Indian Journal of Engineering & Materials Sciences
Vol. 31, August 2024, pp. 487-504
DOI: 10.56042/ijems.v31i4.8853

Artificial Intelligence techniques enabled insights into Leather Defects

Shubhadip Chakrabarti[a], Swamiraj Nithiyanantha Vasagam[b*], Balasundaram Ananthakrishnan[a] & Madasamy Sornam[d]

[a]Vellore Institute of Technology, Chennai, Tamil Nadu, 600127, India

[b]CSIR - Central Leather Research Institute (CLRI), Adyar, Chennai, Tamil Nadu, 600020, India

[c]Center for Cyber Physical Systems, School of Computer Science and Engineering,Vellore Institute of Technology, Chennai, Tamil Nadu, 600127, India

[d]Department of Computer Science, University of Madras, Guindy Campus, Chennai, Tamil Nadu, 600025, India

Received: 16 February 2024; Accepted: 10 July 2024

With the advent of the digital revolution, the detection of leather surface defects has gained immense significance towards automation in the assessment of leather quality, which is of paramount importance in the leather trade that has eventually become global. The proposed work has strived to develop an artificial intelligence-enabled reliable and efficient system for detecting leather surface defects using a leather image dataset. The work has utilized conventional machine learning algorithms and deep learning approaches for distinguishing leather surfaces. However, it has been found that due to the variability in the leather surface and defects, the conventional machine learning algorithms have not been able to satisfactorily distinguish the leather surfaces. As a result, *LeatherNet*, a novel lightweight deep neural network, has been proposed. For better analysis, the performance of *LeatherNet* has been compared with the performances of prominent existing convolutional neural network models, previously experimented machine learning algorithms, and existing state-of-the-art methods in this domain. The performance of *LeatherNet* has been found to outperform all the algorithms, architectures, and existing state-of-the-art methods considered. Accuracy, loss, precision, recall, and AUC score metrics have been used for performance measurement. When trained for 1500 epochs, the proposed model has recorded maximum training accuracy, precision, and recall of 99.78%, 99.69%, and 99.92% respectively, while the maximum testing accuracy, precision, and recall have been recorded at 97.42%, 97.66%, and 99.40% respectively.

Keywords: Smart leather defect Detection, AI in leather industry, Leather quality assessment, Artificial intelligence (AI), Leather image processing, Leather imagedataset, Convolution neural networks (CNN), Deep learning (DL)

1 Introduction

Leather sector occupies a key position in global trade and plays a major role in strengthening the economy of a nation. It has remained as the major foreign exchange earners in India, which happens to be the second largest exporter of leather garments, and 4th largest exporter of leather goods in the world. The year 2020-21 witnessed Indian export of footwear, leather, and leather products to the tune of US $3.68 billion, as indicated in the Council for Leather Exports (CLE) website[1].

Leather making involves a series of chemico-mechanical operations to convert raw hides and skins into a product for industrial as well as high end consumer product applications of immense commercial significance. While the outer covering of the body of a bigger animal like cattle or buffalo is called hide, that of a smaller animal, such as goat, sheep, etc, is called skin. India is bestowed with a rich livestock population. As per the CLE website, India is endowed with 20% of world cattle as well as buffalo and 11% of world goat & sheep population.

While synthetics have emerged as close competitor for leather especially in low value product market, leather maintains its edge especially in the high end as well as fashion market primarily because of its natural look, feel, breathability and other inherent properties as a unique biological material. A flawless surface is thus considered to be of immense importance to fetch a better price in the marketplace.

It is pertinent in this context that as a natural biological material, leather surface is prone to be affected by several defects that degrade its quality and value. These defects may generally be classified into two main categories – ante-mortem as well as post-mortem. Surface defects that commonly lead to deterioration of quality include the following:

*Corresponding author (Email: nvasagam@clri.res.in)

- Wart/ Scar mark
- Sore/Wound/Pox mark
- Rub/Scratch mark
- Yoke/goad mark
- Insect bite/Tick mark
- Brandmark
- Warblehole
- Cockle
- Wrinkle
- Flay cut
- Vein mark
- Stains/Putrefaction

Post-mortem defects generally arise due to improper flaying, curing or processing. While these defects may be controlled to some extent by employing adequate handling as well as care, ante-mortem defects pose a major challenge to the tanners. Since the success of the leather industry depends primarily on the quality of the hides and skins, defect detection seems to be of much significance for the long-term sustenance of the industry.

Quality of leather is adversely affected by defects immensely, leading to substantial reduction in price in the marketplace. This necessitates appropriate corrective measures during leather processing to ensure up-gradation depending on the extent of defects leading to the quality degradation of the hide/skin to be processed, for possible value enhancement. Further, market value of finished leather or its product depends on the quality grade of the leather based on the defects present thereon. Thus, grading for quality forms an important expert activity in the leather industry to gain prior idea about the quality of the material. Skilled professionals normally carry out this activity by visual inspection using their enormous experience, acumen and dexterity.

With the advent of digital revolution, it is considered appropriate to utilize the intervention of artificial intelligence to detect the defects present on a leather surface at a much faster rate to ensure multi fold enhancement of value as well as productivity with better consistency through automation.

Currently, multiple AI based techniques for the defect detection and classification in various domains have been proposed. Most of the algorithms have primarily used machine learning and deep learning algorithms for the defect detection. Significant works have also been done in defect segmentation using various image segmentation techniques.

The primary objective of this manuscript is to develop an artificial intelligence-based system that is both highly reliable and highly efficient in detecting and classifying leather surfaces into defects and non-defects. This manuscript essentially dealt with developing a reliable and efficient automation system for the detection and classification of leather surfaces into defects and non-defects.

1.1 Literature Survey

Various Machine Learning (ML), Deep Learning (DL) and Transfer Learning (TL) concepts have been introduced and employed in the Leather Industry to successfully analyze the defective and non-defective leather images. Some of the computer vision approaches used for the analysis of leather images are stated as follows:

Swamiraj Nithiyanantha Vasagam et al. [2], proposed a methodology that uses the concepts of Black Hat Transform and Hough Transform. The methodology facilitates the selection of features using ensemble algorithms. The proposed methodology recorded an accuracy of 94.25% on the validation portion of the dataset. Y.S.Gan et al. [3], proposed a framework that could successfully differentiate between Tick-Bite Defective and Non-defective leather patches. They used a collection of 1600 calf leather patches to enable the development of such a framework. They proposed a six-step pre-processing procedure to enhance the quality of the leather image dataset. The images were initially subjected to Histogram Matching, followed by resizing of the images from 400x400 to 100x100. These images were then converted into its gray scale equivalent, followed by reduction of noise using Gaussian Filters. Later, these images were subjected to Canny Edge Detector followed by application of Histogram of Gradients (HOG). These processed images were then processed by several classifiers, namely, k-Nearest Neighbors (K-NN), SVC, Linear Support Vector Machine, Decision Trees (DT), Random Forest (RF), AdaBoost, Gradient Boosting, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), ANN and XBoost ANN. ANN and XBoost ANN attained the highest classification accuracy recording a value as high as 94%. Malathy Jawahar et at[4]., proposed a machine learning methodology for detection and classification of leather defects. They used 577 leather defect and 100 leather non-defect images during the study. The images were initially subjected to Weiner Filter to remove the background noises. The Gray Level Co-occurrence Matrix (GLCM), HOG features along with Hu moments and

HSV color features were extracted from these processed images. The extracted features were classified using Machine Learning algorithms like Linear Regression, Linear Discriminant Analysis, K-Nearest Neighbors (kNN), Classification and Regression Tree (CART), Random Forest, Support Vector Machines and Multi-Layer Perceptron. Random Forest was found to record the highest classification accuracy for both GLCM and HOG feature extraction technique. During the next phase of the study, Random Forests were found to record the highest accuracy when GLCM features were combined with Hu moments and HSV color features. During the final phase of the study, Random Forest was found to record the highest accuracy when GLCM, HOG, Hu and HSV color features were used for classification. The authors concluded that the combination of GLCM, Hu and HSV reported a significant enhancement in the accuracy (89.75%) when compared with the existing schemes.

Praveen Kumar Moganam et at. [5], proposed multiple approaches for multiclass leather texture detection and segmentation. They used a collection of 3600 leather images from 6 different classes. The images belonged to different classes of leather surfaces namely, growth marks, loose grains, grain off, folding marks, pin hole and non-defective surfaces. The images were subjected to Histogram Equalization during image acquisition process. These images were later resized into 227x227 dimensions. The leather images were then subjected to color and texture feature extraction using GLCM, Autocorrelation and Color Histogram. The pre-processed images where then fed into the existing machine learning models like the Shallow Feed Forward Neural Network, Support Vector Machine and K-Nearest Neighbors for classification. The images were also classified using pre-trained deep learning models like AlexNet, VGG16, GoogleNet, SqueezeNet and ResNet50. Upon analyzing the confusion matrices, it was found that the AlexNet was found to perform better than all the Shallow Feed Forward Neural Network recording an accuracy score of 99.4%. The authors also demonstrated localizing the defective regions of the surface using the class activation maps of the trained deep neural network. Swamiraj Nithiyanantha Vasagam et al. [6], proposed a methodology that enabled the classification of variations among the Defects and the Non-defects in intermittent leather images. The study involved using a total of 428 leather images. While processing the images, the images were resized, then filtered using Erosion and Dilation operations. These images were then subjected to image segmentation procedure. The study dealt with the extraction of 11 texture features. Three of the features were extracted through Visual Inspectors and histogram analysis, while the remaining eight features were extracted using GLCM. A set of features from the entire collection of features were selected for further analysis. The selected features were then fed into LDA and SVM for classification task. The proposed LDA was found to record a classification accuracy of 92%, while SVM on the other hand recorded a classification accuracy of 89.65%. It was concluded that LDA was found to be relatively more successful at classifying the variations among the defects and non-defects in the intermittent leather images.

Salik Ram Khanal et al. [7], proposed a system of hardware and software platforms to enable leather defect detection using Semantic Segmentation. They proposed a Hardware Prototype to ease the capturing of leather surface images. The captured images were then subjected to image pre-processing procedures followed by application of deep learning algorithms for semantic segmentation and object detection. For experimental purposes, they used the MVTECT leather anomaly database. The images from the database was cropped and resized to 256x256 dimensions. The authors also applied Image Augmentation to increase the number of images. MobileNet and ResNet50 feature extraction models were used in along with UNet and SegNet semantic segmentation models in various combinations during the study. The proposed model was found to achieve an IOU value of more than 95% in 5 class segmentation. Zhongliang Zhang et at. [8], proposed a leather segmentation network to enable improved accuracy in detection of wet leather defects. They used a collection of 1944 (1603 defective and 524 non-defective) wet leather surfaces for the study. The proposed segmentation network uses Kronecker products for the construction of a new semantic information extraction layer. The layer in added to the decoding network of the encoder-decoder network to form new decoding paths. These paths are then used to segment the defective regions of the leather image. The proposed network displayed an improvement of 1.99% in the F1 score compared to UNet for wet leather defects and three times improvement in detection speed.

Malathy Jawahar et al. [9], proposed an inspection system that facilitates the detection of Leather Surface Defects. The acquired images were initially subjected to Image Pre-processing procedures. While pre-processing, the images were resized followed by application of Wiener Filter to enhance the image quality. Fast Convergence Particle Swarm algorithm was later applied on these images. The features extracted from these images were then fed into the neural network to facilitate the classification task. A set of 12 texture features were given as input to the neural network. The performance of the neural network was tested on each of the 10 defect classes of the leather image dataset. The classification of three defects namely the Pox Mark, Scratch and Lime Blast defect was found to record an accuracy of more than 90%. The overall accuracy, specificity, sensitivity, and precision to classify all the leather defects recorded 88.61%, 91.62%, 78.23% and 93.96%. Aashish Ghimire et al. [10], proposed a methodology for leather defect detection using Semantic Segmentation. The authors used MVTECT Leather anomaly dataset for the study. The images were initially resized to 256x256 and then finally converted into its corresponding Gray Scale equivalent. The authors then proposed four segmentation models to facilitate the segmentation of leather images. The four algorithms used in this study correspond to, combination of MobileNet and UNet, combination of ResNet50 and UNet, combination of MobileNet and SegNet and combination of ResNet50 and SegNet. The most optimal results were obtained for the combination of MobileNet and UNet models recording a mean IOU value of 72.10% and mean F1 score of 82.59%.

M. Sornam et al. [11], used K-Means clustering algorithm to segment the leather images. Upon experimentation it was determined that value corresponding to six clusters recorded the highest. This indicated the ideal number of clusters for segmenting the leather images. They also determined the perimeter and the area occupied by the defect in the leather surface. Harshal Piwal et al. [12], proposed a methodology to detect steel surface defects using VGG models. They used North Eastern University (NEU) dataset sources to develop the database for this process. The study tested the performance of VGG16 and VGG19 models on the defect dataset. It was found that VGG19 recorded an accuracy of 97.2%, while VGG16 recorded an accuracy of 93.3%. However, it was noted that VGG16 model was found to have a relatively lesser computation time.D. Soukup et al. [13], proposed an algorithm to detect defects on rail surfaces using photometric images. They used a small image dataset for this study. It was found that problem of smaller dataset could be removed using data augmentation and unsupervised layer-wise pretraining regularization methods. The inclusion of these regularization methods was found to improve the performance of the proposed methodology. They could also show that the efficiency of detecting the defects using CNN was better than the other existing model-based approaches. Honglin Xiong et al. [14], proposed an approach to detect glass surface defects using deep learning. They proposed an image recognition model based on Multiscale Convolution Neural Network (MCNN). They showed that the proposed MCNN model was found to perform better than the traditional CNN methods. They therefore concluded that the proposed model was found to have a superior recognition accuracy of glass defect when compared with the existing CNN recognition methods. Tao Zhang et al. [15], proposed a system for brake pad surface defect detection. They initially developed a mechanism to acquire images, thereby enabling the construction of the brake pad image dataset. They then developed a CNN model for defect detection and classification task. Later, a Fully Convolutional Network (FCN) was also proposed to further improve the performance of the detection system. It was found that CNN had better detection efficiency when compared with FCN. However, FCN was found to display better recognition accuracy when compared with CNN.

Liao D et al. [16], proposed an algorithm to detect and classify the defects on the Si_3N_4 turbine blades. They have used a Si_3N_4 turbine blades dataset for this study. They have used the concepts of both CNN and YOLOv5 to develop the proposed algorithm. They replaced the PAN and FPN structure of YOLOv5 with the BiFPN structure in order to achieve higher levels of feature fusion. This resulted in an accuracy of 97.4%. It is also interesting to note that the detection speed of the algorithm is 16ms. Xinghui Dong et al. [17], proposed an approach to detect small defects using U-Net and Random Forest Classifier. The features from the images were extracted using U-Net, while the Random Forest was trained to classify every pixel in the image. For defect identification, they used Maximally Stable Extremal Region (MSER). They found that their approach produced relatively superior results when compared with the other alternate methods tested by them. Zhong Zhang et al. [18], proposed a methodology to detect defects on Specular Reflection Surface. They proposed an ensemble model by combining two different neural networks to detect the defects present on the surface. The model achieved an accuracy of 97% on the test set, which was found to be 2% higher than single neural network model. Pallavi S. Chandanshive et al. [19], proposed a system to automate the surface defect detection in Hot rolled steel strip. They used the NEU database in this study. The proposed CNN model was used to detect and also classify the defects into its respective categories. The proposed network could achieve an accuracy of 98.9%.K. Sunitha et al. [20], proposed an algorithm for metallic surface defect detection.GC-10 dataset is used in this study. It has 10 different classes of surface defects. They preprocessed this image dataset by performing Data Annotation and Data Augmentation. A neural network model loaded with VGG16 is then used for learning. Mean Squared Error (MSE) is used for evaluating the error during the learning process. An accuracy of 97% was recorded by the authors using the proposed model in this study. Yanghuan Xu et al[21]., proposed a system for surface defect detection using Deep Learning. In this study, the authors used an image dataset having 8 different classes of surface defects. The defect recognition model based on Efficient Net recorded an accuracy of 93.05%. The performance of this model was found to be better than VGG16, MobileNetV2 and ResNet50. Limei Song et al[22]., proposed a methodology for detection of micro-defects on metallic screws. The images of the metallic screws were first captured using cameras. They then located the screw surfaces from these captured images. The extracted images were then fed into a CNN model. The model was found to record an accuracy of 98%. Upon comparison with traditional machine learning algorithms, they found that the proposed architecture had resulted in a better performance.

As observed, the various works undertaken in the field of defect analysis corresponds to defect segmentation, defect identification, defect recognition and defect classification. Research on defect analysis in various domains like steel defect analysis, glass surface defect analysis, leather defect analysis, etc., has been carried out by different authors using CNN networks, segmentation networks, etc. When leather defect analysis is specifically taken into consideration, several research activities on leather defect detection and classification, leather defect segmentation has been carried out.

Table 1 summarizes all the datasets used and the algorithms/models employed in the state of the art.

Table 1 — Summary of the dataset used, and the method applied in the existing works.

Author	Dataset	Model/Method applied
Swamiraj Nithiyanantha Vasagam et al.[2]	Crust Leather Image Dataset	Support Vector Machine
Y.S. Gan et al.[3]	1600 Calf Leather Patches	ANN and XBoost ANN
Malathy Jawahar et al.[4]	577 Leather Defect and 100 Leather Non-Defect Images	Random Forest
Praveen Kumar Moganam et al.[5]	3600 Leather Images	AlexNet
Swamiraj Nithiyanantha Vasagam et al.[6]	428 Leather Images	LDA
Salik Ram Khanal et al.[7]	MVTECT Leather Anomaly Database	Various combinations of MobileNet, ResNet50 and UNet and SegNet
Zhongliang Zhang et al.[8]	1944 Wet Leather Surface Images	Leather Segmentation Network
Malathy Jawahar et al.[9]	Leather Surface Image Dataset	Neural Network
Aashish Ghimire et al[10]	MVTECT Leather Anomaly Database	Various combinations of MobileNet, ResNet50 and UNet and SegNet
M. Sornam et al[11]	Leather Image Dataset	K Means Clustering Algorithm
HarshalPiwal et al.[12]	NEU Dataset	VGG19
D. Soukup et al.[13]	Photometric images of Rail surfaces	CNN
HonglinXionget al.[14]	Glass Surface Defects Image Dataset	Multiscale Convolution Neural Network (MCNN)
Tao Zhanget al.[15]	Brake Pad Surface Defect Image Dataset	CNN, FCN
Liao D et al.[16]	Si_3N_4 turbine blades dataset	CNN + YOLOv5
Xinghui Dong et al.[17]	Defect Image Dataset	U-Net and Random Forest
Zhong Zhang et al.[18]	Defects on Specular Reflection Surface	Ensemble Model
Pallavi S. Chandanshiveet al.[19]	NEU Dataset	CNN
K. Sunitha et al[20]	GC-10 Dataset	Neural Network Model loaded with VGG16
Yanghuan Xu et al[21]	Surface Defect Image Dataset	EfficientNet
Limei Song et al[22]	Metallic Screw Image Dataset	CNN

The major contributions of this paper while meeting the primary objective of designing the most efficient and highly reliable AI based system for the classification of leather surfaces into defective and defect free are as follows:

- This study aims at developing an AI based system for detection and classification of leather surfaces into defective and defect free using a leather image dataset having two classes: defects and non-defects.
- The study involves augmentation of available leather image data to facilitate expansion of number of data instances for the development of a more universal AI based system in order to simulate real time leather images captured during different industrial applications for better performance.
- The study has encountered a challenge of generating satisfactory detection and classification results based on traditional machine learning algorithms due to huge variability in the kinds of leather surface and kinds of defects. Hence, the development of novel lightweight neural network architecture for the generation of an AI based system for the detection and classification of leather surfaces into defects and non-defects has been explored.
- The authors developed *LeatherNet,* a lightweight neural network for the detection and classification of leather surfaces into defects and non-defects. The unique feature of the proposed architectural design is that it is shallower and has limited number of parameters. This architecture has been found to generate outstanding results with maximum training accuracy, precision, and recall recording of 99.78%, 99.69% and 99.92% respectively, with maximum testing accuracy, precision, and recall recording of 97.42%, 97.66% and 99.40% respectively.
- To gain better insights into the performance of *LeatherNet*, the performance of *LeatherNet* was compared with that of some of the prominent existing CNN architectures like VGG16, VGG19, ResNet50 and EfficientNetB0. The *LeatherNet* has been found to perform better than the experimented prominent existing architectures as well as the state of the art.
- Thus, the proposed *LeatherNet* has successfully been validated to be a highly reliable as well as efficient AI model for the detection and classification of leather surfaces into defects and non-defects.

2 Materials and Methods

2.1 Dataset description

This manuscript uses an image dataset having images of different types of leather surfaces. This dataset is custom and solely available with CSIR-CLRI having a total of 384 leather images to meet the primary objective of designing and developing a deep learning model that could automate the detection of defective leather surfaces. The images in the dataset primarily belong to 2 classes, corresponding to defective and non-defective leather surfaces having 283 images and 101 leather images respectively. The dataset has images from huge variety of leather surfaces. Figure 1 shows the structure of the dataset.

The original leather image dataset is observed to have limited number of leather images. Hence, we perform offline data augmentation on the given image dataset to increase the size of image dataset. While performing offline data augmentation, the rotational range, shearing range, width shift range and height shift range is set to 40, 2, 0.2 and 0.1 respectively. Authors also enabled horizontal and vertical flips. The brightness range is set between 0.85 and 1.25. This would enable generation of images under various lighting conditions. We also enabled ZCA whitening parameter in the data augmentation process. Such a wide range of parameters are set to enable the generation of leather images under different conditions. This would enable the development of a more generic and thus better AI system for the detection and classification of leather surfaces into defects and non-defects. The finally obtained image dataset resulted in a total of 6000 images from each class (12000 in total). This leather image dataset is

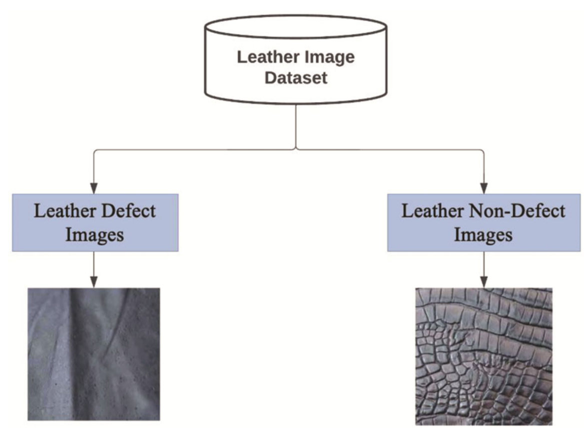


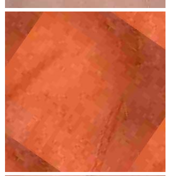
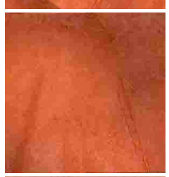
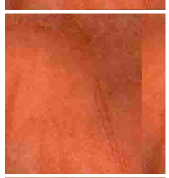
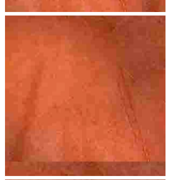
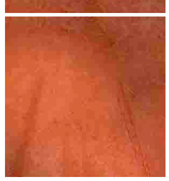
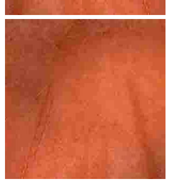
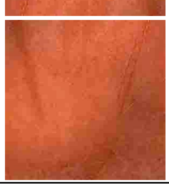
Fig. 1 — Dataset structure.

then used further in this manuscript. Table 2 displays the images obtained after the application of each individual parameter during data augmentation.

2.1 Preprocessing

The augmented dataset with 6000 images from each class is taken up further in this manuscript. Preliminary observations on the dataset revealed that the images present in the dataset were colored and had a dimension of (300, 300, 3). These images were subjected to certain basic pre-processing procedures to enhance the quality of the images present in the dataset.

Table 2— Samples of image variety generated (Data Augmentation).

Image Type/Parameter	Image
Actual Image	
Rotational Range	
Shearing Range	
Width Shift Range	
Height Shift Range	
Brightness Range	
Horizontal Flip	
Vertical Flip	

Several machine learning algorithms were explored during the first phase of the investigation. While preprocessing, the images were initially converted into their corresponding gray scale equivalent images. These images were then denoised using Gaussian Blur to remove the noises from them. For denoising with Gaussian Blur, the Sigma X, Sigma Y, and Kernel sizes were set to 90, 90, and (3,3), respectively. Denoising was found to generate better results under these parameters when experimented on a few test sample images. These images were then subjected to Histogram Equalization, which improved the overall contrast of the images. We then attempted to recognize the edges in these images to improve the model learning. Experimentation revealed that the Prewitt edge detection method detects edges better than other edge detection techniques. As a result, Prewitt Edge Detection was applied to the images. The images were then resized to a (64,64) configuration. This reduction in size of the images enabled smoother processing of many images in the working environment. These processed images were then fed into various Machine Learning algorithms.

During the next phase of this manuscript, the images were fed into various deep learning architectures. While pre processing the images, they were initially subjected to the process of denoising. This enabled the removal of noises from these leather images. Denoising using Gaussian Blur with SigmaX, SigmaY and Kernel size parametric values set to 90, 90 and (3,3) respectively was found to generate better results when experimented on a few test sample images. Hence, in this manuscript, authors used the above said parameters to denoise all the images in the image dataset. These denoised images were then downsized to (64,64,3) configuration. This reduction in the image size from (300, 300, 3) to (64,64,3) enabled smoother processing of images. These resized images were later normalized. The normalized images were finally fed into the various deep learning architectures taken up during this work. Figure 2 elaborates the preprocessing approaches taken during the different phases of the work.

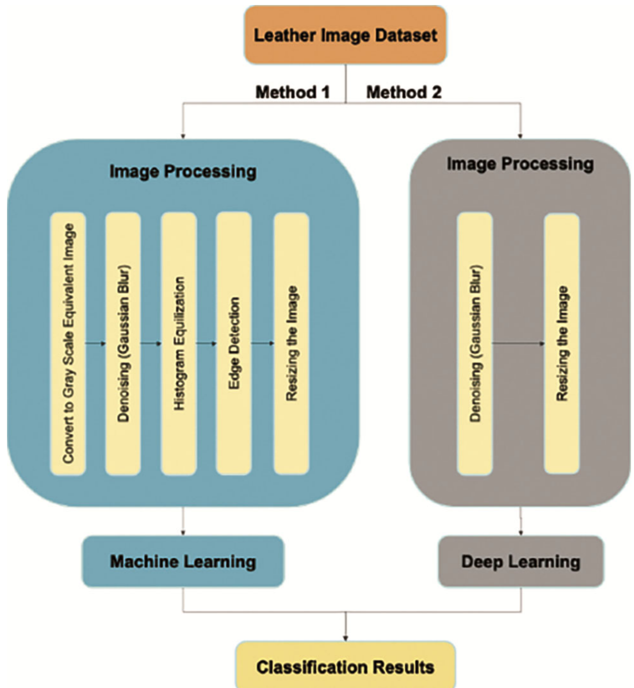


Fig. 2 — Proposed Workflow for Pre processing.

2.2 Approaches

The authors primarily divided the work into two different phases. During the first phase of the work, exploration of various machine learning models like Logistic Regression (LR), Naïve Bayes (NB), K Nearest Neighbors (kNN), Decision Trees (DT), Random Forest Classifier (RF) and Support Vector Machine (SVM) for detection and classification of leather images into defect and non-defect were carried out. While in the second phase of the work, the impacts of deep learning models on the same were explored. It is pertinent to note that, it is during this phase of the work a novel Convolution Neural Network (CNN) architecture for detection of defective and non-defective leather surfaces was proposed. To gain better insights into the performance of the proposed architecture, certain prominent existing CNN architectures like VGG16, VGG19, ResNet50 and EffecientNetB0using transfer learning were also taken into consideration in this work. Figure 3 summarizes the models/algorithms that were taken into consideration in this work.

2.1.1 Machine learning approaches

Singh, Y. K., et al., found that logistic regression was able to generate better results when compared with the other experimented algorithms. Similarly, Kumar, P. R., et al., found that Naïve Bayes classifier was able to quite accurately able to assess the level of deformation of thin

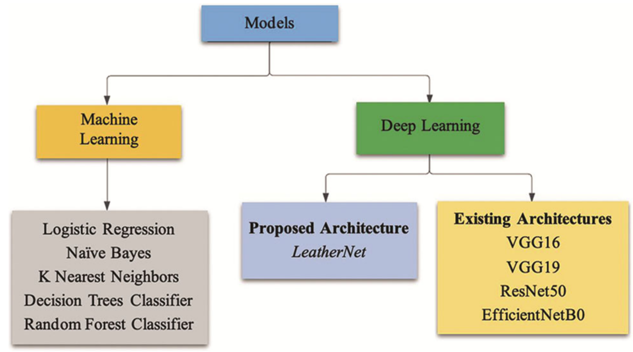


Fig. 3 — Models experimented in this work.

components and perform image classification. Likewise, Patidar, D. et al., experimented the application of kNN algorithm for image-based classification task. Alassar, Z. I on the other hand, used Decision tree classifier for the task of image classification. This algorithm was found to generate better results than the other experimented algorithms. Similarly, Jawahar, Malathy, L. et al., were able to generate good classification results using random forest. As observed, several conventional machine learning algorithms were able to generate quite impressive results in different domains involving image classification task. Hence, we experiment the performance of the conventional machine learning algorithms on the given dataset. The various machine learning algorithms explored during this phase of work for the image classification task include Logistic Regression, Naïve Bayes, K Nearest Neighbors, Decision Trees Classifier and Random Forest Classifier.

Logistic Regression: It is a type of statistical model that is used to facilitate classification. It basically estimates the parameters of a statistical model that models the probability of an event by having the log odds of an event be a linear combination of one or more independent variables.

Naïve Bayes: It is a supervised machine learning algorithm used for classification. The algorithm utilizes the concepts of Bayes theorem with strong independence assumptions among the features.

K Nearest Neighbors: It is a non-parametric supervised machine learning algorithm. This algorithm uses the concepts of proximity or similarity to facilitate the classifications of data points.

Decision Trees: It is a supervised machine learning algorithm used for classification of data points. It is a tree like model that has a root node where the decisions are made, branches where the possible action paths are mapped and the leaves where the potential decision outcomes are defined.

Random Forest Classifier: It is a supervised machine learning model that operates by constructing a multitude of decision trees for the classification tasks. In classification task, the class selected by most of the trees in a random forest is the output class of the random forest.

2.11.1 Deep Learning Approaches

2.11.1.1 Proposed Architecture – LeatherNet

LeatherNet– a lightweight neural network architecture is proposed to automate the detection and classification of leather surface images into defect and non-defect surface images. The architecture is designed with an intension of enhancing the ability to detect and classify the leather surfaces into its respective categories, when compared with the other existing state of the art works and the above experimented conventional machine learning models.

LeatherNet is designed to be lightweight so that minimal computational resources can be used to develop and utilize this AI based system. Itis designed to be lightweight by making the neural network architecture less deep and thus have relatively lesser number of parameters. This designing is done without compromising the performance of the proposed *LeatherNet* architecture.

The proposed *LeatherNet* architecture has a series of convolution layers, max pooling layers for image feature extraction and fully connected layers for image classification task. The architecture uses sigmoid activation function to facilitate the classification of the image dataset.

The *LeatherNet* architecture is designed to input images having image dimensions same as (64,64,3). The first layer in the architecture corresponds to a convolution layer followed by max pooling layer. This convolution layer has 16 filters with the kernel size defined as (3,3), while the max pooling layer has a pool size of (3,3). The next two layers in the proposed architecture corresponded to convolution layers having 32 and 64 filters respectively. The output from this layer is then given as an input to a convolution layer having 128 filters with the kernel size defined as (3,3). The output from this convolution layer is then fed into the max pooling layer. The features map obtained because of the above layers is flattened before finally feeding the same into a sequence of fully connected layers to facilitate classification. Sigmoid classifier is used in the final dense layer to successfully classify the image into one of the two possible output classes. The architecture primarily uses ReLU activation function and Adam Optimizer. It also uses L2 Regularizers facilitate better learning. Figure 4 shows the pictorial representation of the proposed *LeatherNet* architecture. Table 3 summaries the hyper parameters used while designing this architecture.

2.11.1.2 Transfer Learning

The authors corresponding to the papers [27-30], could display the generation of reliable results by applying the concepts of transfer learning in various domains for image classification task. To better analyze the performance of the proposed deep neural network, its performance is compared with other existing prominent deep learning architectures. Hence, we apply concepts of transfer learning on the models like VGG16[27], VGG19[28], Res Net50[29] and EffecientNetB0[30] for the given leather image dataset. Authors have used ImageNet weights in this process.

VGG16: It is prominent CNN architecture developed by K. Simonyan and A. Zisserman in 2014. This architecture is a 16 layers deep neural network having 13 convolution layers and 3 fully connected layers. The use of smaller 3x3 receptive fields in its entire network makes it different from the then existing prominent architectures like AlexNet. The presence of two non-linear activation functions makes its decision functions more discriminative in nature. In this work, sigmoid classifier was used to facilitate the classification of leather surfaces.

VGG19: It is another prominent CNN architecture developed by Simonyan and Zisserman. This

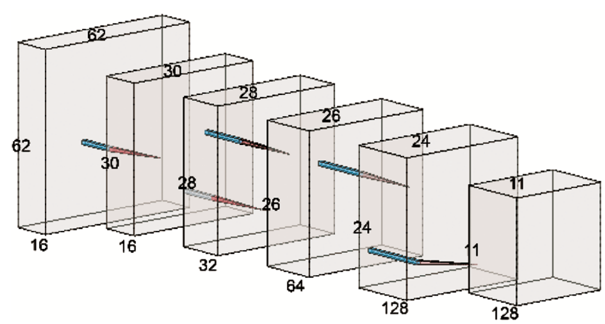


Fig. 4 — *LeatherNet* architecture (Convolution Layers).

Table 3 — Specifications of Hyperparameters of the Proposed *LeatherNet* architecture

Hyperparameters	Values
Number of Convolution Layers	4
Number of Max-Pooling Layers	2
Activation Function	ReLU and Sigmoid
Optimization Function	Adam
Regularizers	L2

architecture is a 19-layer deep neural network having 16 convolution layers and 3 fully connected layers. It uses (3,3) filters with padding and strides set at 1 along with (2,2) max-pooling layers. In this work, sigmoid classifier was used to facilitate the classification of leather surfaces.

ResNet50: It is yet another prominent CNN architecture developed by Kaiming et al in 2015. It is a 50-layer deep neural network structure. The presence of Skip Connections helps it overcome a major problem of vanishing gradient faced by most of the Convolution Neural Networks. In this work, sigmoid classifier was used to facilitate the classification of leather surfaces.

Efficient NetB0: It is another prominent CNN architecture used in this work. It is a 237-layer deep architecture. This is the deepest architecture used in this work. It is also an architecture that can successfully perform the task of image classification.

2.12 Performance Measuring Metrics

The work primarily uses the accuracy, the precision and the recall metric to measure and compare the performances of various machine learning and deep learning algorithms used during the work.

Accuracy is defined as the ratio of number of correct predictions to total number of predictions made by the model as specified in eq 1.

$$Accuracy\ \% = \frac{True\ Positive\ (TP)\ +\ True\ Negative\ (TN)}{True\ Positive\ (TP) + False\ Positive\ (FP)\ +\ True\ Negative\ (TN) +\ False\ Negative\ (FN)}\ x\ 100 \quad \ldots (1)$$

Precision on the other hand is defined as the ratio of correctly classified positive instances to the total number of classified positives instances as specified in eq 2.

$$Precision\ \% = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)}\ x\ 100 \quad \ldots (2)$$

Further, recall is defined as the ratio of correctly classified positive instances to the total number of ground-truth positive instances as specified in eq 3.

$$Sensitivity/Recall\ \% = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) +\ False\ Negative\ (FN)}\ x\ 100 \quad \ldots (3)$$

However, it is pertinent to note that this work also uses loss metric to help understand the learning curve of the algorithm in a better way.

Figure 5 depicts a pictorial representation of the overall workflow undertaken in this work.

3 Results and Discussion

The models developed in this work were run in the Google Colab environment. The work utilized the power of GPU provided by the Google Colab environment to facilitate faster processing of such large amounts of image data.

Several machine learning and deep learning python libraries were used in various phases of the work to make the entire process more efficient. Image Generator was used during the Data Augmentation process to increase the number of data instances. Further, certain inbuilt functions from Scikit-learn were used while dealing with machine learning algorithms. Additionally, libraries like Tensor flow and Keras were used while dealing with deep learning architectures. It is pertinent to note that libraries like Open CV, Numpy, Matplotlib among others werealso used in various phases of thiswork.

During the first phase of the work, several conventional machine learning algorithms were taken into consideration for automating the detection and classification of leather surfaces into its respective categories. However, the performance of the machine learning algorithms was found to be poor for industrial applications. Hence, in the next phase of the work, application of deep learning algorithms was explored for automating the detection and classification of leather surfaces. In this process, *LeatherNet,* a novel CNN architecture was designed to enhance the performance of automating the detection and classification of leather images into defects and non-defects. While designing the neural network, we also took an attempt to propose an architecture that was shallower andhad lesser number of parameters. This enabled the development of lightweight neural network architecture. This also facilitated faster learning in theneural network. Thus, during this process, several iterations of parameter tweaking were performed to obtain the most optimal architecture which is both shallower and had lesser number of parameters. This finally resulted in the proposed architecture. To further gain better insights into the performance of the *LeatherNet*, multiple existing prominent CNN architectures using Transfer Learning were also taken into consideration.

During both the first and the second phase of the work, that corresponded to dealing with machine learning and deep learning models (*LeatherNet* and prominent existing architectures), the images were subjected to respective pre-processing techniques as

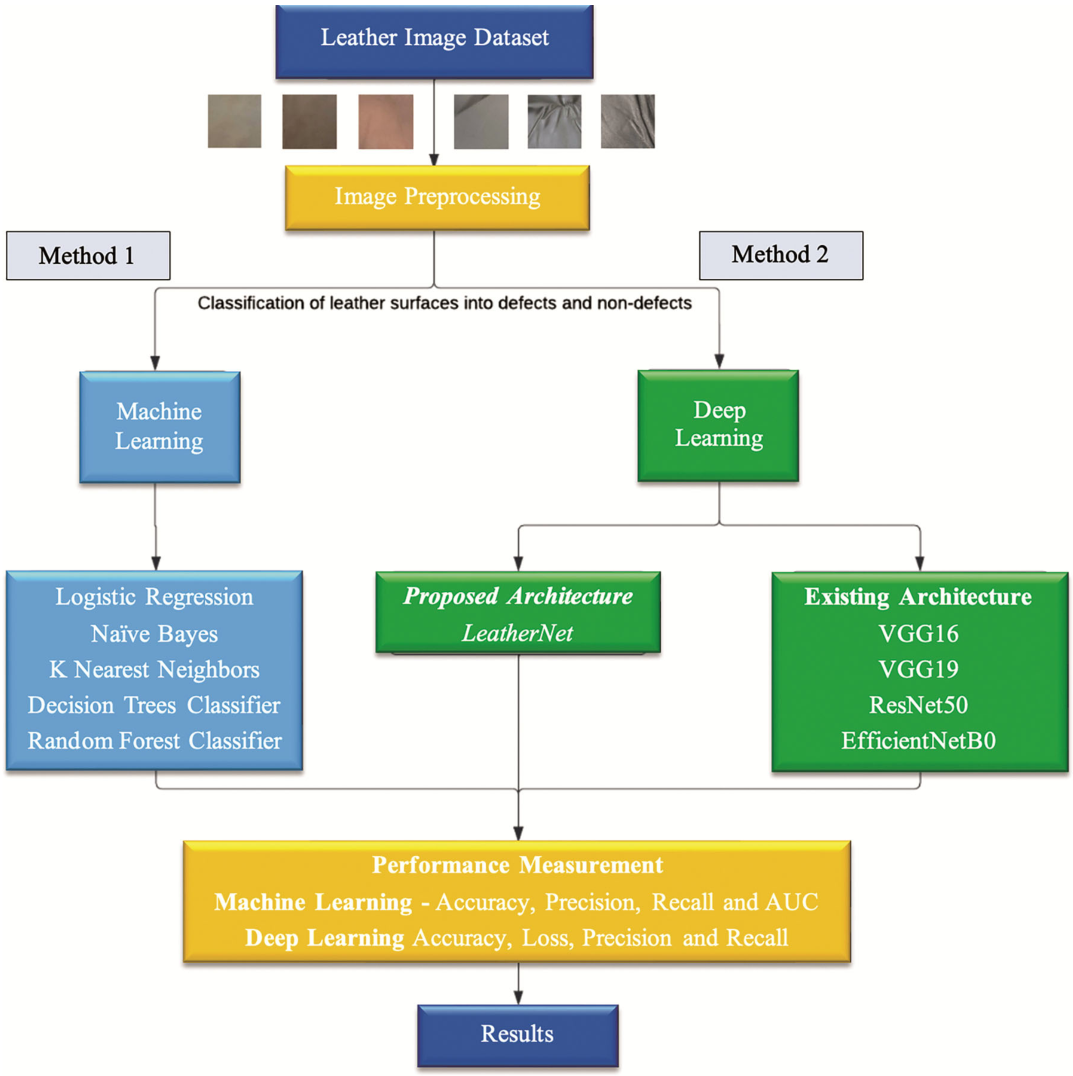


Fig. 5 — Overall Workflow.

stated earlier. After performing initial operations on the images, the image dataset was then split in 80:20 training and testing image set ratio. This resulted in 9600 training images and 2400 testing images. The images present in the training set was then used for training the models, while the images in the testing set were used to determine the performance of the model on unseen data instances. The performance of the model was quantified using several performance metrics as discussed earlier.

In this section, we discuss and summarize the results that were obtained on application of varies machine learning and deep learning algorithms on the leather image dataset.

Table 4 summarizes the performances of the various machine learning models that were trained on a subset of the entire dataset.

Table 4 — Performance Comparison among various machine learning

Machine Learning Models	Accuracy	Precision	Recall
Logistic Regression (LR)	53.29%	53.41%	51.41%
Naïve Bayes (NB)	52.66%	52.81%	50%
K Nearest Neighbors (KNN)	53.625%	53.86%	50.58%
Decision Trees Classifier (DTC)	53.625%	53.85%	50.85%
Random Forest Classifier (RFC)	57.79%	59.06%	50.75%

As quite evident from the above Table 4, the highest accuracy was attained by Random Forest Classifier, recording 57.79%. This is followed by K Nearest Neighbors and Decision Tress Classifier, both recording 53.625% accuracy. These are followed by Logistic Regression and Naïve Bayes recording an accuracy of 53.29% and 52.66% respectively. In this context it is pertinent to note that all the ML models except Random Forest Classifier recorded accuracies between 52% and 54%. The Random Forest Classifier recorded the highest precision of 59.06%. While the remaining machine learning algorithms recorded precision values between 52% and 54%. However, the highest recall was recorded by Logistic Regression. It is also pertinent to note that recall values recorded by all the machine learning algorithms lie in the range of 50% to 52%. In addition to accuracy, precision and recall metrics, AUC score was also calculated. The highest AUC value was found for Random Forest Classifier recording 0.60. This is closely followed by Logistic Regression recording an AUC value of 0.57. While the AUC score of the remaining machine learning models was found to be in the range 0.50 to 0.53. The ROC curves with the corresponding AUC scores are displayed in Figure 6.

As observed, Random Forest gives better results when compared to the other machine learning algorithms. This is mainly due to its ability to combine the results of multiple classifiers for better performance. However, it is noted that the performance of all the machine learning models is not quite satisfactory, indicating their inefficiency in detecting and classifying the leather surfaces into defect and non-defects. This inefficiency is mainly attributed to their inability in properly identifying the leather surface defects. This might be mainly due to the huge variability in the type of leather surfaces and the kind of leather defects (ranging from physical defects to textural defects) that has been taken into consideration. Hence, designing a single machine learning algorithm compatible to all kinds of leather surfaces and leather defects for the purpose of defect detection might be difficult to develop. This motivated us to explore the development of deep learning architectures for the detection and classification of leather surface images with a greater efficiency for all types of leather surfaces and leather defects. Hence, a novel deep learning architecture namely, *LeatherNet* was designed and constructed.

Table 5 summaries the performance of *LeatherNet* architecture trained on the leather image dataset. The architecture was trained for 1000 epochs, 1250 epochs

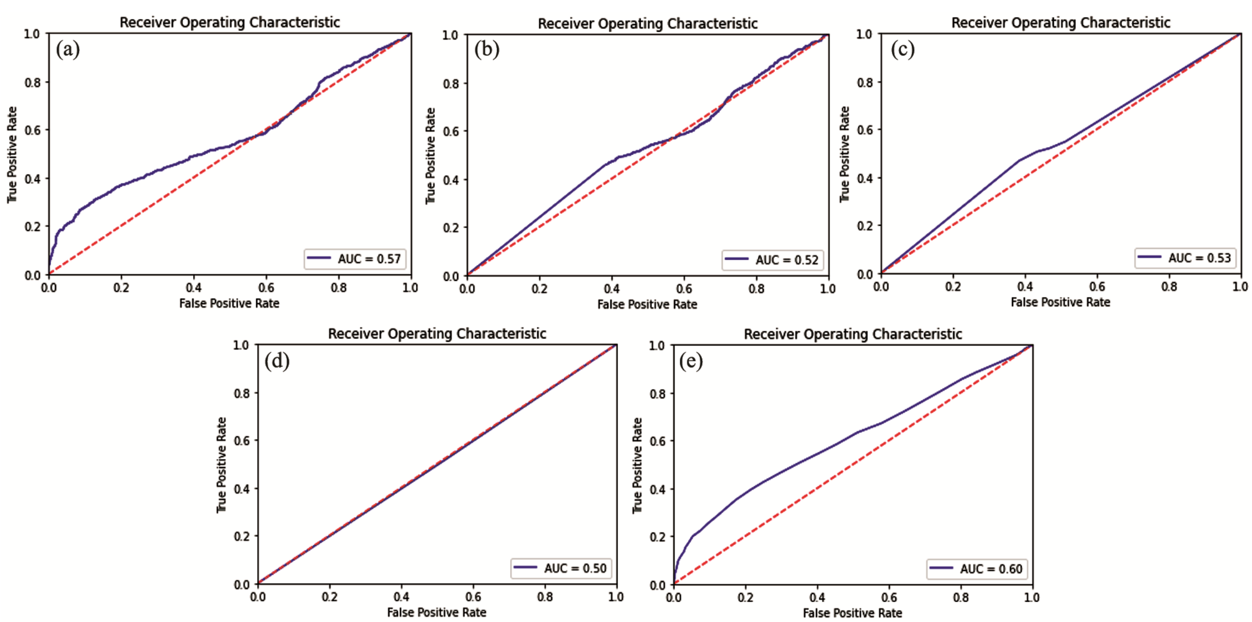


Fig. 6 — ROC Curves (for Machine Learning Models).

Table 5 — Performance of the proposed *LeatherNet*

	Epochs	Accuracy	Loss	Precision	Recall
Training	1000	99.11%	0.0753	98.95%	99.23%
Testing		97.04%	0.1512	97.38%	98.88%
Training	1250	99.80%	0.1420	99.77%	99.85%
Testing		97.33%	0.0453	98.47%	98.97%
Training	1500	99.78%	0.0328	99.69%	99.92%
Testing		97.42%	0.1464	97.66%	99.40%

and 1500 epochs. The results of the same are summarized as follows.

The proposed deep learning network, *LeatherNet,* was trained on the leather image dataset having 6000 images from each of the two classes. The proposed model was trained on 80% of the image dataset while the remaining 20% of the data was used for testing the model performance. The model was trained for 1000, 1250 and 1500 epochs.

The model when trained for 1000 epochs recorded a maximum training accuracy, maximum training precision and maximum training recall values of 99.11%, 98.95% and 99.23% respectively. While the model on the other hand, recorded a maximum testing accuracy, maximum testing precision and maximum testing recall values of 97.04%, 97.38% and 98.88% respectively. The training loss and validation loss after training the model for 1000 epochs recorded the lowest values of 0.0753 and 0.1512 respectively.

In the next stage, the model was again trained for 1250 epochs. This resulted in a maximum training accuracy, maximum training precision and maximum training recall values of 99.80%, 99.77% and 99.85% respectively. The corresponding maximum testing accuracy, maximum testing precision and maximum testing recall values recorded 97.33%, 98.47% and 98.97% respectively. The training loss and validation loss after training the model for 1250 epochs recorded the lowest values of 0.1420 and 0.0453 respectively.

When the number of training epochs was increased from 1000 to 1250, an improvement in the performance of the proposed deep neural network was observed. Hence, in the next stage a further increase in the number of epochs to 1500 was taken into consideration.

In the next stage, the model was again trained for 1500 epochs. This time the model recorded a maximum training accuracy, maximum training precision and maximum training recall values of 99.78%, 99.69% and 99.92% respectively. The corresponding maximum testing accuracy, maximum testing precision and maximum testing recall values recorded 97.42%, 97.66% and 99.40% respectively. The training loss and validation loss after training the model for 1500 epochs recorded the lowest values of 0.0328 and 0.1464 respectively.

As evident, with an increase in the epochs, the model was found to record better validation accuracy and validation recall despite negligible decrease in the validation precision. Overall, the proposed model was found to generate very good results. Generation of good results on the validation set, indicated its superior efficiency in the detection of leather defects. This development of such a high-performing model, *LeatherNet*, resulted in the evolution of a highly reliable artificial intelligence-based system which could potentially also be deployed for industrial application. The accuracy, loss and precision graphs of the proposed architectures are shown in Fig. 7, Fig. 8 and Fig. 9 respectively.

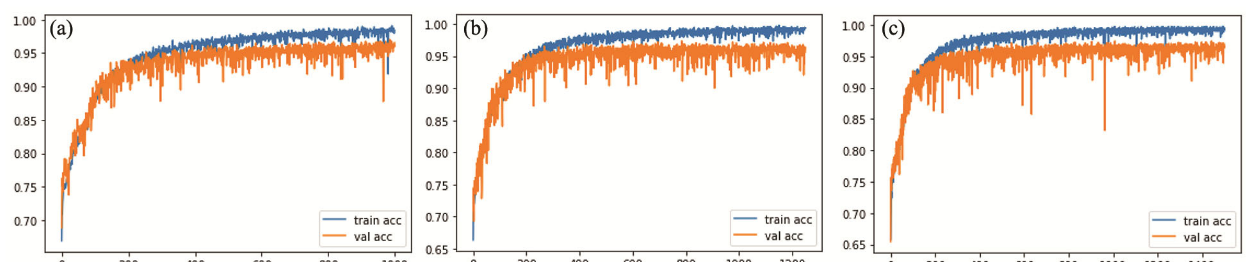


Fig. 7 — Accuracy graphs of *LeatherNet* trained for (a)1000, (b) 1250, and (c)1500 epochs respectively.

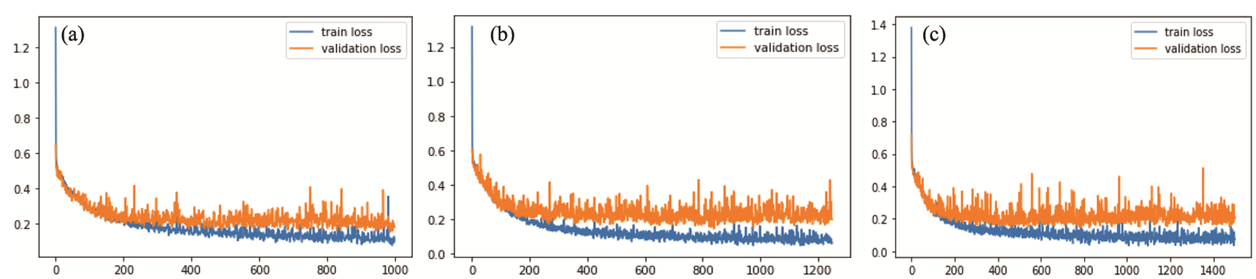


Fig. 8 — Loss graphs of *LeatherNet* trained for (a)1000, (b)1250, and (c)1500 epochs respectively.

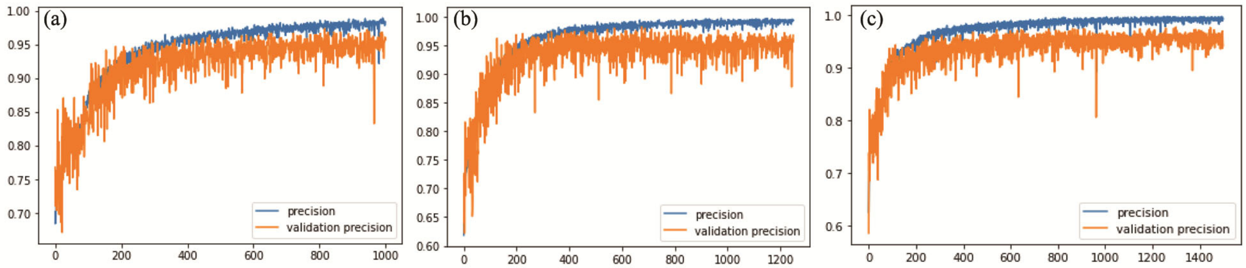


Fig. 9 — Precision graphs of *LeatherNet* trained for (a)1000, (b)1250, and (c)1500 epochs respectively.

Table 6 — Performance Comparison for Existing Prominent Architectures

Model		Accuracy (%)	Loss	Precision (%)	Recall (%)
VGG16	Accuracy	78.19	0.4821	77.54	91.34
	Validation Accuracy	81.41	0.4385	89.98	99.24
VGG19	Accuracy	76.64	0.5213	78.09	76.04
	Validation Accuracy	78.50	0.4867	90.87	98.14
ResNet50	Accuracy	50.44	0.6931	50.45	100
	Validation Accuracy	50.70	0.6931	49.29	100
EffecientNetB0	Accuracy	50.24	0.6931	50.28	100
	Validation Accuracy	50.70	0.6931	49.29	100
Proposed Architecture	Accuracy	99.78	0.0328	99.69	99.92
(LeatherNet - at 1500 epochs)	Validation Accuracy	97.42	0.1464	97.66	99.40

It is pertinent to note that the above stated results are obtained when the proposed *LeatherNet* neural network was trained using Adam optimizer. However, to gain better insights, the study also considered using RMSProp optimizer for training the proposed *LeatherNet* model. To begin with, the model was first trained for 1000 epochs using RMS Prop optimizer. During training, it was observed that the performance of the *LeatherNet* model improved until it reached a maximum training accuracy of 84.48%. However, it was noted that further training resulted in a decrease of both training and validation accuracy. Hence, it was noted that in similar working conditions, the Adam optimizer was found to perform better than the RMS Prop optimizer. Thus, it was concluded that the proposed model, *LeatherNet*, displayed good results when the Adam optimizer was used for training the neural network.

The Adam optimizer was used for training all the prominent existing models that were taken into consideration, in further phases of this study.

During the next phase of the study, we tried gaining more detailed insights into the performance of the proposed model, *LeatherNet*. To achieve this, the performance of various prominent architectures trained on the dataset was compared with performance of *LeatherNet* architecture. Accuracy, Precision, Recall and Loss metrics were primarily used for comparing the efficiency of various existing deep learning architectures with the proposed architecture, *LeatherNet*.

Table 6 summaries the performance of the prominent existing architectures trained in conditions like that of the proposed *LeatherNet* architecture. The results of the same are summarized as follows.

Several existing architectures like VGG16, VGG19, Res Net 50 and Efficient Net B 0 were taken into consideration during this phase of the study.

To begin with, the VGG16 architecture was taken to consideration. When this architecture was initially trained for 1000 epochs, the maximum training and validation accuracy reached by the model was found to be 78.19% and 81.41% respectively. The training precision, validation precision, training recall and validation recall on the other hand recorded 77.54%, 89.98%, 91.34% and 99.24% respectively. However, the learning graph was not found to display a promising learning curve with quite haphazard learning observed during the model training. The learning curve was also found to deteriorate after reaching its most optimal performance. With such inconsistent learning curve observed for VGG16, a relatively deeper architecture was next taken into consideration.

Hence, during the next phase, a deeper VGG architecture, namely the VGG19 model was taken into consideration. The architecture when trained for 1000 epochs recorded the maximum training accuracy, precision, and recall of 76.64%, 78.09% and 76.06% respectively. The maximum testing accuracy, precision and recall on the other hand recorded 78.50%, 90.87% and 98.14% respectively. However, there was no improvement in the results of the deep neural network observed at the end of the training process. Careful observations of the learning graphs also revealed that the learning curve during the model training process was found to deteriorate after a certain point of time. Further analysis on the learning path also revealed the presence of huge spikes indicating inconsistent learning happening by the neural network. We next considered relatively deeper neural network architecture.

We then took ResNet50 into consideration. Like the previous cases, this architecture was also trained initially for 1000 epochs. The maximum training accuracy, precision and recall recorded 50.44%, 50.45% and 100% respectively. Also, the maximum testing accuracy, precision and recall recorded 50.70%, 49.29% and 100% respectively. This time the deep neural network was found to display a poorer performance in comparison with the VGG architectures that were taken into consideration. It is pertinent to note that the recall values recorded 100% in both training and testing set of the dataset. However, the overall performance of this model was not found to be quite satisfactory for the development of a reliable AI system.

During the final phase of the study, a much deeper architecture, namely the Efficient NetB0, was taken into consideration. The model was again trained on 1000 epochs. This time the maximum training accuracy, precision and recall recorded 50.24%, 50.28% and 100% respectively. Also, the maximum testing accuracy, precision and recall recorded 50.70%, 49.29% and 100% respectively. However, this time again the performance of the model was not found to display significant improvement when compared with ResNet50.

Upon further analysis, it was found that certain deep learning architectures like VGG16, VGG19 displayed much better performance than all the conventional machine learning models tested in this study. On the other hand, deep neural network architectures like ResNet50 and Efficient NetB0 displayed very good recall values, outperforming the recall capacity of all the conventional machine learning algorithms. However, it is also pertinent to note that the overall performance of these architectures (ResNet50 and Efficient Net B 0) was not found to be quite satisfactory. The overall performance of these architectures was found to underperform when compared with the other conventional machine learning models. It is also observed that as the number of layers in the prominent existing CNN model is increased, the performance of the model in detecting and classifying the leather defects and non-defects decreases.

It is pertinent to note that though the performance of the VGG architectures was found to be better than ResNet50, EfficientNetB0 and the conventional machine learning algorithms, the performance of the model was not enough to develop a reliable automated system for industrial application.

On comparing the performance of the existing models with the proposed deep learning architecture, *LeatherNet,* it was found that the proposed *LeatherNet* performed much better than all the other existing algorithms experimented in this study. Thus, it was concluded that the proposed AI based system led to the evolution of a highly reliable and highly efficient automated system for the detection and classification of the leather surfaces into defects and non-defects.

During the final phase of the study, we compare the performance of our proposed architecture *LeatherNet*, with the existing studies done in this field. Table 7 summaries the performances of the existing state of the art that authors listed out in Table 1 to have a better insight into the performance of *LeatherNet*.

It is noted that the reference papers[2-7,9-10] corresponds to developing a leather defect detection system, and the reference papers[8,11] corresponds to developing leather segmentation algorithm. Since this work essentially deals with leather defect detection, we will take into consideration the first set of papers for detailed comparative analysis. The algorithm proposed in 2 recorded an accuracy of 94.25% on the validation set of the image data. Similarly, the combination of ANN and XBoost ANN proposed in 3 recorded an accuracy as high as 94%. However, it is pertinent to note that this study was carried out on 1600 calf leather patches. On the other hand, the work carried out in 4recorded an accuracy of 89.75%. However, it is pertinent to note that this study involved using 577 leather defect and 100 leather non-defect images. Similarly, the authors of reference

Table 7 — Comparison of existing works

Author	Performance Measurement/Remarks by the authors
Swamiraj Nithiyanantha Vasagam et al.[2]	94.25% validation accuracy was achieved on the image dataset.
Y.S.Gan et al.[3]	An accuracy of 94% accuracy on an image dataset having 1600 calf leather patches was achieved.
Malathy Jawahar et al.[4]	An accuracy of 89.75% on an image dataset having 577 leather defect and 100 leather non-defect images was recorded.
Praveen Kumar Moganam et al.[5]	An accuracy of 99.4% was recorded on an image dataset having 3600 leather images from 6 different classes.
Swamiraj Nithiyanantha Vasagam et al.[6]	Proposed LDA recorded an accuracy of 92%. While SVM recorded an accuracy of 89.65%.
Salik Ram Khanal et al.[7]	IOU score of more than 95% in 5 classes was recorded
Zhongliang Zhang et al.[8]	The proposed network displayed an improvement of 1.99% in the F1 score when compared to UNet for 1944 wet leather surface image dataset.
Malathy Jawahar et al.[9]	The overall accuracy, specificity, sensitivity, and precision of 88.61%, 91.62%, 78.63% and 93.36% respectively was recorded on leather surface defects image dataset used by the authors
Aashish Ghimire et al[10]	Optimal mean IOU value of 72.10% and mean F1 score of 82.59% was recorded.
M. Sornam et al[11]	The ideal number of clusters for segmenting the leather images was found to be 6.
HarshalPiwal et al.[12]	The VGG16 and VGG19 architectures were found to record 93.3% and 97.2% accuracy respectively.
D. Soukup et al.[13]	They could show that the efficiency of detecting the defects on rail surfaces using CNN was found to be better than the existing model-based approaches.
HonglinXionget al.[14]	The proposed MCNN model was found to perform better than the traditional CNN models.
Tao Zhanget al.[15]	FCN was found to display better recognition accuracy than CNN.
Liao D et al.[16]	Recorded an accuracy of 97.4% on the Si_3N_4 turbine blades dataset
Xinghui Dong et al.[17]	Their algorithm was found to produce relatively superior results when compared with other alternative methods.
Zhong Zhang et al.[18]	Recorded an accuracy of 97% on the test set.
Pallavi S. Chandanshive et al.[19]	The proposed network was found to achieve an accuracy of 98.9%
K. Sunitha et al[20]	The proposed model was found to record an accuracy of 97%.
Yanghuan Xu et al[21]	The recognition model was found to record an accuracy of 93.05%. This performance was found to be better than VGG16, MobileNetV2 and ResNet50.
Limei Song et al[22]	The model was found to record an accuracy of 98%.
This Work (Proposed *LeatherNet* Model)	Maximum training accuracy, precision, recall of 99.78%, 99.69% and 99.92% respectively. Maximum validation accuracy, precision, and recall of 97.42%, 97.66% and 99.40% respectively

paper[5] developed anAI based system that could record an accuracy of 99.4% on a dataset having 3600 leather images from 6 different classes. On the other hand, the authors of reference paper[6], recorded a classification accuracy of 92% using LDA. They used a leather image dataset having 428 leather images in their study. The authors of reference paper[7], on the other hand used MVTECT dataset for leather defect detection using semantic segmentation. They could achieve an IOU score of more than 95% in 5 classes. Similarly, the authors of reference paper[9] also proposed a system that could detect the leather defects with an accuracy of 88.61%. The authors of reference paper[10], developed a leather defect detection system based on Semantic Segmentation and recorded the optimal mean IOU value of 72.10% and mean F1 score of 82.59%. However, the proposed *LeatherNet* model presented in this work is capable to achieving the maximum training accuracy, precision, recall of 99.78%, 99.69% and 99.92% respectively. The validation accuracy, precision, and recall for the proposed *LeatherNet* model on the other hand recorded 97.42%, 97.66% and 99.40% respectively. This indicates its superior capability in detecting and classifying the leather images into defects and non-defects. Thus, we could conclude the development of a highly efficient and highly reliable AI based system for automating the detection and classification of the leather surfaces into defects and non-defects.

4 Conclusion

The work essentially deals with developing an automated system for the detection and classification of leather surfaces into defects and non-defects. To facilitate the process of development of an automated system, Authors have used a leather surface image dataset having images from both defect and non-defect classes. However, since the number of images present in the image dataset was found to be limited, hence data augmentation was performed. This enabled the generation of images under different conditions, which would thereby increase the number of data instances and thus also facilitate the development of a better AI based model. The features were then extracted from these images using various image processing techniques. The extracted features were then fed into various conventional machine learning algorithms. However, the performance of the conventional machine learning algorithms was not found to be satisfactory. Upon further analysis, it was concluded that the inefficient and unreliable artificial intelligence-based systems developed using machine learning algorithms were mainly due to the huge variability in the types of leather surface and kinds of leather defects. Hence, in the next phase of the manuscript, the performance of the deep learning models in detection and classification of leather surfaces into defects and non-defects were explored.

A novel lightweight CNN architecture, *LeatherNet* for the detection and classification of leather surfaces into defects and non-defects was thus proposed. The *LeatherNet* is designed in such a way that the model could generate extremely reliable results despite being lightweight. The proposed model was specially designed to be shallower and have lesser number of parameters to enable faster training of the model. To see that the performance of the automation system is not compromised due to the application of such constraints while designing the model, a comparative work with the performance of the prominent existing CNN architectures (like VGG16, VGG19, ResNet50 and EfficientNetB0) using transfer learning concepts is also done. It is found that the performance of the proposed model is in no way compromised in view of the above designing constraints, indicating its superior efficiency and reliability in detecting and classifying the leather surfaces into defects and non-defects. The performance of *LeatherNet* model is then finally compared with the existing state of the art in this domain to better analyze the performance efficiency of the proposed model. It is found that the performance of the proposed lightweight architecture is found to generate outstanding results when compared with the existing state of the arts works.

It is finally concluded that the proposed lightweight *LeatherNet* neural network architecture evolved as a highly reliable and highly efficient AI based system for detection and classification of leather surfaces into defects and non-defects.

With highly reliable results obtained by the proposed model for detection and classification of leather surfaces into defects and non-defects, the work could be extended to identifying the specific type of leather surface defect that the leather has incurred. The work could also be deployed into an easy-to-use application to facilitate easy deployment in the leather industry.

Acknowledgement

Authors are thankful to CSIR-Central Leather Research Institute (CSIR-CLRI), Chennai, University of Madras, Chennai and Vellore Institute of Technology Chennai Campus for their extensive support during this research work. The authors also sincerely acknowledging the support extended from Tanners Association. The authors also acknowledge the CSIR-CLRI communication No. 1825. This work was supported by the CSIR-CLRI (internal funding) under MLP14. The data used to support the findings of this work may only be obtained from the CSIR-CLRI after a reasonable request.

References

1 Indian Leather Industry. https://leatherindia.org/indian-leather-industry (accessed 25 November 2022)

2 Vasagam, S.N., Sornam, M. *ICT Analysis and Applications. Lecture Notes in Networks and Systems*, Springer Singapore, 314, (2022). pp. 35-45.

3 Gan, Y. S., Wei-ChuenYau, Sze-Teng Liong, and Chih-Cheng Chen. *Mathematical Problems in Engineering* 2022, 1 (2022): 5549879.

4 Jawahar, Malathy, L. Jani Anbarasi, S. Graceline Jasmine, Modigari Narendra, R. Venba, and V. Karthik. In *Inventive Computation and Information Technologies: Proceedings of ICICIT 2020*, pp. 189-202. Springer Singapore, 2021.

5 Moganam, Praveen Kumar, and Denis Ashok SathiaSeelan. *Journal of Leather Science and Engineering* 4, 1 (2022).

6 S. Nithiyanantha Vasagam and M. Sornam. *Int. J. Innov. Technol. Exploring Eng* 8, 12 (2019), pp. 2253-2259.

7 Khanal, Salik Ram, Jorge Silva, Luís Magalhães, João Soares, Dibet Garcia Gonzalez, Yusbel Chavez Castilla, and Manuel J. Ferreira. *Procedia Computer Science* 204 (2022). pp. 573-580.

8 Zhang, Zhongliang, Yao Fu, Huiling Huang, Feng Rao, and Jun Han. *Mathematical Biosciences and Engineering* 19, 12 (2022). pp. 13782-13798.

9 Jawahar, Malathy, and K. Vani. *Journal of the American Leather Chemists Association* 114, 1 (2019).

10 Ghimire, Aashish, Aman Mahaseth, Ramesh Thapa, and Suraj Ale. *Journal of Artificial Intelligence* 4, 2 (2022), 131-138.

11 Sornam, M., and S. Nithiyanantha Vasagam. *International Journal of Engineering Research and Application* 8(9)(2018), pp.12-16.

12 Piwal, H., Dhokale, M., Biswas, R., Raut, S., & Malge, A. *International Journal of Mechanical Engineering*. 7(3), 2022.

13 Soukup, Daniel, and Reinhold Huber-Mörk. In *International symposium on visual computing*. Cham: Springer International Publishing, 2014. pp. 668-677.

14 Xiong, Honglin, Chongjun Fan, Shan Zhao, and Ying Yu. *JisuanjiJichengZhizaoXitong/Computer Integrated Manufacturing Systems* (2020).

15 Zhang, Tao, Yuting Liu, Yaning Yang, and YingguJin. In *2020 39th Chinese Control Conference (CCC)*, IEEE, 2020. pp. 7310-7315.

16 Liao, Dahai, Zhihui Cui, Xin Zhang, Jun Li, Wenjie Li, Zuoxiang Zhu, and Nanxing Wu. *Advances in Mechanical Engineering* 14, 2 (2022): 16878132221081580.

17 Dong, Xinghui, Chris J. Taylor, and Tim F. Cootes." In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.

18 Zhang, Zhong, Borui Zhang, and Takuma Akiduki. In *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*, 2019. pp. 6-10.

19 Chandanshive, P. S., Chavan, A. H., Kale, P. D., Sutar P. D., Siddha, U. Y. *International Research Journal of Modernization in Engineering Technology and Science*. 3(3). 2021.

20 Sunitha, K., & Reddy, S. M. *International Journal of Electrical Engineering and Technology (IJEET)*, 12(5), May 2021, pp. 28-37.

21 Xu, Y., Wang, D., Duan, B., Yu, H., & Liu, H. (2021). *Applied Sciences* 11(19). 2021. p. 8945

22 Song, L., Li, X., Yang, Y., Zhu, X., Guo, Q., & Yang, H. *Sensors*, *18*(11), 2018. 3709.

23 Singh, Y. K., & Singh, N. D. *International Conference on Energy, Communication, Data Analytics and Soft ComputingIEEE*2017. 2017. pp. 3883-3888

24 Kumar, P. R., & Vijaya, A. (2022). *Materials Today: Proceedings*, 68, 2265-2274.

25 Patidar, D., Shah, B. C., & Mishra, M. R. 2014 *International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE). IEEE.*

26 Alassar, Z. I. Decision tree as an image classification.

27 Tammina, S. *International Journal of Scientific and Research Publications*, 9(10), pp. 143-150.

28 Bansal, M., Kumar, M., Sachdeva, M., & Mittal, A. (2021). *Journal of Ambient Intelligence and Humanized Computing. Springer Science and Business Media LLC*.

29 Reddy, A. S. B., & Juliet, D. S. *International Conference on Communication and Signal Processing.*IEEE2019. pp. 0945-0949.

30 VijayaKumari, G., Vutkur, P., & Vishwanath, P. *Global Transitions Proceedings*, 3(1), 2022. pp. 225-229.