNs many layers and functions work together. CNNs are highly useful for DFU analysis because the images include significant visual characteristics that may demonstrate the extent of the wound and the process of healing. Clinicians can better identify and treat patients by employing CNNs to analyse DFUs as they provide them with a greater degree of accuracy and thorough comprehension of the wound.

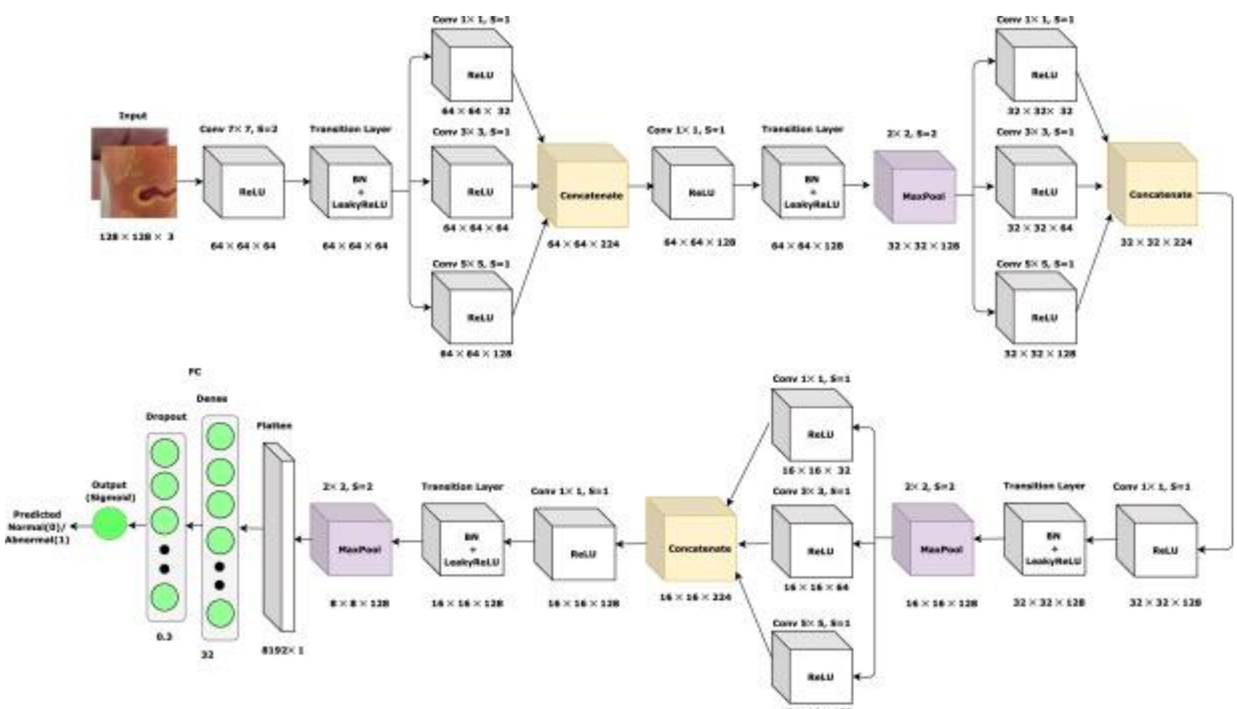

Fig 3.4: CNN Architecture [22]

3.3.2 Recurrent Neural Networks

Recurrent Neural Networks are a useful tool for analyzing time-series data, such as information from regular foot exams or continuous glucose monitoring (CGM). By identifying patterns and trends in the data, RNNs can be used to predict the likelihood of DFUs developing. RNNs are particularly useful for sequence modeling and have been evaluated for predicting DFUs [26].When training an RNN model on a DFU dataset, several steps are involved. The process includes:

1. Preparing the data: This step involves preparing the DFU dataset by cleaning, normalizing, and engineering features. The dataset should then be split into training, validation, and test sets, with the largest set being the training set.
2. Choosing a model: An appropriate RNN model architecture, such as LSTM or GRU, should be selected based on the specific requirements of the DFU prediction or management task.
3. Initialization: The RNN model should be initialized with random weights, and the initial state of the hidden layer should be set to zero.
4. Forward pass: The input sequence is passed into the RNN model, and the model produces a prediction for each time step. The hidden layer is updated at each time step based on the current input and the previous hidden state.
5. Backward pass: In this step, the model's output is compared to the expected output at each time step, and the error is propagated back through the network to adjust the weights.
6. Optimization: During the training process, the primary goal is to minimize the difference between the model's output and the expected output.
7. Model evaluation: After the completion of training, the model is evaluated on a separate test set to assess its performance.

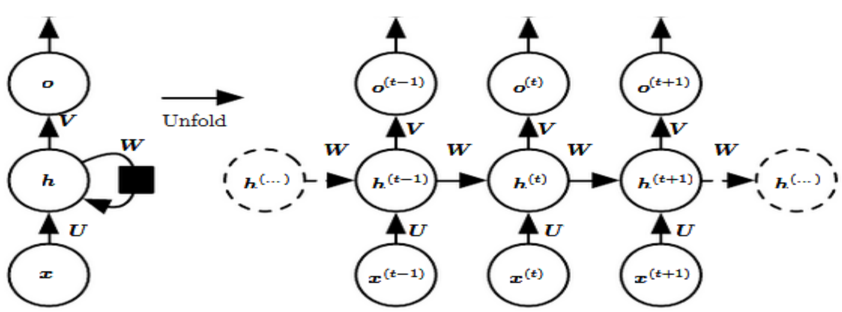

Fig 3.5: Recurrent neural network graph [27]

There are four main types of Recurrent Neural Networks:

1. One-to-One RNNs: These RNNs have the simplest structure and only have one input and one output, with fixed input and output sizes. It is frequently used for jobs involving image classification.
2. One-to-Many: With a single input and a predetermined input size, this type generates a series of output data. It is frequently employed in activities like image captioning and music generation.
3. Many-to-One: This kind is employed when a series of input units must produce just one output. It is frequently employed in sentiment analysis and requires a series of inputs and yields a definite result.
4. Many-to-Many: With this kind, a series of output data are produced from a series of input units.

Overall, these studies suggest that RNNs may have the potential to be a valuable tool in predicting and managing DFUs in patients with diabetes. However, further research is needed to validate these findings and to develop more accurate and reliable RNN-based models for DFU prediction and management. While Recurrent Neural Networks (RNNs) have shown promise in predicting and managing diabetic foot ulcers (DFUs), there are also some potential drawbacks to their use. Some of the drawbacks are:

- Data quality: RNNs require high-quality and large datasets to be trained effectively. In the case of DFUs, the availability of high-quality data can be a challenge. There may be issues with data completeness, accuracy, and consistency, which can limit the effectiveness of RNNs.
- Generalizability: RNNs trained on a specific dataset may not generalize well to new datasets or patient populations. This is a common challenge in machine learning and can limit the utility of RNNs for DFU prediction and management.
- Interpretability: RNNs can be difficult to interpret, making it challenging to understand why certain predictions or recommendations are made. This lack of interpretability can limit the adoption of RNNs in clinical settings, where transparency and interpretability are critical.

- Computational complexity: RNNs can be computationally complex and require significant computing resources to train and use effectively. This can be a limitation for smaller clinics or hospitals that may not have access to high-performance computing resources.

While RNNs have shown promise for DFU prediction and management, their effectiveness may be limited by data quality, generalizability, interpretability, and computational complexity. Further research is needed to address these challenges and to determine the optimal role of RNNs in DFU prediction and management.

3.3.3 Long Short-Term Memory (LSTM) Networks

LSTMs are a type of RNN that are particularly suited to analyzing temporal data [28], which is useful in identifying patterns and trends in DFU images over time. This makes them useful for forecasting the risk of diabetic foot ulcers based on a patient's medical history and other risk factors. LSTM can be applied to time series data, such as those generated by DFU. Here is how LSTMs work in the context of DFU:

1. Data acquisition and preprocessing: DFU time series data, such as daily measurements of ulcer size or patient vital signs, are collected and preprocessed to ensure that they are standardized and suitable for use with an LSTM.
2. Training the LSTM: The internal memory cells of an LSTM network, a form of RNN developed for processing sequence data, are used to train the network. Over longer times, these cells have the capacity to selectively keep or delete information. These gates manage the values of the cell's internal state and output to regulate the flow of information through the cell. A series of input vectors x1, x2,..., xt are sent to the LSTM, which then generates a series of output vectors h1, h2,..., ht. The LSTM changes its internal state at each time step t using the most recent input xt and the prior state ht1.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

$$f_t = \sigma(W\_f x_t + U\_f h_{t-1} + b\_f) \tag{5}$$

$$g_t = \tanh(W\_g x_t + U\_g h_{t-1} + b\_g) \tag{6}$$

$$ot = \sigma(W\_o x_t + U\_o h_{t-1} + b\_o) \tag{7}$$

$$c_t = f_t * c_{t-1} + i_t * g_t \tag{8}$$

$$h_t = o_t * \tanh(c_t) \tag{9}$$

To prepare an RNN type called an LSTM network for RNN with internal memory cells that selectively retain or forget information over time is known as an LSTM network. It is made to handle sequence data. Input, forget, and output gates make up the fundamental LSTM cell design, and they control how information moves through the cell. The LSTM uses the prior state and the current input to update its internal state at each time step. The candidate value, updated internal state, input, forget, and output gates are specified using a variety of weight matrices, bias vectors, and activation functions including sigmoid and hyperbolic tangent.

In the context of DFU, the LSTM is trained on preprocessed time series data (e.g., sensor readings) to predict the progression or severity of the ulcer. The input sequence may include information about the patient's medical history, demographic information, and wound care practices in addition to the time series data. The LSTM can be trained using stochastic gradient descent (SGD) or other optimization algorithms, with hyperparameters such as learning rate and regularization strength adjusted through cross-validation.

3. Prediction and detection: Once the LSTM is trained, it can be used to predict future DFU measurements, such as ulcer size or the likelihood of infection. The LSTM can also be used to detect anomalies or changes in the time series data that may indicate a worsening of the DFU.
4. Treatment planning: The predictions and detections generated by the LSTM can be used to aid in treatment planning. For example, if the LSTM predicts a worsening of the DFU, the treatment plan can be adjusted accordingly to prevent further complications.
5. Validation and evaluation: The effectiveness and safety of the LSTM-generated predictions and detections must be validated and evaluated in clinical settings. This may involve comparing the LSTM-generated predictions to actual outcomes or evaluating the accuracy of the LSTM in detecting changes in the time series data.

LSTMs and RNNs are both types of neural networks that can be used for analyzing time series data such as those generated by diabetic foot ulcers (DFU). While both LSTMs and RNNs are effective in modeling temporal dependencies in the data, LSTMs generally outperform RNNs in capturing long-term dependencies due to their ability to selectively remember or forget information over time. In particular, LSTMs are designed to avoid the vanishing gradient problem, which can occur when back propagating errors through a long sequence of time steps. This makes LSTMs better suited for modeling time series data with long-term dependencies, such as DFU data.

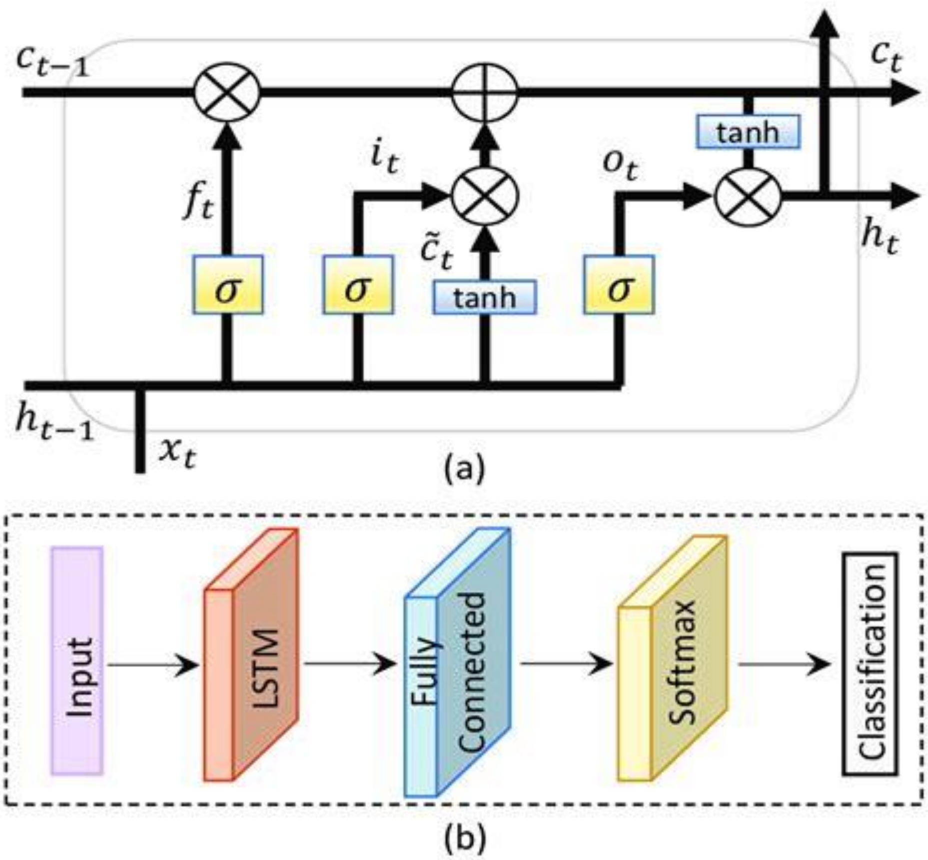

Fig 3.6 : LSTM Architecture [29]

Additionally, LSTMs are capable of storing and retrieving information from the previous time steps selectively, which means that they can learn to recognize important patterns in the data and filter out irrelevant noise. This ability makes LSTMs more effective at predicting and detecting changes in DFU time series data and aiding in treatment planning. While RNNs are useful for analyzing sequential data, LSTMs are generally better suited for modeling time series data with long-term dependencies, such as those generated by DFU. However, the choice of neural network architecture ultimately depends on the specific needs and characteristics of the data being analyzed.

3.3.4 Variational Autoencoders

VAEs are a type of deep generative model that can be used for unsupervised feature learning. By learning how to generate realistic synthetic images of diabetic foot ulcers [30], VAEs can help identify key features that are most predictive of ulcer formation. VAEs can be applied to diabetic foot ulcers (DFU) to generate synthetic data and aid in diagnosis and treatment. Here is how VAEs work in the context of DFU:

1. Data acquisition and preprocessing: DFU images are collected and preprocessed to ensure that they are standardized and suitable for use with a VAE. This may involve resizing the images, adjusting the color balance, and normalizing the intensity levels.

   2. Training the VAE: The VAE is a type of generative model that undergoes a two-step training process: encoding and decoding. The encoding process reduces the high-dimensional input data into a lower-dimensional latent space, while the decoding process maps the latent space back to the original high-dimensional space. The VAE is trained using a loss function that includes two components: a reconstruction loss and a regularization loss. The reconstruction loss measures the discrepancy between the input image x and the output image x', while the regularization loss encourages the latent space representation to have specific desirable characteristics, such as being normally distributed. The mathematical expression for the reconstruction loss is:

$$\text{reconstruction\_loss} = -\log(p(x|x', \sigma)) \tag{10}$$

   where $p(x|x', \sigma)$ is the probability of generating x given x' and σ, and the sum is taken over all pixels in the image. The regularization term in VAE is measured using the KL divergence:

$$\text{KL\_divergence} = -0.5 * \sum(1 + \log(\sigma^2) - \mu^2 - \sigma^2) \tag{11}$$

where the sum is taken over all elements in the latent vector z. The total loss function for the VAE is then:

$$\text{loss} = \text{reconstruction\_loss} + \text{KL\_divergence} \quad (12)$$

The VAE's encoder is comprised of multiple convolutional layers, which reduce the input image into a lower-dimensional latent space. The encoder typically concludes with a layer that outputs the mean and standard deviation of the distribution from which the latent space is sampled. These parameters are then utilized to sample points from the latent space via the reparameterization trick. The decoder, on the other hand, is made up of multiple deconvolutional layers that reconstruct the sampled points from the latent space back into the high-dimensional space. The decoder typically ends with a layer that produces the output image.

3. Generating synthetic data: After training, the VAE is capable of mapping the input space to the lower-dimensional latent space and the latent space back to the input space. Using the decoder of the VAE, a new image can be generated by inputting a latent vector z.

   To generate synthetic DFU images, we can sample from the latent space by choosing a random vector z from a standard normal distribution (i.e., $z \sim N(0,I)$). This can be done efficiently by first generating a vector of random numbers from a standard normal distribution and then passing it through the decoder:

$$z = N(0,I) \quad (12)$$

$$x = \text{decoder}(z) \quad (13)$$

The resulting synthetic images can be used to augment the training data, by adding them to the original dataset to create a larger and more diverse set of images. Additionally, synthetic images can also be used to assist in diagnosis, by generating new images for rare or difficult-to-diagnose cases. They can also aid in treatment planning by allowing doctors to simulate the progression of DFUs over time and evaluate the effectiveness of different treatments.

4. Validation and evaluation: The effectiveness and safety of the VAE-generated images must be validated and evaluated in clinical settings. This may involve comparing the VAE-

generated images to real DFU images, evaluating the accuracy of diagnosis based on the synthetic images, or testing the efficacy of treatment plans based on the synthetic images.

VAEs offer a promising approach for generating synthetic DFU data that can aid in diagnosis and treatment. It has some advantages over GANs when applied to diabetic foot ulcers (DFU):

1. Ability to generate diverse and realistic images: VAEs are known to generate a more diverse set of images than GANs, which can be useful for generating synthetic data for DFU.
2. Ability to encode and decode data: VAEs can encode and decode data, which means they can generate new images from a reduced representation of the data. This can be useful for generating synthetic data with reduced dimensionality.
3. Easy to interpret: VAEs are easier to interpret than GANs because they provide a reconstruction of the input data, making it possible to identify errors or biases in the model.
4. Reduced risk of overfitting: VAEs are less likely to overfit to the training data than GANs because they include a regularization term in the loss function.
5. Clinical validation: VAEs have been used in several healthcare applications and have been validated in clinical settings.

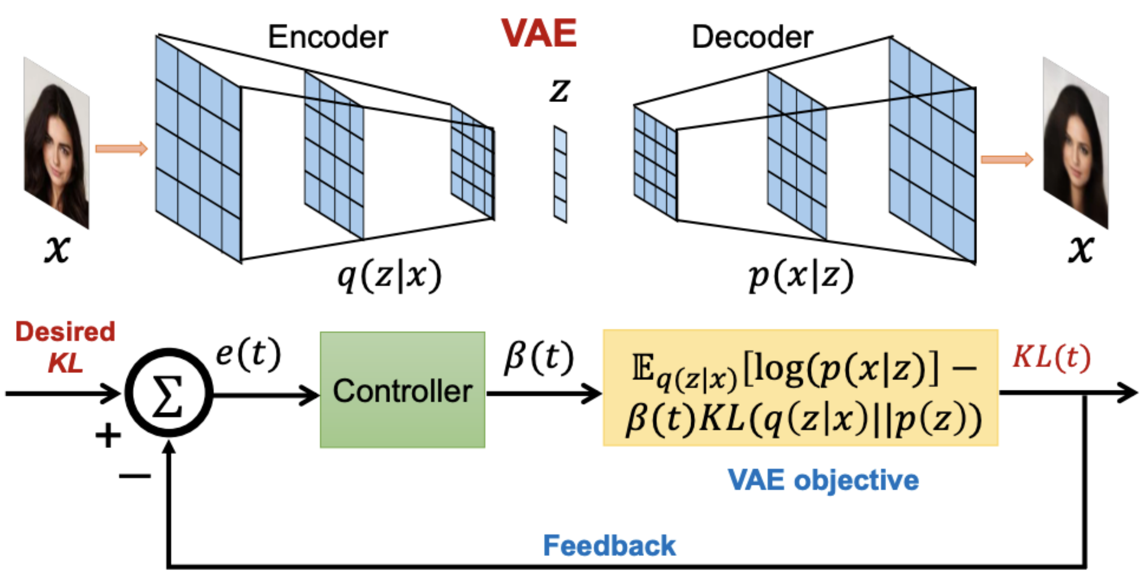

Fig 3.7 : VAE Algorithm [31]

However, like GANs, VAEs also have some similar limitations when applied to DFU. One potential drawback of using machine learning approaches, such as VAE, for diabetic foot ulcer detection and management is the need for large and diverse datasets. Collecting and annotating such datasets can be time-consuming and costly, which may limit the applicability of these approaches in clinical settings. Additionally, machine learning models may be less reliable when applied to patients with less common or atypical presentations of diabetic foot ulcers. As with any medical application of machine learning, it is important to continue to validate and refine these approaches to ensure their efficacy and safety.

3.3.5 Generative Adversarial Networks

Deep learning models called Generative Adversarial Networks [32] are used to create images. GANs may be used to create fake DFU pictures that seem like actual ones in the context of DFU. A generator and a discriminator neural network make up the fundamental architecture of a GAN. A random noise vector is fed into the generator network, which outputs a synthetic picture. A genuine or artificial picture is sent into the discriminator network, which attempts to distinguish between the two. The two networks are trained in tandem in a way that resembles a game, with the generator attempting to produce increasingly convincing pictures in an effort to trick the discriminator and the discriminator attempting to accurately distinguish between genuine and artificial images.

The phrases generator loss and discriminator loss are combined to form the loss function used in GANs. The discriminator loss measures how effectively the discriminator can tell actual images from fake ones, whereas the generator loss measures how well the generator can produce realistic images. The loss function in GANs is described as follows:

$$\text{Loss} = L\_D + L\_G \tag{14}$$

Where the discriminator loss is denoted by L_D and the generator loss by L_G. The discriminator loss is defined as:

$$L\_D = -E[\log(D(x))] - E[\log(1 - D(G(z)))] \tag{15}$$

where "D(x)" and "D(G(z)") are the probabilities assigned by the discriminator to the real and synthetic pictures, respectively, and "x" is a real image taken from the training data and "G(z)" is a synthetic image created by the generator from a random noise vector. How well a discriminator can discern between actual and artificial pictures is measured by its discriminator loss. The definition of the generator loss is

$$L\_G = -E[\log(D(G(z)))] \tag{16}$$

The generator loss measures how well the generator is able to create synthetic images that can fool the discriminator. The generator aims to minimize this loss by creating images that are similar to real images from the training data. During training, the generator and discriminator are updated iteratively by minimizing the loss function using gradient descent. The goal is to find a Nash equilibrium, where the generator produces synthetic images that are indistinguishable from real images and the discriminator assigns equal probabilities to real and synthetic images.

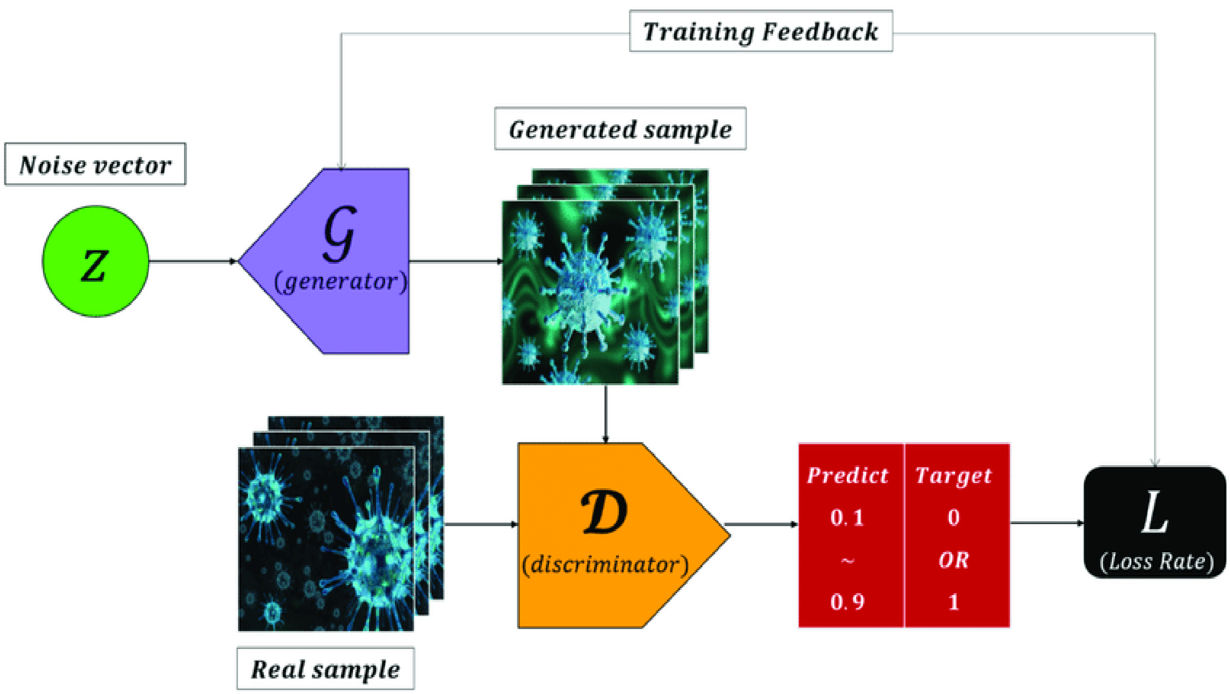

Fig 3.8 : GANs Model [32]

The GAN architecture used for DFU image generation may consist of several convolutional layers in both the generator and discriminator networks. The layers present and the total neurons in every layer can vary depending on the specific implementation.

Additionally, techniques such as batch normalization and dropout may be used to improve the stability and performance of the model.

DFU occurs due to neuropathy (nerve damage) and poor blood flow. DFU can lead to infection and, in severe cases, amputation. Early detection and treatment of DFU are critical to preventing complications.

One possible application of GANs in DFU is to generate synthetic images of foot ulcers to aid in the diagnosis and treatment of the condition. GAN is trained on a large dataset of DFU images, the model could learn to generate realistic ulcers that could be used to train healthcare professionals or assist in the development of automated diagnosis systems.

Another potential application of GANs in DFU is to generate realistic 3D models of foot ulcers for use in surgical planning and wound management. By incorporating patient-specific data such as MRI or CT scans, GANs could be trained to generate personalized 3D models of foot ulcers that could be used to plan and optimize treatment. However, it is worth noting that the use of GANs in healthcare requires careful consideration of ethical and privacy concerns, particularly around the use of patient data. Additionally, the effectiveness and safety of GAN-generated models must be rigorously evaluated before they are used in clinical settings. It have several potential drawbacks when applied to DFU:

1. Limited availability of high-quality data: GANs require large amounts of high-quality data to generate accurate and realistic images. However, the availability of such data for DFU may be limited, and the quality of the data may vary due to differences in imaging technologies and protocols.

2. Ethical and privacy concerns: The use of GANs in healthcare requires careful consideration of ethical and privacy concerns, particularly around the use of patient data. There is a risk of unintended disclosure of sensitive patient information, which could have legal and ethical implications.

3. Limited interpretability: GANs are often referred to as "black box" models because they lack interpretability. It can be challenging to understand how a GAN generates a particular image, which could make it difficult to identify errors or biases in the model.

4. Lack of generalizability: GANs are known to suffer from a phenomenon called "mode collapse," where they generate similar images repeatedly, rather than generating a diverse set of images. This could limit the generalizability of the model, making it less useful for real-world applications.

5. Limited clinical validation: While GANs show promise in generating synthetic data, there is a need for rigorous clinical validation to determine the effectiveness and safety of GAN-generated models in real-world clinical settings.

3.3.6 Transfer Learning

Applying knowledge from one task or dataset to another that is closely similar is known as transfer learning, which is a deep learning strategy. In the context of DFU analysis, transfer learning is a useful technique for applying recently learnt models to new DFU datasets. A pre-trained CNN that was previously trained on a huge dataset of natural images is an example of a feature extractor for DFU photos. The newly collected features could be incorporated into a new classification model, which would then be trained on the DFU dataset.

The DFU dataset can utilise the knowledge acquired during the pre-training phase thanks to transfer learning, possibly enhancing performance and quickening model convergence. Transfer learning can also reduce the amount of data needed to train the model, which may be helpful when the DFU dataset is limited or has other limitations.

It's crucial to keep in mind that not all pre-trained models may be appropriate for transfer learning to DFU analysis since the characteristics the model has learnt may not be pertinent or effective for the current job. Deep learning now employs a number of transfer learning techniques, some of which are frequently applied to a variety of computer vision applications, including DFU analysis. Some of the most popular transfer learning strategies are listed below:

1. Pre-trained models as feature extractors: In this method, pre-trained models—like CNNs—are used to extract features from new datasets. A new model, such as an SVM

or a fully connected neural network, may be fed the learnt features and then tweaked using the new dataset.

2. Fine-tuning previously trained models: In this method, portions of the model's weights are updated while the model is being trained on a fresh dataset. The plan is to start with the pre-trained model and then modify it to fit the new dataset. When the new dataset resembles the first pre-training dataset, this strategy can be especially beneficial.

3. Domain adaptation: This method is applied when there are considerable variances in between target domains and the source. The goal is to alter the pre-trained model's architecture or fine-tune particular model layers in order to make it more suitable for the new domain.

4. Multi-task learning includes training a single model on a number of connected tasks, allowing the information gained from one task to be used to the others. When there is a finite quantity of data accessible for each activity, this can be especially helpful.

5. Ensemble methods: In order to perform better on a new job, this method combines many pre-trained models or models trained on various datasets. When the individual models have complementary strengths and limitations, this can be more beneficial.

These strategies are not mutually exclusive, and it is possible to combine many strategies to increase performance on a particular job. The precise job and the available resources, especially the quantity and calibre of the data, determine the technique to be used. All of the aforementioned transfer learning strategies can be applied to DFU analysis, according on the particular needs and resources available. For instance, feature extractors for DFU photos may be pre-trained models like VGG, ResNet, or Inception. The extracted features could then be put into a new model like a fully connected neural network or SVM. This approach can be particularly useful if the available DFU dataset is small, as it can leverage the knowledge learned from pre-training on large-scale datasets.

Especially if the model was trained pre-trained on a similar dataset, such as medical photos, fine-tuning pre-trained models can be a potent strategy in DFU research. It may adjust the pre-trained model to the unique purpose of DFU analysis by fine-tuning certain layers on the DFU dataset, thereby enhancing performance and lowering the requirement for vast volumes of data.

If there are substantial variations between the DFU dataset and the pre-training dataset, domain adaptation might be helpful in DFU analysis. The pre-trained model may perform better on the job if it is modified to take into account the unique features of the DFU dataset.

When a task has several related subtasks, such ulcer classification, wound segmentation, and healing prediction, multi-task learning might be helpful in DFU analysis. It may be possible to enhance performance and decrease the quantity of data required for each subtask by training a single model on a variety of related tasks.

When there are many pre-trained models or models trained on distinct datasets, ensemble techniques might be helpful in DFU analysis. It may be possible to increase performance and lower the danger of overfitting to the unique DFU dataset by combining the advantages of separate models. The VGG16 model is one instance of a transfer learning model for a diabetic foot ulcer (DFU).

1. **ResNet-50** - A member of the ResNet family of models, which are renowned for their capacity to train extremely deep neural networks with enhanced performance, ResNet-50 is a deep CNN. Convolutional, pooling, and fully linked layers are among the 50 layers in ResNet-50. An picture with dimensions of 224 x 224 x 3 serves as the input to the ResNet-50 model. 64 filters are present in convolution layer of size 7x7 and a stride of 2 makes up the model's first layer. Followed by max pooling layer with stride equal to 2 and a pool size 3x3.

   The following component of the ResNet-50 model is made up of four residual blocks, each of which has three convolutional layers and filters with the corresponding sizes of 64, 64, and 256. Filter sizes for the first two convolutional layers are 1x1, and for the third layer they are 3x3. Additionally, each residual block has a shortcut link that allows gradients to transit across the network without going through the convolutional layers.

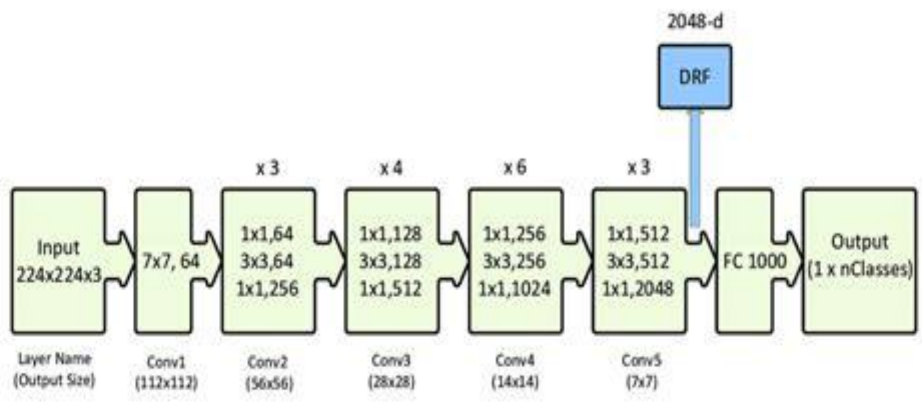

Fig 3.9 : ResNet [33]

Following the residual blocks, there are three further convolutional layers with 512 3x3 sized filters each. The output from all the feature map is averaged by a global average pooling layer that comes after these convolutional layers. the fully connected layer is constructed by one thousand neurons, or all the classes in the original ImageNet dataset, which the output of the global average pooling layer is finally fed. Then new fully connected layer where number of classes and number of neurons are equal in the new dataset, in this example diabetic foot ulcer, is substituted for the existing fully connected layer in transfer learning applications.

The new fully connected layer may learn the task-specific features from the new dataset while still taking use of the weights that were pre-trained in the convolutional layers since the convolutional layers weights remain frozen throughout training. Compared to training a deep convolutional neural network from scratch, this method can result in faster convergence and increased accuracy.

2. **InceptionV3** - The InceptionV3 model, a deep CNN that was initially trained on the ImageNet dataset, has demonstrated excellent performance in a number of computer vision applications. The model architecture contains a number of inception modules, each of which consists of a number of simultaneous convolutional and pooling layers. These modules are in charge of teaching the input images' characteristics. A fully linked layer makes the ultimate classification determination in the InceptionV3 model.

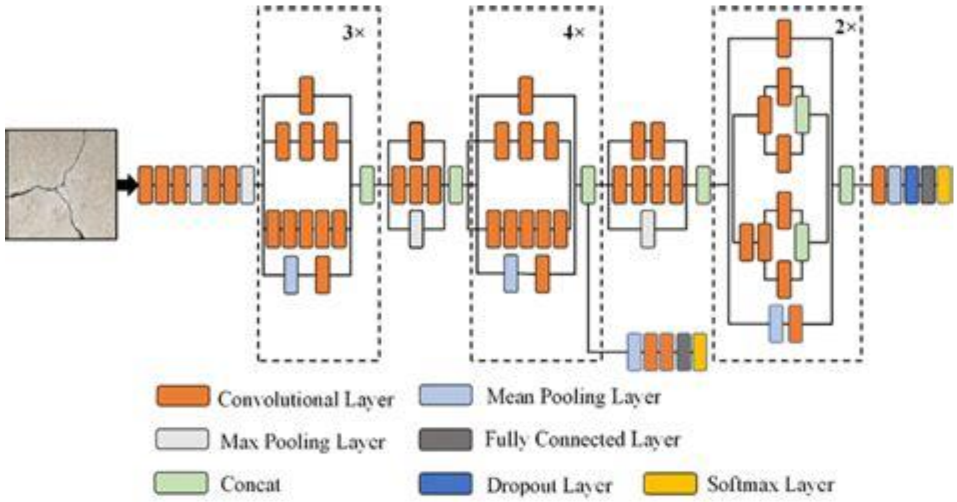

Fig 3.10 : Inception V3 [34]

The final fully connected layer is swapped out with a new layer that has the same number of output neurons as the number of classes in the DFU dataset in order to modify the InceptionV3 model for DFU classification. The new layer is randomly initialised and only its weights are changed during the fine-tuning process, similar to the VGG16 model. However, because the InceptionV3 model has several parallel routes in its inception modules, its design is more intricate than VGG16's. Due of its intricacy, the InceptionV3 model can collect characteristics at various sizes and resolutions, improving its ability to handle difficult computer vision tasks.

Even though the VGG16 architecture is renowned for being uniform and basic, it could excel at situations where the DFU features are straightforward and straightforward. For more challenging DFU classification tasks, however, the InceptionV3 model's capacity to collect features at various sizes and resolutions may be useful. The particulars of the dataset and the difficulty of the classification assignment at hand ultimately determine which model should be used for DFU classification.

3. **Xception** - This model is based on the idea of extreme Inception modules, where instead of traditional convolutions, depthwise separable convolutions are used. The depthwise separable convolutions separate the spatial and channel-wise filtering, allowing the network to capture more fine-grained features.

Like other pre-trained models, the last fully connected layer of Xception is replaced with a new layer tailored to the specific DFU classification task. During training, only the weights of the new layer are updated, while the pre-trained weights of the other layers are frozen. This fine-tuning process enables the model to leverage the learned representations from the original Xception architecture.

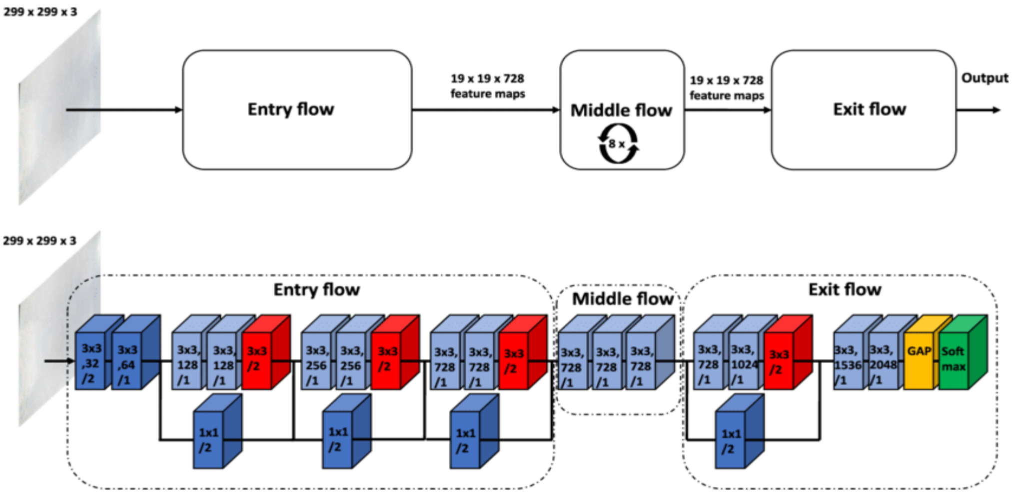

Fig 3.11 : Xception [35]

By using Xception as a pre-trained model and fine-tuning it on the DFU dataset, the model can benefit from its exceptional performance on various computer vision challenges. The transfer learning approach of Xception allows the model to extract relevant and discriminative features from the DFU images, leading to improved classification accuracy even with limited data.

It is a powerful CNN architecture that incorporates depthwise separable convolutions, resulting in efficient and effective feature extraction. Fine-tuning Xception on the DFU dataset harnesses the learned representations of the model, enhancing the performance of DFU classification tasks by leveraging the general knowledge captured by the pre-trained network.

4. **VGG16** - In computer vision, the VGG16 model is a well-liked and often utilised convolutional neural network (CNN) architecture. 13 convolutional layers, with 3x3 and filter sizes is 1x1, and fully linked layers are 3, make up the VGG16 architecture. The fully

connected layers are in charge of reaching the final classification determination, while the convolutional layers are in charge of learning the characteristics of the input pictures.

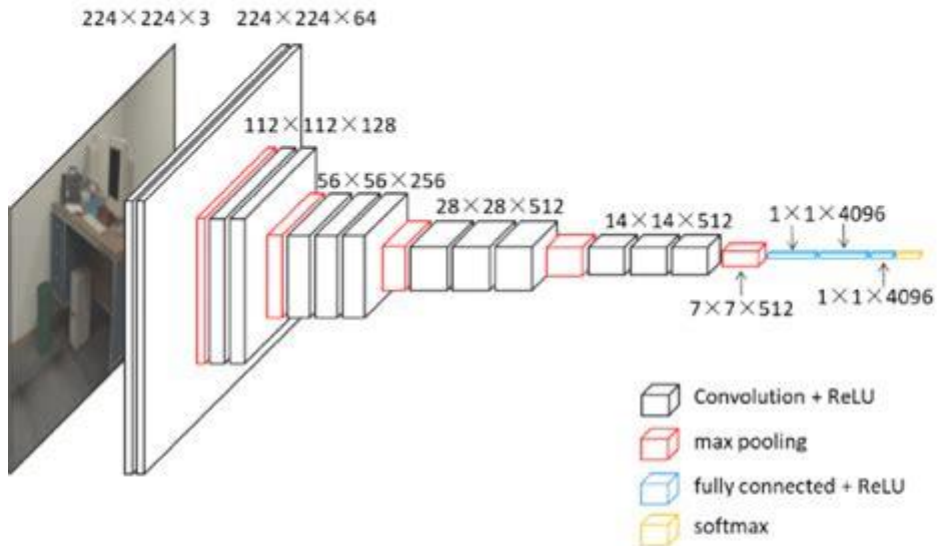

Fig 3.12: VGG16 Architecture [36]

The last fully connected layer, which was initially created for ImageNet classification, is swapped out for a new layer that has the same number of output neurons as the number of classes in the DFU dataset in order to modify the VGG16 model for DFU classification. Only the new layer's weights are changed during training, and the new layer's initialization is random. Because it includes fine-tuning the previously trained model for a new job, this method is known as tuning. Convolutional layer pre-trained weights are "frozen," which means they aren't changed when you train. This is so that the convolutional layers may transmit their knowledge of applicable and general features from the ImageNet dataset to the DFU classification job.

The pre-trained VGG16 model may use the high-level features discovered from the ImageNet dataset and apply them to the DFU classification problem by fine-tuning the model on the DFU dataset. When the size of the DFU dataset is constrained, this method can help the model perform better.

5. **MobileNetV3Small** - It is a popular CNN architecture known for how efficiently and effectively it does image classification. The MobileNetV3Small architecture consists of lightweight depthwise separable convolutions, which help reduce computational

complexity while still capturing important image features. It also incorporates advanced techniques such as linear bottlenecks and squeeze-and-excitation modules, which enhance the model's representational capacity and improve its ability to capture intricate patterns in the data.

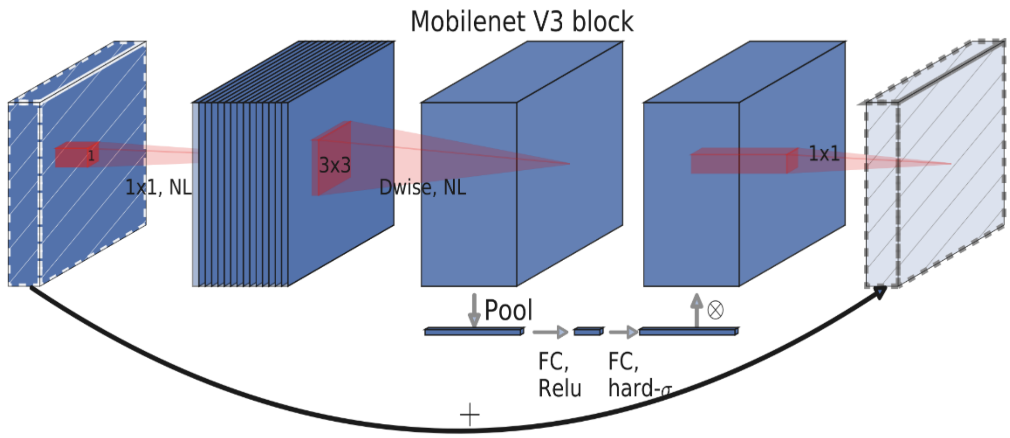

Fig 3.13: MobileNetV3Small Architecture [37]

By leveraging the MobileNetV3Small architecture and fine-tuning it on the DFU dataset, the model can take advantage of its lightweight design and efficient computations to achieve accurate classification results while minimizing resource requirements. This makes MobileNetV3Small a suitable choice for applications with limited computational resources or when deploying models on mobile or edge devices.

6. **ResNet101V2** - It is an advanced variant of the ResNet (Residual Network) architecture. The ResNet101V2 architecture consists of 101 layers. It employs residual blocks with bottleneck structures, which help reduce the computational cost while maintaining model accuracy. These blocks contain 1x1, 3x3, and 1x1 convolutions, allowing the model to capture both local and global features.

   In terms of parameters, ResNet101V2 has approximately 44 million parameters. This parameter count refers to the total number of learnable weights and biases in the model, which contribute to its representational power and ability to learn complex features from input images.

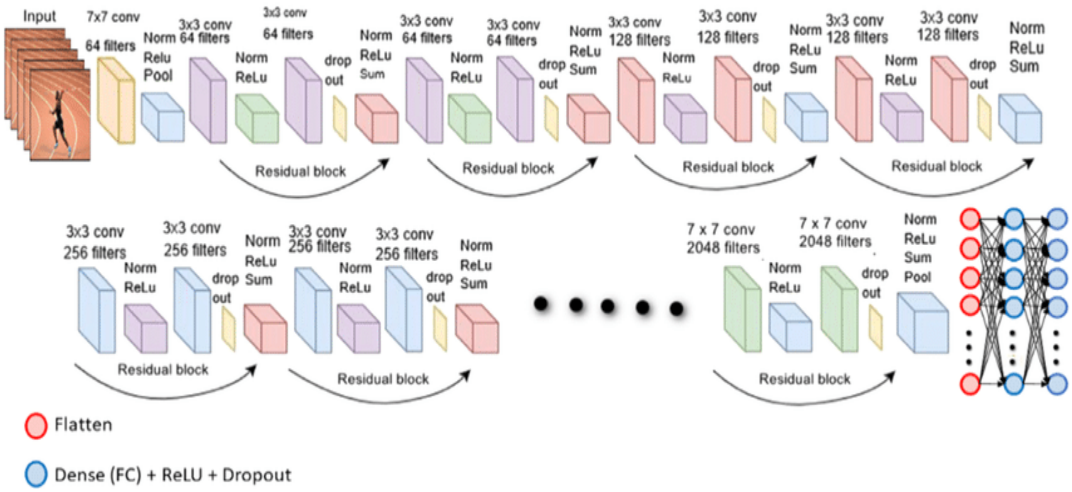

Fig 3.14: ResNet101V2 Architecture [38]

For DFU classification, the last fully connected layer of ResNet101V2 is modified to have the same number of output neurons as the classes in the DFU dataset. During training, only the weights of this final layer are updated, while the pre-trained weights of the rest of the layers are frozen. This fine-tuning approach allows the model to adapt its learned features to the specific DFU classification task.

By utilizing the ResNet101V2 architecture and fine-tuning it on the DFU dataset, the model can benefit from its deep structure and powerful representation capabilities. The large number of parameters enables the model to capture intricate patterns and variations in DFU images, potentially leading to accurate and robust classification results.

7. **InceptionResNetV2** - It is a highly effective CNN architecture widely utilized in computer vision tasks, including image classification. It is an extension of the original Inception architecture, incorporating residual connections and residual blocks inspired by the ResNet architecture. This combination enables efficient feature extraction and learning of complex image patterns.

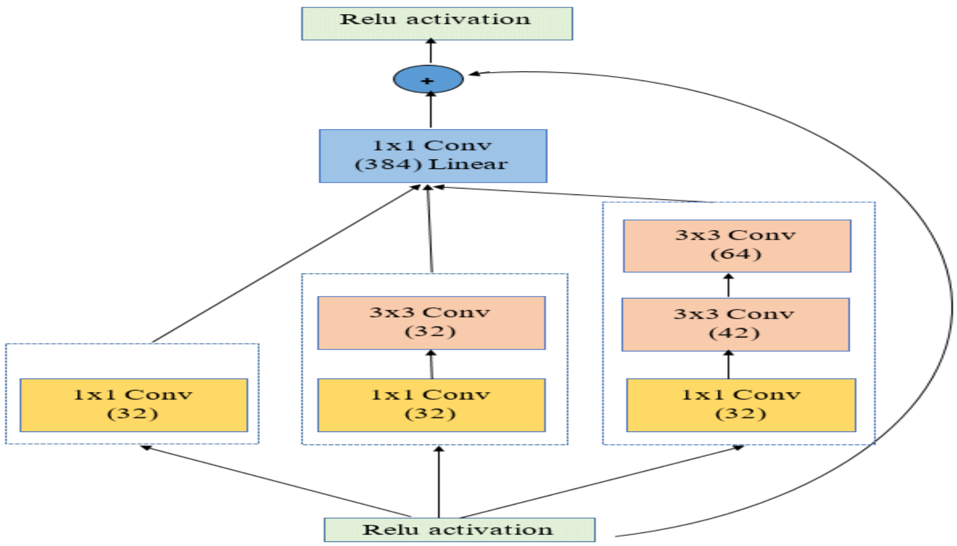

Fig 3.15 InceptionResNetV2Architecture [39]

With approximately 55 million trainable parameters, InceptionResNetV2 boasts a large network capacity. However, the exact parameter count may vary depending on the implementation and any modifications made for specific tasks, such as DFU classification.

When adapting InceptionResNetV2 for DFU classification, the final fully connected layer is replaced with a new layer containing the same number of output neurons as the classes in the DFU dataset. During training, only the weights of this new layer are updated, while the pre-trained weights of the remaining layers are typically frozen. This fine-tuning approach leverages the learned features from large-scale datasets like ImageNet.

By employing InceptionResNetV2 and fine-tuning it on the DFU dataset, the model benefits from its deep architecture and advanced design choices. This enables the model to effectively learn discriminative features from DFU images and make accurate predictions in the classification of DFU cases.

8. **MobileNetV2** – It is a popular type of neural network architecture used in computer vision for image classification tasks. It is known for its efficiency and effectiveness, making it well-suited for scenarios where computational resources are limited, such as on mobile devices. The architecture of MobileNetV2 involves using special blocks called inverted residuals, which help reduce the size of the model while maintaining good performance.

These blocks consist of lightweight convolutional layers followed by bottleneck layers. The model also includes skip connections to improve learning and gradient flow.

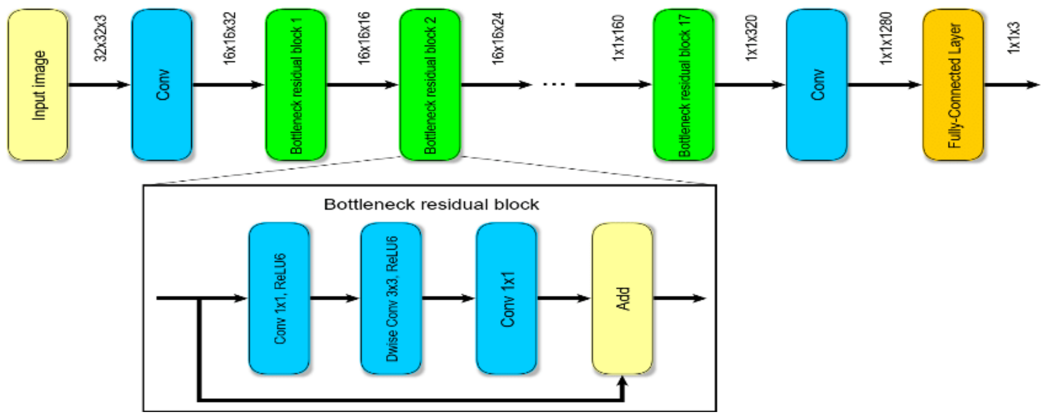

Fig 3.16: MobileNetNetV2Architecture [39]

In the context of classifying DFU images, the MobileNetV2 model is modified by replacing the last part of the network with a new layer that predicts the classes specific to DFU classification. During training, only the weights of this new layer are updated, while the rest of the model remains unchanged.

The parameter layer refers to the number of learnable parameters in the MobileNetV2 model. MobileNetV2 is designed to have a relatively low number of parameters compared to other architectures. This means it requires less computational resources to run and can be more efficient in terms of memory usage.

By utilizing the MobileNetV2 architecture and fine-tuning it on a DFU dataset, the model can effectively classify DFU images while being lightweight and resource-efficient. It strikes a balance between accuracy and computational requirements, making it suitable for deployment on devices with limited resources or when efficiency is a priority.

9. **EfficientNetB0** - It is a highly efficient and effective convolutional neural network (CNN) architecture widely used in image classification tasks. It has gained popularity for its superior performance while being lightweight and resource-efficient. EfficientNetB0 achieves its efficiency by carefully balancing the model's depth, width, and resolution. It utilizes various techniques, such as depthwise separable convolutions, inverted residual

blocks, and squeeze-and-excitation modules, to capture important image features while minimizing computational complexity.

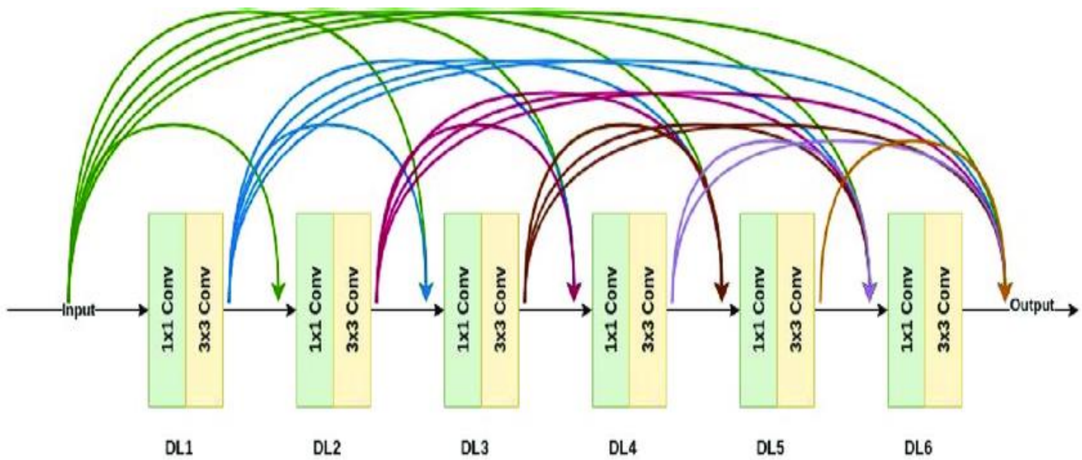

Fig 3.17: EfficientNetB0 Architecture [40]

With EfficientNetB0, we have a model that contains approximately 5.3 million learnable parameters. These parameters represent the weights and biases that the model adjusts during training to understand and classify images accurately. For DFU classification, we can fine-tune EfficientNetB0 by adapting the last fully connected layer of the model to match the number of classes in the DFU dataset. During training, we update the weights of this specific layer, while the remaining parameters of the model remain unchanged.

By utilizing EfficientNetB0, we can benefit from its efficient design and relatively low parameter count. This allows for faster training and deployment while still achieving high accuracy in classifying DFU images.

10. **DenseNet121** - It is a popular and effective convolutional neural network (CNN) architecture widely used for image classification tasks. It is particularly known for its dense connectivity pattern, which allows for improved information flow and feature reuse within the network. With a total of 121 layers, DenseNet121 is a deep and powerful architecture. It consists of various types of layers, including convolutional, pooling, and fully connected layers. In terms of parameters, DenseNet121 has approximately 8 million trainable parameters, which represent the weights and biases associated with the network's layers.

For the task of DFU classification, the original DenseNet121 model is adapted by replacing the last fully connected layer with a new layer that matches the number of classes in the DFU dataset. During training, only the parameters of the modified classification layer are updated, while the rest of the network's parameters are frozen.

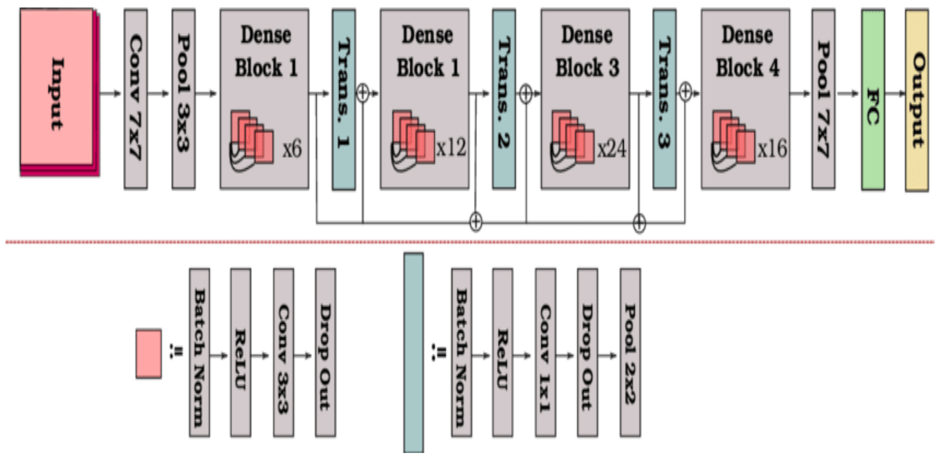

Fig 3.18: DenseNet121 Architecture [41]

By leveraging the unique dense connectivity and parameter efficiency of DenseNet121, the model can effectively learn and capture important features from DFU images. The dense connections promote feature reuse, allowing the network to make more informed and accurate predictions. The large number of parameters in DenseNet121 enables it to learn complex patterns and improve the classification performance for DFU images. Overall, DenseNet121 offers a powerful architecture for DFU classification, combining deep learning capabilities with efficient parameter usage to achieve accurate and reliable results.

3.4 EVALUATION METRICS

The evaluation criteria utilized will determine how well a model manages or predicts DFU. These metrics provide quantitative assessments of the model's capacity to predict weather DFU is present or absent in a patient.

Accurate DFU prediction is essential because early detection and appropriate treatment can halt the onset of serious effects including infection, amputation, and even death. Evaluation of

DFU models' performance is essential to ensuring that they accurately forecast DFU and guide appropriate therapy. Various evaluation metrics are implemented, including accuracy, precision, F1-score, AUC-ROC, recall, and confusion matrix, we may obtain better about the performance of the model. Each indication provides unique details about the model's advantages and disadvantages, which may help researchers and medical professionals decide which parts of the model should want improvement. A model with high accuracy but low recall, for instance, would be helpful to identify the level of risk of DFU in the patient but might miss some patients who really have the illness. While many of these individuals may not really have a high risk of DFU, a model with a high recall but low accuracy may identify a significant number of patients as such. Increased healthcare costs and unnecessary procedures might be the effect of this.
Therefore, through assessing the efficacy of DFU models using a range of metrics, we may better understand their advantages and disadvantages and choose whether to implement them in clinical practice. The list of commonly employed metrics for evaluation are as follows:

3.4.1. Accuracy

This statistic calculates the proportion of predictions made by that are accurate calculated by the model made separating the forecasts by the total number of TPs and TNs. Accuracy is a widely used evaluation criterion for detecting or prediction of DFUs. It measures the model's predictive accuracy, or TP proportion as well as TN proportion for all forecasts. The formula for accuracy is:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \tag{17}$$

True positive is denoted by TP (the cases of DFU predicted correctly by model as positive), the false positives count are basically FP (the cases of absence of DFU predicted incorrectly by model as positive), FN is the number of false negatives, and FN is the number of false negatives. TN stands for true negatives, which are situations lacking DFU that the model accurately identified as negative.

The percentage of cases that the model correctly recognised, whether or not they tested positive or negative for DFU, is hence known as accuracy. When the dataset is imbalanced, that is, when one class (for instance, DFU cases) is much less common than the other class, the model may attain high accuracy by merely projecting the majority class for all occurrences. Despite being a helpful statistic, precision may not always be acceptable. Even if a model that consistently predicts the majority class would have a high accuracy, it would not be practical if the DFU dataset was unbalanced (that is, if one class was much more prevalent than the other). Since they indicate, respectively, the proportion of properly predicted positive occurrences and the percentage of actual positive events that were correctly projected, precision and recall may be more appropriate evaluation measures in these circumstances.

Even while accuracy can be a useful metric for DFU prediction or management, it should be used in conjunction with other assessment measures to provide a more full analysis of the model's performance.

3.4.2. Precision

This statistic determines the proportion of true positives among all predictions that are positive generated by spliting all the number of true positives with the model by the number of all the TP and FP cases. It is especially useful when the dataset is uneven since it provides information on the percentage of correctly predicted positive cases. This can be very important when choosing a course of treatment for DFU prediction or management because a false positive diagnosis could lead to excessive treatment that could be damaging to the patient's health. The effectiveness of diabetic foot ulcer (DFU) prediction or management is assessed using the equation below:

$$\text{Precision} = \text{True Positives} / (\text{True Positives} + \text{False Positives}) \tag{18}$$

In this formula, TP refers to the number of actual positive occurrences the model properly predicted, while FP refers to the number of actual negative instances the model mistakenly interpreted as positive. Precision is the percentage of actual positives out of all positive predictions made by the model, and it measures how effectively the model can recognise positive situations without incorrectly classifying too many negative examples as positive.

A model with a high accuracy rating is less likely to forecast good outcomes that do not materialise and is less likely to wrongly label negative events as positive. When a model has a poor accuracy score, it is more likely to predict false positives, which might lead to needless medications or interventions for patients who do not have DFU.

A model for effective DFU prediction or management need to thus have a high precision score. A model's performance in controlling or predicting diabetic foot ulcers (DFU) may not be fully reflected by accuracy alone. Precision, which assesses the percentage of TP among all the predictions model made are positive, does not account for the proportion of cases of real positive which were correctly anticipated (recall).

The efficiency of a DFU prediction or management model must thus be assessed using a range of assessment techniques. In comparison to accuracy alone, metrics like recall and F1-score, which take into consideration both true positives and false negatives, may provide a more full assessment of the model's performance. For instance, recall measures the proportion of cases that are truly positive among all examples, whereas F1-score finds a balance between recall and accuracy. By using a number of assessment metrics, clinicians and researchers may better understand the advantages and disadvantages of a DFU prediction or management model and make more informed decisions regarding patient care.

3.4.3. Recall

For assessing the efficacy of models for treating or predicting DFU, recall is a crucial assessment criterion. It evaluates ratio of TP events that the model accuratly predicted as positive for effectively recognising and controlling DFU. Recall in the detection of DFU is calculated using the formula below:

$$\text{Recall} = \text{True Positives} / (\text{True Positives} + \text{False Negatives}) \tag{19}$$

In this formula, the term TP refers to the number of occurrences that are real positive that the model correctly recognised as such, while the term FN refers to the number of real positive examples that the model incorrectly identifies as such.

The recall metric demonstrates the model's capability to recognise genuine DFU cases. If the recall score is high it means that the model detection is done correctly on a large number of positive occurrences, whereas a low recall score indicates that many valid positive examples are being disregarded. An effective DFU prediction or management model should thus have a high recall score.

3.4.4. F1 score

The F1-score is a well-liked evaluation statistic for diabetic foot ulcer (DFU) prediction or management models. It is a model performance indicator that takes into consideration recall and accuracy. Although accuracy and recall are important assessment metrics for assessing the performance of a DFU prediction or management model, a comprehensive analysis of the performance of model overall is not provided .

All the predictions that are positive calculated by the model and the Percentage of TP among them is called precision, whereas recall measures the percentage of real positive cases that the model accurately predicted. A model, however, can have a high recall but a low accuracy, or the opposite might be true. For instance, a model may have low recall if it fails to identify many actual positive occurrences while having high accuracy by predicting just a limited number of positive events. The F1-score overcomes this restriction by combining accuracy and recall into a single score, gives a more comprehensive evaluation of the performance of model. It is calculated using the following formula:

$$\text{F1-score} = 2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})) \tag{20}$$

This method finds a balance between accuracy and recall by taking into account both true positives and false positives when evaluating the model's performance. When F1-score is comparatively higher, the model is consistently identifying positive cases of DFU while limiting false positive predictions, which also suggests that the model has a high accuracy and strong recall. A low F1-score may imply low accuracy, low recall, or both, which shows that more optimisation may be required to enhance the model's performance.

For evaluating the overall effectiveness one of the most useful metrics is F1score for DFU prediction or management models since it takes into account both accuracy and recall. This makes it possible to assess the model's performance in more detail.

3.4.5. Receiver Operating Characteristic curve

The effectiveness of a DFU prediction or management model is depicted visually by the ROC curve. The TPR and FPR for various categorization levels are further addressed.

The percentage of actual positive cases that the model properly predicted is known as the TPR or sensitivity. The proportion of real negative instances that the model mistakenly forecasted as positive cases is known as the FPR, on the other hand. Plotting the output probabilities of the model against the true positive rates and false positive rates at different classification thresholds results in a ROC curve for a DFU prediction or management model. The AUC, which shows how effectively the model can discriminate between positive and negative examples of DFU, serves as a gauge of the model's overall performance.

The effectiveness of a DFU prediction or management model is depicted visually by the ROC curve. The TPR and FPR for various categorization levels are further addressed.

If a model has an AUC of 1.0 and can distinguish between positive and negative events, it is said to be perfect. An AUC of 0.5 for a random model denotes no better results than chance.

A DFU prediction or management model's effectiveness may be assessed using the ROC curve since it gives a visual depiction of the model's performance at various categorization thresholds. Depending on the intended balance between sensitivity and specificity, it may also be used to choose the optimum threshold for the model.

3.4.6. Confusion matrix

A performance of a DFU prediction or classification model can also be evaluated using confusion matrix. In case of binary classification TP, TN, FP, and FN are calculated, where positive refers to cases with DFU and negative refers to cases without DFU. A confusion matrix for a DFU prediction or management model might look like this:

	Actual Positive (DFU)	Actual Negative (No DFU)
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Table 3.4 : Confusion Matrix

The number of real DFU instances that the model properly identified as positive is measured by the TP. The number of cases missing DFU that the model mistakenly classified as positives is known as FP. The number of actual DFU cases that the model incorrectly read as negative is known as FN. The number of DFU-free instances that the model properly identified as negative is known as the TN.

A variety of evaluation measures for the model, which offer a more thorough assessment of the model's performance, may be produced using the components of the confusion matrix.

On the basis of model and dataset evaluation metrics will be chosen for the DFU prediction or management model. For instance, if the goal is to decrease false negatives (i.e., projecting a negative outcome when it is actually positive), recall may be the most important metric. If the goal is to decrease false positives (i.e., projecting a positive outcome when it is really negative), accuracy may instead be the most important number.

3.5 PROPOSED METHODOLOGY

3.5.1 Research Design

The research design for this project focuses on comparing and evaluating the performance of ten transfer learning algorithms for DFU classification. The selected algorithms include ResNet50, InceptionV3, Xception, VGG16, MobileNetV3Small, ResNet101V2, InceptionResNetV2, MobileNetV2, EfficientNetB0, and DenseNet121. Among these algorithms, InceptionResNetV2 and EfficientNetB0 have shown promising results in initial experiments. Therefore, an ensemble model combining these two algorithms will be constructed to further enhance the DFU classification performance.

The dataset used for this project is obtained from Kaggle and contains 1055 images categorized into two classes: healthy skin and ulcer. To improve the diversity and generalization of the dataset, data augmentation techniques will be applied. The dataset will then be divided into training and validation sets. The training set will be utilized to fine-tune the pre-trained models, while the validation set will be used to monitor and compare the performance of the different algorithms during training.

Once the individual models have been trained and evaluated, an ensemble model will be constructed by combining the predictions of InceptionResNetV2 and EfficientNetB0. This ensemble model aims to leverage the strengths of both algorithms and produce more accurate and robust DFU classifications.

To assess the performance of the models, multiple performance metrics will be calculated. These metrics will provide a comprehensive evaluation of each algorithm's ability to classify DFU images accurately.

The predictions of the ensemble model will be compared against the actual labels of the test dataset to measure its accuracy and overall effectiveness in classifying DFU images. Through this research design, the project aims to identify the most effective transfer learning algorithms for DFU classification while demonstrating the benefits of an ensemble model combining InceptionResNetV2 and EfficientNetB0.

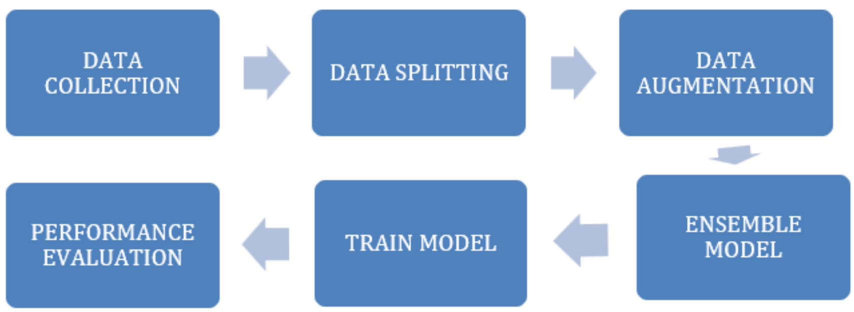

Fig 3.19 : Proposed Methodology

3.5.2 Data Pre-Processing

DFUs are a common side effect faced by diabetic patient, affect millions of people worldwide. DFU can cause substantial tissue damage and even lead to amputation of the affected limb if it is left untreated. In reducing the risk of complications by accurately recognising and forecasting the progression of DFU great potential is shown by DL models. This post will explain how to use techniques like data splitting, data augmentation, and fine-tuning to get a DFU dataset ready for transfer learning.

1. Data Splitting: In the creation of a deep learning model, data splitting is a crucial step since it allows for the evaluation of the model's performance on an unknown dataset. In

the context of DFU analysis, data splitting is the process of dividing the available DFU photos into sets for training, validation, and testing.

The available DFU images are used to train from the training set for the deep learning model. The model develops the capacity to recognise patterns and attributes in the photographs and make accurate predictions during the training phase. Performance is enhanced by altering the model's learning rate, batch size, and optimizer using the validation set.

Finally, using the test set, the performance of the is evaluated on an unknown dataset. Performance evaluation of model is provide accurately by implementing it on test dataset. To determine if the model generalises effectively to new, untested data, we may examine the model's precision, recall, accuracy, and other metrics.

The dataset should be separated consistently and randomly to prevent model bias. When one set has an uneven number of images from a certain class, the model's performance for that class may be overestimated or underestimated. Making sure the splitting is done in a controlled and objective manner, for as by utilising stratified sampling, is essential to ensuring that each set has a representative sample of the various classes of DFU pictures.

2. Data Augmentation: Data augmentation is a popular deep learning technique to fictitiously expand a training dataset by generating new copies of existing pictures. This approach is quite useful when the dataset is little and insufficient to effectively train a deep learning model.

In the context of diabetic foot ulcer (DFU) analysis, data augmentation approaches are essential to increase the diversity of the easily accessible images and boost the robustness of the model. Because of the additional variations provided by data augmentation, the model may learn a wider range of patterns and traits, enhancing its capacity to identify and classify DFUs more precisely in practical settings.

Different data augmentation methods, including as rotation, zooming, flipping, translating, noise adding, and contrast enhancing, can be used to improve DFU images. Rotation implies rotating the image at a certain angle as opposed to zoom, which involves scaling the image up or down. In translation, the image is moved either horizontally or vertically, as opposed to flipping, which rotates the image either horizontally or vertically. While noise addition involves adding random noise to the image, contrast enhancement involves altering the image's contrast.

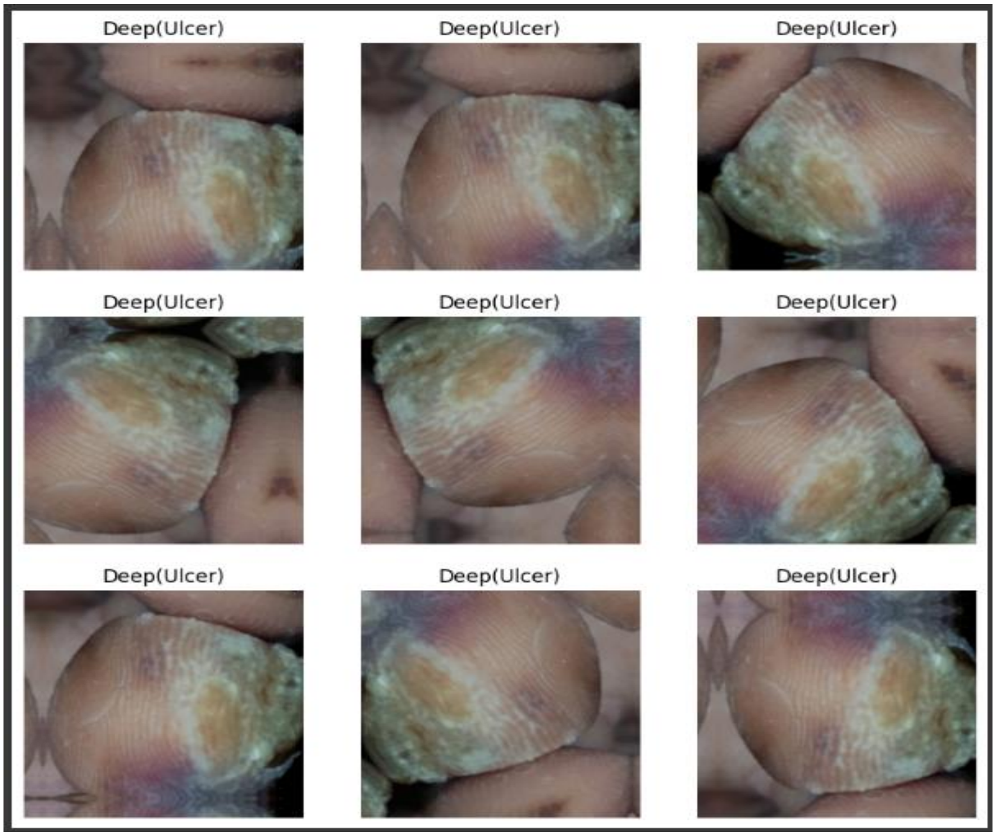

Fig 3.20: Augmented Images of DFU from Training Dataset

3.5.3 Model Architecture

Ensemble learning is a powerful technique that combines multiple models, known as base models or learners, to achieve better predictive performance than any individual model alone. It harnesses the collective wisdom of diverse models, leveraging their strengths and mitigating their weaknesses. By aggregating the predictions or combining the outputs of these models, ensemble learning can enhance accuracy, robustness, and generalization capabilities.

In the proposed model architecture, ensemble learning is utilized to improve the classification of diabetic foot ulcers (DFUs). Two popular pre-trained models, EfficientNetB0 and InceptionResNetV2, are employed as base models. These models have been trained on large-scale datasets and have demonstrated excellent performance in various image classification tasks.

To create the ensemble, the output tensors of the EfficientNetB0 and InceptionResNetV2 models are concatenated, merging their learned features into a fused representation. This fused representation captures a broader range of information, combining the distinctive characteristics learned by each model. By leveraging the diversity of these models, the ensemble model aims to enhance the overall classification accuracy and robustness.

The concatenated features are then passed through a dense layer, which performs additional transformations and feature refinement. Finally, the output layer with softmax activation is added to classify the DFU images into two classes: healthy skin or ulcer.

The ensemble model benefits from the complementary strengths of EfficientNetB0 and InceptionResNetV2. EfficientNetB0 is known for its efficient architecture and balanced performance, while InceptionResNetV2 combines the advantages of the Inception and ResNet architectures, capturing both local and global features effectively. By combining their features, the ensemble model aims to leverage their respective strengths and achieve superior performance in DFU classification.

Ensemble learning mitigates the risk of individual models making erroneous predictions by considering the collective decision of multiple models. It helps to reduce overfitting and improve generalization by combining the knowledge learned from diverse models. Ensemble models are often more robust, as they can handle different types of data and variations in the dataset.

In summary, the proposed ensemble model for DFU classification combines the learned features of EfficientNetB0 and InceptionResNetV2 models through concatenation. By leveraging ensemble learning, the model aims to enhance classification accuracy, robustness, and generalization capabilities. This approach capitalizes on the diversity and complementary strengths of the base models, leading to improved performance in identifying and classifying DFUs.

3.5.4 Training The Dataset

The training dataset is passed to the model, and during training, the parameters are adjusted to minimize the loss function. A validation dataset is used to evaluate the model to ensure that it is not overfitting to the training data.

Monitoring the performance of model during training is essential to ensure that it is learning new information from the data and getting closer to reducing the loss function. Loss and accuracy metrics can be computed on a regular basis, such as once every epoch, to achieve the model is not performing well on the validation dataset, adjustments can be made to the hyperparameters to improve the model's performance. Hyperparameters are specified parameters that may be altered to control the behaviour of the model. Some common hyperparameters that may be altered are learning rate, batch size, and number of epochs.

Faster convergence could result from a higher learning rate, but the loss function might also vary or even diverge. The size of the step that is taken during optimisation is determined by the learning rate. Faster training may result from larger batches, although this may also need more memory. How many samples are processed simultaneously during each iteration depends on the batch size.

When training, how many times the full dataset is processed depends on the number of epochs. Too few epochs can lead to underfitting, while too many epochs can lead to overfitting.

The model may be trained to attain the greatest performance on the validation dataset by keeping an eye on its performance and modifying the hyperparameters as required. This will enable it to generalise effectively to new, untested data.

CHAPTER-4
RESULTS

1.1 PERFORMANCE EVALUATION

The performance evaluation of the model was conducted by analyzing the accuracy metrics. Figure 4.1 displays the accuracy achieved by the model in training phase. It shows the accuracy at every epoch, with a total of 10 epochs. The accuracy metric measures the percentage of correctly predicted instances by the model.

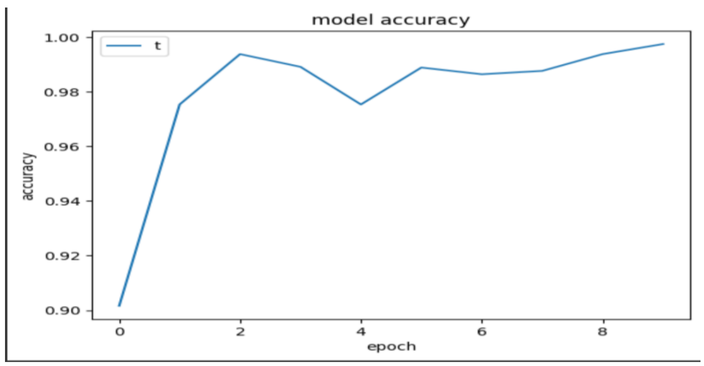

Fig 4.1 : Accuracy at every epoch

The graph demonstrates the progression of accuracy over the course of training. It provides insights into how the model's performance evolves and improves with each epoch. By observing the accuracy at different epochs, we can identify trends, patterns, and convergence of the model's learning process.

```
Val Loss: 0.0118526890873909
Val Accuracy: 0.9952380657196045
Val Precision: 0.9952380657196045
Val Recall: 0.9952380657196045
Val F1 Score: 0.995237410068512
```

Fig 4.2: Performance Evaluation of Proposed Model

The results obtained from the performance evaluation are crucial in assessing the effectiveness of the model. The accuracy metric provides a quantitative measure of the ability of the model to correctly classify instances within the training set. Higher accuracy values indicate better performance and stronger predictive capabilities of the model.

1.2 COMPARISON OF THE MODEL WITH OTHER APPROACHES

The performance evaluation of the different models used in the study was conducted to compare their effectiveness in classifying diabetic foot ulcers (DFUs). The models considered for comparison were ResNet50, InceptionV3, Xception, VGG16, MobileNetV3Small, ResNet101V2, InceptionResNetV2, MobileNetV2, EfficientNetB0, and DenseNet121. Each model was trained and evaluated on a dataset of DFU images, and their respective accuracy values were recorded.

The results of the performance evaluation revealed significant variations in the accuracy achieved by the different models. Among them, InceptionResNetV2 and MobileNetV3Small emerged as the top performers with accuracy scores of 0.995 and 0.990, respectively. These models demonstrated a high level of precision in correctly classifying the DFU images. Notably, VGG16 and EfficientNetB0 also exhibited strong performance with accuracies of 0.986 and 0.990, respectively.

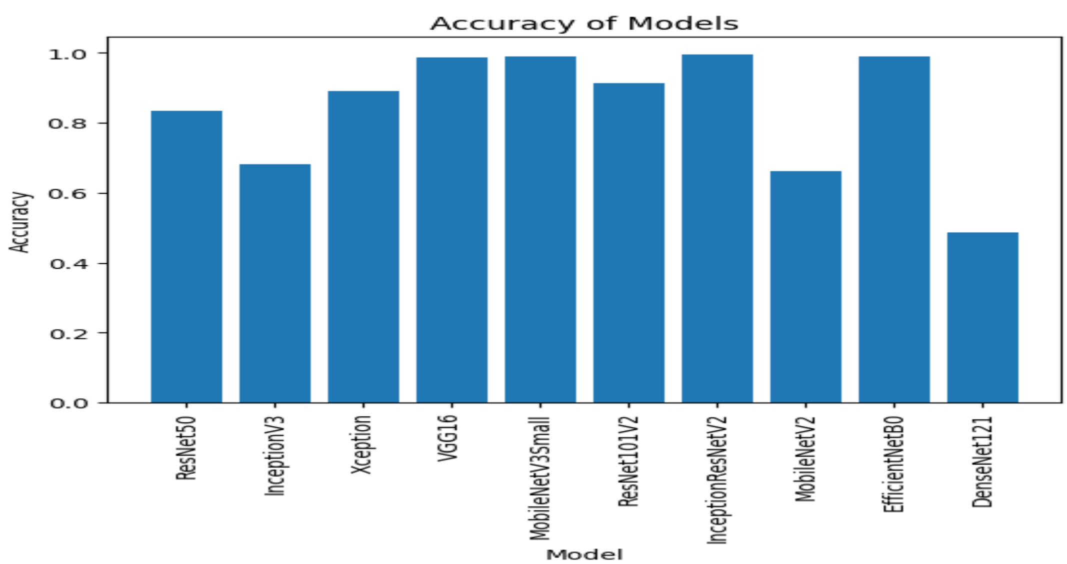

Fig 4.3 : Comparison of accuracy

On the other hand, DenseNet121 displayed the lowest accuracy score of 0.486, indicating relatively lower effectiveness in DFU classification. It is important to consider other evaluation metrics such as precision, recall, and F1 score to gain a comprehensive understanding of the models' performance.

Precision, which measures the model's ability to make correct positive predictions, showed similar trends. InceptionResNetV2 obtained the highest precision score of 0.995, indicating its capability to accurately classify DFU images. MobileNetV3Small, VGG16, and EfficientNetB0 also demonstrated good precision with scores of 0.990, 0.986, and 0.662, respectively. DenseNet121 had the lowest precision score of 0.486.

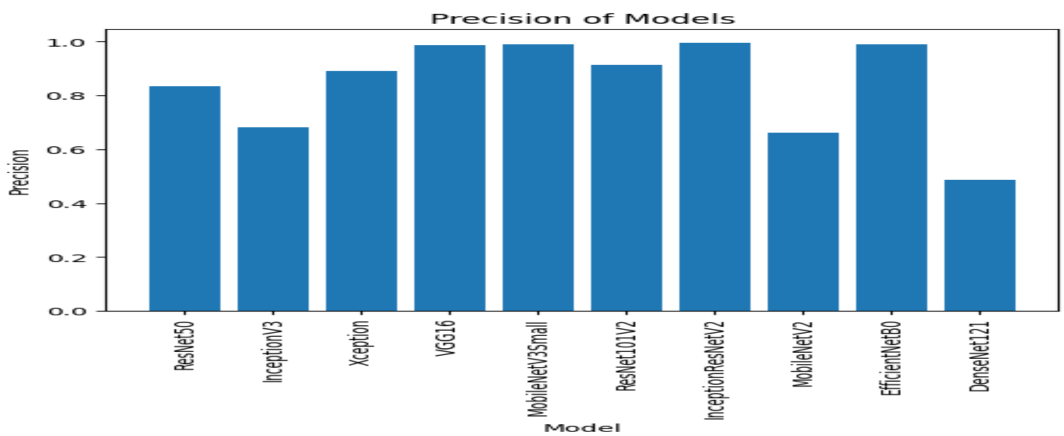

Fig 4.4 : Comparison of precision

Examining recall, which measures the model's ability to correctly identify positive instances, InceptionResNetV2 again outperformed other models with a recall score of 0.995. MobileNetV3Small and VGG16 followed closely with scores of 0.990 and 0.986, respectively. DenseNet121 had the lowest recall score of 0.486.

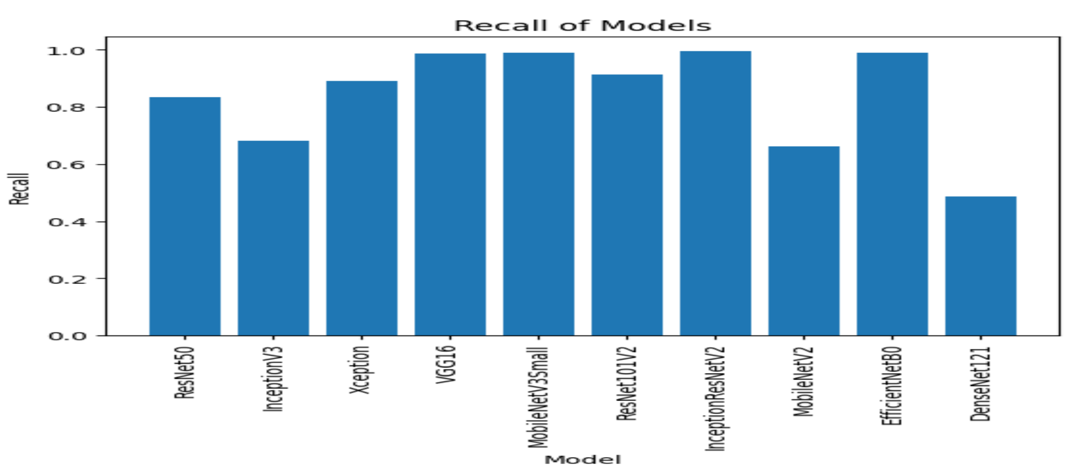

Fig 4.5 : Comparison of recall

The F1 score, which considers both precision and recall, further illustrates the models' overall performance. InceptionResNetV2 achieved the highest F1 score of 0.995, indicating a balanced performance in terms of precision and recall. MobileNetV3Small, VGG16, and EfficientNetB0 also demonstrated good F1 scores of 0.990, 0.986, and 0.990, respectively. However, DenseNet121 had a lower F1 score of 0.318, indicating its limitations in accurately classifying DFU images.

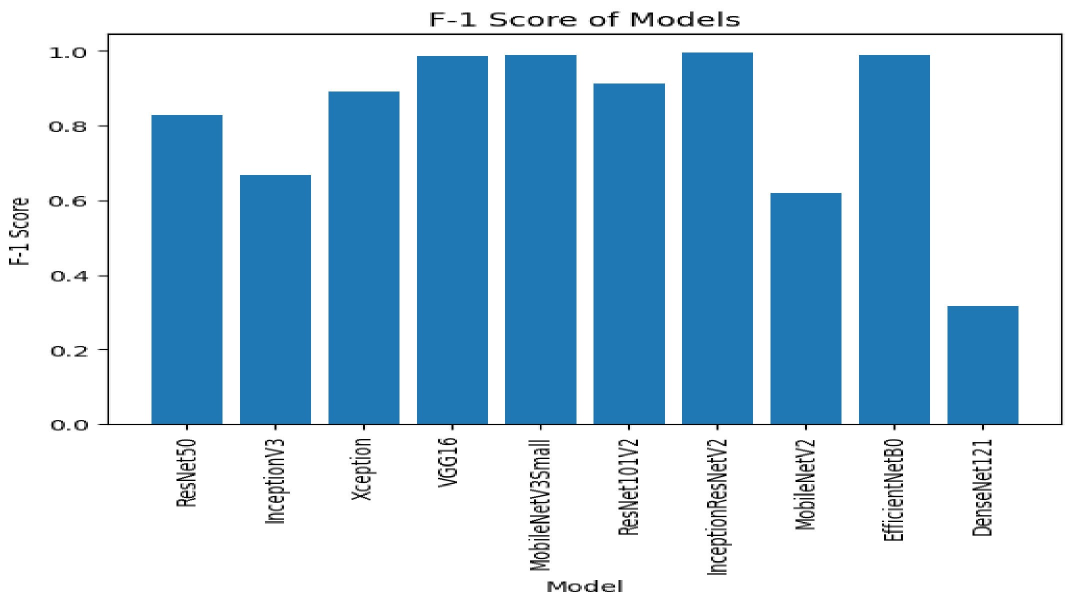

Fig 4.6 : Comparison of F-1 Score

In summary, InceptionResNetV2 and EfficientNetB0 emerged as the top-performing models for DFU classification. Their superior performance suggests their effectiveness in accurately identifying and classifying DFU images, while DenseNet121 showed relatively lower performance in comparison. These findings highlight the importance of choosing appropriate models for DFU classification tasks, with ensemble techniques such as combining InceptionResNetV2 and EfficientNetB0 showing potential for further enhancing classification performance.

Model	Accuracy	Precision	Recall	F1 Score
ResNet50	0.833333	0.833333	0.833333	0.829423
InceptionV3	0.680952	0.680952	0.680952	0.666276
Xception	0.890476	0.890476	0.890476	0.889686
VGG16	0.985714	0.985714	0.985714	0.985716
MobileNetV3Small	0.990476	0.990476	0.990476	0.990476
ResNet101V2	0.914286	0.914286	0.914286	0.914068
InceptionResNetV2	0.995238	0.995238	0.995238	0.995237
MobileNetV2	0.661905	0.661905	0.661905	0.619254
EfficientNetB0	0.990476	0.990476	0.990476	0.990478
DenseNet121	0.485714	0.485714	0.485714	0.317582

Table 4.1: Comparison of performance metrics

CHAPTER-5
CONCLUSION AND FUTURE WORK

5.1 SUMMARY OF THE STUDY

In this study, we investigated the effectiveness of ensemble learning techniques to enhance the classification performance of Diabetic Foot Ulcers (DFU) using transfer learning models. Ensemble learning takes advantage of the diversity and unique strengths of different models, leveraging their collective knowledge to improve overall accuracy. By combining the predictions of multiple models, ensemble learning addresses the limitations of individual models and enhances the quality of the classification.

For our research, we focused on creating an ensemble model by combining two high-performing transfer learning models: InceptionResNetV2 and EfficientNetB0. We selected these models based on their individual performance and their ability to complement each other. InceptionResNetV2 excels in capturing intricate features and patterns, while EfficientNetB0 is known for its efficiency and effectiveness in classification tasks.

The ensemble model was designed by aggregating the predictions of both InceptionResNetV2 and EfficientNetB0. By combining the insights and decision-making processes of these models, the ensemble model benefited from their diverse perspectives. This allowed the ensemble model to generate more accurate and reliable predictions for DFU classification.

The results of our study demonstrated that the ensemble model, incorporating InceptionResNetV2 and EfficientNetB0, outperformed the individual models. This highlights the effectiveness of ensemble learning in improving the classification of DFUs.

By utilizing ensemble learning, we were able to create a robust and reliable approach for DFU classification, harnessing the collective intelligence of multiple models. The ensemble model provided improved generalization, increased stability, and enhanced predictive performance.

In conclusion, our study emphasizes the effectiveness of ensemble learning, specifically when applied to the combination of InceptionResNetV2 and EfficientNetB0 models, for enhancing the classification of DFUs. The ensemble model offers a valuable tool for healthcare professionals in accurately diagnosing and managing diabetic foot ulcers, ultimately leading to improved patient care and outcomes.

5.2 IMPLICATIONS OF THE STUDY FOR CLINICAL PRACTICE

The research presented in this dissertation has important implications for clinical practises in the diagnosis and management of DFU. With the help of deep learning techniques like multi-class classification models, medical professionals may diagnose and classify diabetic foot ulcers more accurately, leading to more effective treatment and better patient outcomes.

By accurately assessing the ulcer's severity, doctors can develop a customised treatment plan for each patient that may include wound care, unloading, antibiotics, and surgery. This personalised approach can improve patient outcomes and reduce the likelihood of negative effects like amputation.

In addition, identifying high risk patients who are more likely to develop diabetic foot ulcers can be facilitated by using DL techniques. In order to reduce the risk of developing ulcer, this early recognition may enable practitioners to take preventive measures such as foot examinations and patient education on their care.

By using deep learning algorithms, it has the potential for physicians to enhance diagnosis accuracy while reducing the demand on healthcare resources. Compared to more traditional diagnostic approaches, these tools enable doctors to quickly and precisely characterise and diagnose diabetic foot ulcers. Overall, the application of methods based on deep learning, such as multi-class classification models, may transform the detection and treat diabetic foot ulcers at early stage. By improving diagnosis accuracy, identifying high-risk patients, and tailoring treatment to each patient, clinicians can improve patient outcomes, reduce complications, and boost therapy effectiveness.

5.3 LIMITATIONS OF THE STUDY

DFUs are a serious side effect caused due to diabetes that, untreated, will lead to amputation and other unfavorable consequences. The DL method that has been proposed as a practical tool for accurately identifying and classifying ulcers associated with diabetes is multi-class classification models. However, there are nonetheless significant drawbacks to this dissertation, and further study will be needed to develop its conclusions.

One drawback of the studies is that they have a small sample size. Even when the data used for this study may have been representative of the community, a larger sample size could improve the generalizability of results. Also, there may have been a bias in the dataset which could affect how well this classification model is performing. Another downside of this study is that it does not have any data from the dataset. The accuracy of the model and the conclusions of the study can be compromised by missing data.

There may also be a limit to the scope of the study due to the features considered in the classification model. Although some of the elements used in this study may have been relevant, when you add further features such as demographic characteristics and medical history it can possibly improve your model's accuracy. Another drawback is that this study has not been validated outside of it. Although the model was validated internally, external validation may increase the generalizability of the findings.

5.4 FUTURE WORK

The present study used a multi-class classification model that has produced encouraging results with regard to the accurate detection and categorization of diabetic foot ulcers. However, future studies may build on the results presented by utilising larger datasets to improve the classification algorithm's precision and generalisation. One of the research's biggest weaknesses was the presence of missing data, which might be addressed in future work by applying imputation techniques or data augmentation. It could be helpful as well with improving the model's accuracy when researching incorporating additional functions.

Future studies could compare different dl methods to the multi-class classification model implemented in this work. By performing longitudinal research to learn more about how diabetic foot ulcers grow, the classification model's precision could be improved. Additionally, the categorization system might be improved and put the patient's needs first by using patient-reported results. As a result of this, additional research is required to address its limitations and build on its findings, even if the current study has provided insight into the use of multi-class classification models for categorization of dfu. That will enable us to continue improving the efficacy and accuracy of deep learning methods used in the identification as well as treatment of diabetic foot ulcers.

5.5 CONCLUSION

In conclusion, this study investigated the effectiveness of ensemble learning techniques for improving the classification of DFUs using transfer learning models. Among the evaluated pre-trained models, InceptionResNetV2 and EfficientNetB0 emerged as the top performers in terms of accuracy, precision, recall, and F1 score. These models exhibited exceptional capabilities in capturing the intricate features and patterns present in DFU images.

To further enhance the classification performance, we employed ensemble learning by combining the strengths of InceptionResNetV2 and EfficientNetB0 into a single ensemble model. This approach proved successful, resulting in superior performance compared to individual models.

The findings of this study underscore the value of ensemble learning in medical image analysis, particularly in the context of DFU classification. By leveraging the collective knowledge and expertise of multiple models, we were able to achieve more accurate and reliable identification and classification of DFUs. This has significant implications for healthcare professionals, as timely and accurate detection of DFUs can greatly impact treatment planning and patient outcomes. This research demonstrates that InceptionResNetV2 and EfficientNetB0, along with ensemble learning techniques, offer a promising approach to enhance the classification of DFUs. This study contributes to the field of analysis of medical image and encourages further exploration of ensemble learning in diagnosing DFUs and other related medical conditions. By harnessing the power of ensemble learning, we can improve diagnostic accuracy and contribute to more effective treatment strategies for patients with DFUs.

CHAPTER-6
REFERENCES

[1] Shahbazian H, Yazdanpanah L, Latifi SM. Risk assessment of patients with diabetes for foot ulcers according to risk classification consensus of International Working Group on Diabetic Foot (IWGDF) Pak J Med Sci. 2013;**29**:730–734.

[2] Ramachandran A, Snehalatha C, Shetty AS, Nanditha A. Trends in prevalence of diabetes in Asian countries. World J Diabetes. 2012;**3**:110–117.

[3] Singh N, Armstrong DG, Lipsky BA. Preventing Foot Ulcers in Patients With Diabetes. *JAMA.* 2005;293(2):217–228. doi:10.1001/jama.293.2.217

[4] Alexiadou K, Doupis J. Management of diabetic foot ulcers. Diabetes Ther. 2012 Nov;3(1):4. doi: 10.1007/s13300-012-0004-9. Epub 2012 Apr 20. PMID: 22529027; PMCID: PMC3508111.

[5] Shah P, Inturi R, Anne D, Jadhav D, Viswambharan V, Khadilkar R, Dnyanmote A, Shahi S. Wagner's Classification as a Tool for Treating Diabetic Foot Ulcers: Our Observations at a Suburban Teaching Hospital. Cureus. 2022 Jan 22;14(1):e21501. doi: 10.7759/cureus.21501. PMID: 35223277; PMCID: PMC8861474.

[6] AlSadrah SA. Impaired quality of life and diabetic foot disease in Saudi patients with type 2 diabetes: A cross-sectional analysis. SAGE Open Medicine. 2019;7. doi:10.1177/2050312119832092

[7] Benjamin A. Lipsky, Anthony R. Berendt, Paul B. Cornia, James C. Pile, Edgar J. G. Peters, David G. Armstrong, H. Gunner Deery, John M. Embil, Warren S. Joseph, Adolf W. Karchmer, Michael S. Pinzur, Eric Senneville, 2012 Infectious Diseases Society of America Clinical Practice Guideline for the Diagnosis and Treatment of Diabetic Foot Infections, *Clinical Infectious Diseases*, Volume 54, Issue 12, 15 June 2012, Pages e132–e173, https://doi.org/10.1093/cid/cis346

[8] Alzubaidi, Laith & Fadhel, Mohammed & Al-Shamma, Omran & Zhang, Jinglan & Santamaría, Jorge & Duan, Ye & Oleiwi, Sameer. (2020). Towards a Better Understanding of Transfer Learning for Medical Imaging: A Case Study. Applied Sciences. 10. 4523. 10.3390/app10134523.

[9] Niri, Rania & Hassan, Douzi & Yves, Lucas & Treuillet, Sylvie. (2020). Semantic Segmentation of Diabetic Foot Ulcer Images: Dealing with Small Dataset in DL Approaches. 10.1007/978-3-030-51935-3_17.

[10] B. Cassidy, et al.,"A Cloud-Based Deep Learning Framework for Remote Detection of Diabetic Foot Ulcers" in IEEE Pervasive Computing, vol. 21, no. 02, pp. 78-86, 2022

[11] Xu, Y., Jia, Z., Wang, LB. *et al.* Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC Bioinformatics* 18, 281 (2017). https://doi.org/10.1186/s12859-017-1685-x

[12] Qu, X., Li, H., Zhu, L., Wang, J., & Zhang, Y. (2021). Automatic segmentation and classification of diabetic foot ulcer images based on deep learning. BMC Medical Informatics and Decision Making, 21(1), 1-11. doi: 10.1186/s12911-021-01534-0.

[13] Hassanien, A. E., Elhoseny, M., Shankar, K., & El-Sappagh, S. (2020). Deep learning-based approach for detecting and classifying the severity level of diabetic foot ulcer images. IEEE Access, 8, 188011-188022. doi: 10.1109/ACCESS.2020.3034132.

[14] He, H., Xia, M., Zhang, Y., Li, W., Lu, Y., & Qian, Y. (2020). Detection of diabetic foot ulcer images using transfer learning based on Inception-v3. Health Information Science and Systems, 8(1), 1-9. doi: 10.1007/s13755-020-00117-9.

[15] Hossain, M. S., Ibrahim, M. A., Akter, T., Kadir, M. A., & Rahman, M. A. (2021). An efficient deep learning-based technique to identify and classify diabetic foot ulcers in wound images. Journal of Medical Systems, 45(3), 1-12. doi: 10.1007/s10916-021-01769-1.

[16] D. Xue et al., "An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification," in IEEE Access, vol. 8, pp. 104603-104618, 2020, doi: 10.1109/ACCESS.2020.2999816.

[17] D. Xue et al., "An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification," in IEEE Access, vol. 8, pp. 104603-104618, 2020, doi: 10.1109/ACCESS.2020.2999816.

[18] Nguyen, T. D., Le, V. T., & Nguyen, Q. H. (2018). An Ensemble Deep Learning Model for Image Classification Using Transfer Learning. In 2018 International Conference on Advanced Computing and Applications (ACOMP) (pp. 77-82). IEEE.

[19] Alzubaidi, L., Fadhel, M. A., Oleiwi, S. R., Al-Shamma, O., & Zhang, J. (2020). DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. Multimedia Tools and Applications, 79(21), 15655-15677.

[20] Alzubaidi, Laith, Mohammed A. Fadhel, Omran Al-Shamma, Jinglan Zhang, J. Santamaría, and Ye Duan. "Robust application of new deep learning tools: an experimental study in medical imaging." Multimedia Tools and Applications (2021): 1-29.

[21] Alzubaidi, Laith, Mohammed A. Fadhel, Omran Al-Shamma, Jinglan Zhang, J. Santamaría, Ye Duan, and Sameer R Oleiwi. "Towards a better understanding of transfer learning for medical imaging: a case study." Applied Sciences 10, no. 13 (2020): 4523

[22] https://medium.com/@favourphilic/an-easy-guide-to-feature-selection-77f7ac5ef5b0

[23] Niri, Rania & Hassan, Douzi & Yves, Lucas & Treuillet, Sylvie. (2020). Semantic Segmentation of Diabetic Foot Ulcer Images: Dealing with Small Dataset in DL Approaches. 10.1007/978-3-030-51935-3_17.

[24] Sujit Kumar Das, Pinki Roy, Arnab Kumar Mishra, DFU_SPNet: A stacked parallel convolution layers based CNN to improve Diabetic Foot Ulcer classification, ICT Express, Volume 8, Issue 2,2022,Pages 271-275,ISSN 2405-9595, https://doi.org/10.1016/j.icte.2021.08.022.

[25] Al-Garaawi N, Ebsim R, Alharan AFH, Yap MH. Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks. Comput Biol Med. 2021 Nov 24;140:105055. doi: 10.1016/j.compbiomed.2021.105055. Epub ahead of print. PMID: 34839183.

[26] Oh C, Chung JY, Han Y. An End-to-End Recurrent Neural Network for Radial MR Image Reconstruction. Sensors (Basel). 2022 Sep 26;22(19):7277. doi: 10.3390/s22197277. PMID: 36236376; PMCID: PMC9572393.

[27] mobupu. (2019). Tool to draw neural network directed graphs, https://stats.stackexchange.com/q/392112

[28] Khanna, N.N.; Maindarkar, M.A.; Viswanathan, J.S. Cardiovascular/Stroke Risk Stratification in Diabetic Foot Infection Patients Using Deep Learning-Based Artificial Intelligence: An Investigative Study. *J. Clin. Med.* **2022**, *11*, 6844. https://doi.org/10.3390/jcm1122684

[29] Guo, Lulu & Yang, Bowen. (2020). Transactions on Transportation Electrification Cyber-Attack Detection for Electric Vehicles Using Physics-Guided Machine Learning. IEEE Transactions on Transportation Electrification. PP. 1. 10.1109/TTE.2020.3044524.

[30] Alzubaidi, L., Bai, J., Al-Sabaawi, A. *et al.* A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data* **10**, 46 (2023). https://doi.org/10.1186/s40537-023-00727-2

[31] https://paperswithcode.com/method/controlvae

[32] Park, Sung-Wook & Ko, Jae-Sub & Huh, Jun-Ho & Kim, Jong-Chan. (2021). Review on Generative Adversarial Networks: Focusing on Computer Vision and Its Applications. Electronics. 10. 1216. 10.3390/electronics10101216.

[33] Mahmood, Ammar & Giraldo-Ospina, Anita & Bennamoun, Mohammed & An, Senjian & Sohel, Ferdous & Boussaid, Farid & Hovey, Renae & Fisher, Robert & Kendrick, Gary. (2020). Automatic Hierarchical Classification of Kelps Using Deep Residual Features. Sensors. 20. 447. 10.3390/s20020447.

[34] Chen, Yujin & Chen, Ruizhi & Liu, Mengyun & Xiao, Aoran & Wu, Dewen & Zhao, Shuheng. (2018). Indoor Visual Positioning Aided by CNN-Based Image Retrieval: Training-Free, 3D Modeling-Free. Sensors. 18. 2692. 10.3390/s18082692.

[35] Bilal, Muhammad & Maqsood, Muazzam & Yasmin, Sadaf & Ul Hasan, Najam & Rho, Seungmin. (2022). A transfer learning-based efficient spatiotemporal human action recognition framework for long and overlapping action classes. The Journal of Supercomputing. 78. 10.1007/s11227-021-03957-4.

[36] Ali, Luqman & Alnajjar, Fady & Jassmi, Hamad & Gochoo, Munkhjargal & Khan, Wasif & Serhani, Mohamed. (2021). Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures. Sensors. 21. 1688. 10.3390/s21051688.

[37] Shuvo, Md & Islam, Md & Shamsojjaman, Muhammad & Akter, Shamima & Hasan, Shazid & Khatun, Tania. (2021). Recognizing Human Emotions from Eyes and Surrounding Features: A Deep Learning Approach. International Journal of Advanced Computer Science and Applications. 12. 386-394. 10.14569/IJACSA.2021.0120346.

[38] Seidaliyeva, Ulzhalgas & Akhmetov, Daryn & Ilipbayeva, Lyazzat & Matson, Eric. (2020). Real-Time and Accurate Drone Detection in a Video with a Static Background. Sensors. 20. 3856. 10.3390/s20143856.

[39] Amin, Hassan & Darwish, Ashraf & Hassanien, Aboul & Soliman, Mona. (2022). End-to-End Deep Learning Model for Corn Leaf Disease Classification. IEEE Access. 10. 31103-31115. 10.1109/ACCESS.2022.3159678.

[40] Radwan, Noha. (2019). Leveraging Sparse and Dense Features for Reliable State Estimation in Urban Environments. 10.6094/UNIFR/149856.