Moganam and Sathia Seelan
Journal of Leather Science and Engineering (2022) 4:7
https://doi.org/10.1186/s42825-022-00080-9

Journal of Leather Science and Engineering

RESEARCH Open Access

Deep learning and machine learning neural network approaches for multi class leather texture defect classification and segmentation

Praveen Kumar Moganam and Denis Ashok Sathia Seelan*

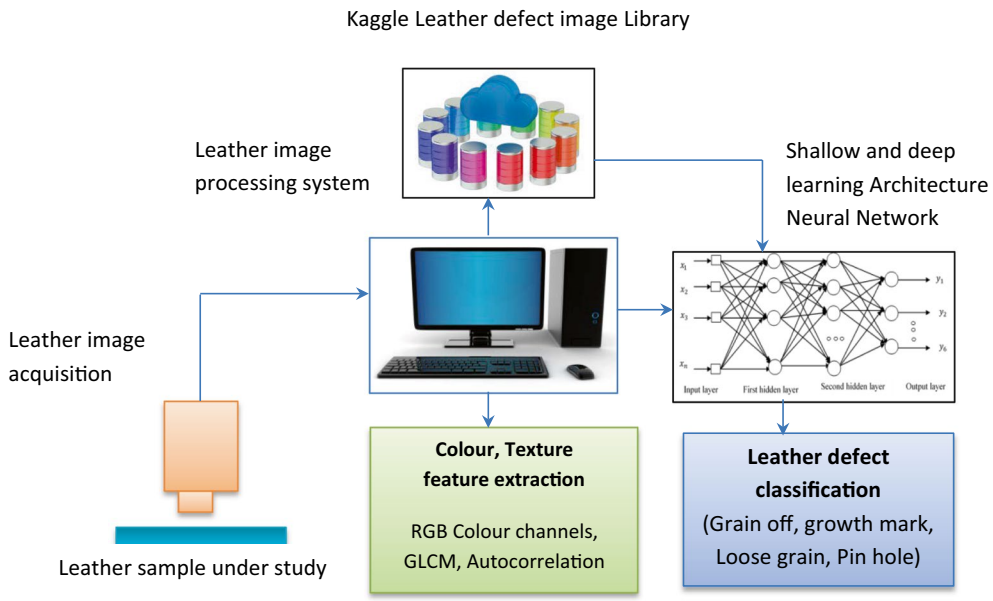
Abstract

Modern leather industries are focused on producing high quality leather products for sustaining the market competitiveness. However, various leather defects are introduced during various stages of manufacturing process such as material handling, tanning and dyeing. Manual inspection of leather surfaces is subjective and inconsistent in nature; hence machine vision systems have been widely adopted for the automated inspection of leather defects. It is necessary develop suitable image processing algorithms for localize leather defects such as folding marks, growth marks, grain off, loose grain, and pinhole due to the ambiguous texture pattern and tiny nature in the localized regions of the leather. This paper presents deep learning neural network-based approach for automatic localization and classification of leather defects using a machine vision system. In this work, popular convolutional neural networks are trained using leather images of different leather defects and a class activation mapping technique is followed to locate the region of interest for the class of leather defect. Convolution neural networks such as Google net, Squeeze-net, RestNet are found to provide better accuracy of classification as compared with the state-of-the-art neural network architectures and the results are presented.

Keywords: Convolution neural networks, Machine learning classifier, Leather defects, Multi class classification, Class activation map, Segmentation

*Correspondence: denisashok@vit.ac.in
Department of Design Automation, Cyber Physical Systems Laboratory, School of Mechanical Engineering, Vellore Institute of Technology, Vellore 632014, India

Graphical Abstract



1 Introduction

Modern leather manufacturers and designers have given major focus on aesthetic perception, visual appearance and hand feel of leather garments as it affects the purchasing decision. However, various defects are introduced in the leather surface during the pre-tanning and post tanning processes in the leather industries. Hence, the identification and classification of leather defects is an essential process for maintaining the quality of finished products. As the manual inspection methods are slow, error-prone, and labor-intensive, machine vision-based automated inspection techniques are widely adopted for improving the productivity of the leather inspection process [1]. Due to the ambiguous texture pattern and tiny size of the defect, it is difficult to distinguish the localized defect and the background in the leather images. Hence, there is a need for developing a suitable image processing approach for the improved classification and perception of leather defects.

Various image processing techniques were proposed by many researchers for leather grading, defect identification, and classification. Quality inspection for grading is an important step in assessing the usable area of leathers. Each piece of leather is graded based on its effective cutting value, which is decided to take into consideration the number, size, and location of surface defects [2]. Grayscale image processing techniques using thresholding and morphological operations are applied for defect detection applications [3]. A histogram-based identification method is proposed for detecting defective leather images [4]. Edge detection along with morphological operations is applied to the leather images for segmenting the defect locations in the leather images [5]. Texture analysis technique using wavelet transform provides a collective spatial analysis of local pixel regions for leather defect detection [6]. A multi-level thresholding algorithm with the texture feature extraction method is proposed to segment defective and non-defective regions of leather for objectively quantification of the leather surface defects [7]. Sobral et al. introduced a new wavelet-based method using optimized filter banks for leather defect detection [8]. An optimization approach with a filtering process is applied for isolating the defective regions from the complex and not homogeneous background by analyzing their strongly oriented structure [9]. For defect detection and classification process, several image processing algorithms are employed to provide the quantitative descriptions of defective and non-defective leather images, various descriptors like first-order statistics, Contrast characteristics, Haralick descriptors, Fourier and Cosine transform, Hu moments with information about the intensity, local binary patterns, Gabor features are extracted to locate defect's positions on the leather surface [10]. Haralick features are derived from gray-level co-occurrence matrix (GLCM) which extract the local patterns in the image and count their distribution

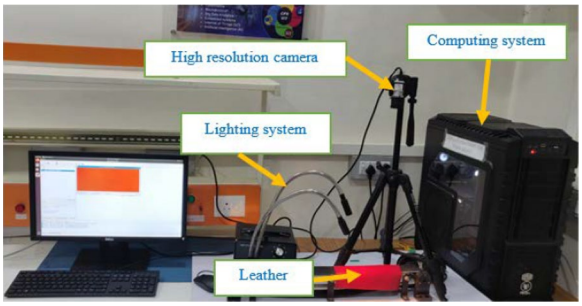


Fig. 1 Machine vision inspection system for defect identification in leather samples

Table 1 Major specifications of camera used for leather image acquisition

Specifications	Description
Make	Basler acA 4600
Sensor type	CMOS
Sensor size	6.5 mm × 4.6 mm
Resolution (H × V)	4608 px × 3288 px
Resolution	14 MP
Pixel size (H × V)	1.4 µm × 1.4 µm
Frame rate	10 fps
Mono/color	color
Interface	USB 3.0

across the entire image. They provide a good discriminative encoding of the textural and gradient-based information, in the form of feature values [11]. A texture analysis method using wavelet statistical features and wavelet co-occurrences matrix features such as entropy, energy, contrast, correlation, cluster prominence Standard Deviation, Mean, and local homogeneity is proposed for leather defect classification [12]. Color-based models and Co-occurrence matrix-based texture analysis is reported for defect detection in raw leather [13]. Though the digital image processing approaches are applied for leather defect inspection applications, the accuracy classification is limited due to the presence of noise.

Recently, machine learning and deep learning methods have gained attention for image classification, detection, and segmentation applications. Kwak et al. proposed a three-stage sequential decision tree for the classification of defects such as lines, holes, stains, wears, and knots [14]. Viana et al. presented an empirical evaluation of support vector machine against AdaBoost and MLP, for solving the leather defect classification problem [15]. Supervised classification using the multi-layer perceptron (MLP), Decision trees (DT), SVM, Naïve Bayes, KNN, and Random forest (RF) classifiers were used to classify the defective and non-defective leather regions [16]. The neural network classifier is proposed by multilayer perceptron neural networks for recognizing leather defects like open cut, closed cut, and Fly Bite [17]. Amorim et al. presented linear discriminate analysis techniques for attribute reduction to four different classifiers such as C4.5, KNN, Naïve Bayes, and SVM classifiers for leather defect classification [18]. With recent advancements in computing and graphical processing units, deep learning neural networks are developed for automated inspection applications [19]. ResNet and VGG architectures based on convolution neural networks are capable of automated surface inspection and image classification applications using transfer learning [20]. Liong et al. proposed an integrated machine vision system using an artificial neural network and deep learning neural network for leather defect classification [21]. Region convolution neural network-based deep learning approach is used for defect detection and segmentation of defective regions in the leather image [22]. Based on the visual-tactile sense perception of the consumers, the back propagation neural network is developed for selecting the suitable leather materials to manufacture the user specified leather products [23].

It is found that there are many research works have been contributed for the leather defect detection and classification. The existing classification approaches has limited human perception using leather images and the accuracy of the classification is also found to be limited due to the vagueness, randomness, and size of the leather defects in the background texture pattern of the leather surface. In order to provide the improved accuracy of leather defect classification and human perception; this paper presents deep learning convolution neural network and machine learning classifier approaches for multi class classification and segmentation of leather defects. Classification performance of state-of-the-art deep learning and machine learning classifiers are compared for the leather data sets with texture defects and the results are presented in this paper.

2 Machine vision-based leather inspection system

A typical leather surface consists of different types of defects such as scars, growth marks, grain off, loose grain, pinholes, and folding marks. A machine vision system consisting of a high-resolution camera (BASLER acA4600), lighting system, computing system with an image processing software (MATLAB Version. 2020a) is established in the present work for identifying and classifying the leather defects and it is shown in Fig. 1.

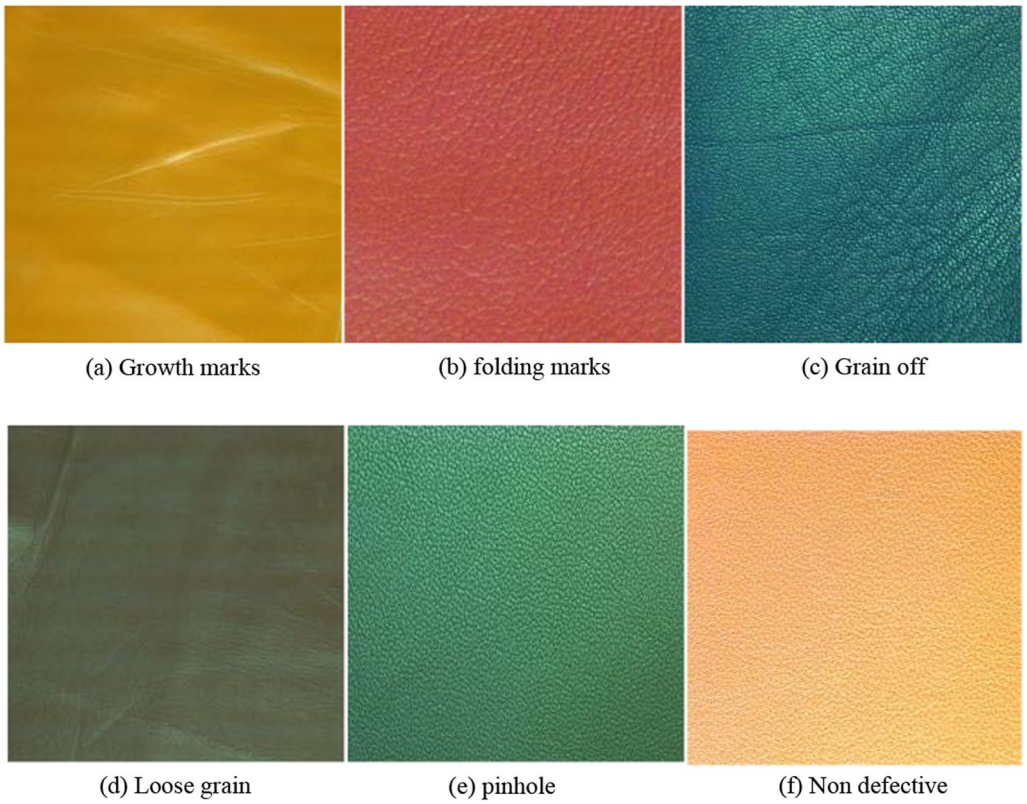


Fig. 2 Samples of leather surfaces with different colours and defects

A high-resolution camera with a resolution of 14 MP is used for acquiring the images of the leather surface with a resolution of 4608 × 3288 pixels. Table 1 shows the specifications of the camera used in the machine vision system.

Lighting plays an important role in vision inspection applications for illuminating the object of interest. In this work, a fiber optic illumination system is used for providing uniform illumination on the leather surface. The magnitude of luminance is measured using a Lux meter and it is controlled using a light controller knob.

2.1 Leather Image acquisition

A comprehensive data set of 3600 leather images is developed with different defects such as folding marks, grain off, pinhole, growth marks, loose grain, and non-defective leather surfaces. It's placed in open data science environment Kaggle for exploring suitable machine learning and deep learning-based image processing technique to classify the leather defects. Figure 2 shows the sample leather images of leather images with different leather defects.

It can be noted from Fig. 2a, b that folding marks and growth marks have a better visual perception of change in color and texture as compared to other defects. Also, the grain off, pinhole, and loose grain are found to have a finer texture pattern as identified in Fig. 2c–e respectively.

2.2 Leather texture defects

The leather images of different colours and defects as shown Fig. 2 are processed to obtain the grey scale intensity maps for analysing the texture variations due to different leather defects and the results are shown in Fig. 3. Folding marks, grain off, pinhole and loose grain has a coarse texture and they disturb the regular texture pattern which leads to many abrupt variations and peaks in the intensity of the pixels as shown in Fig. 3a–d respectively.

Growth mark and non-defective leather showed finer texture and uniform intensity variation is seen in Fig. 3e, f respectively. The visual perception of leather defects is limited by the ambiguous texture pattern and tiny nature of different leather defects. In order to the distinguish the type of leather defects, there is a need for developing suitable image processing algorithms for classification of different leather defects.

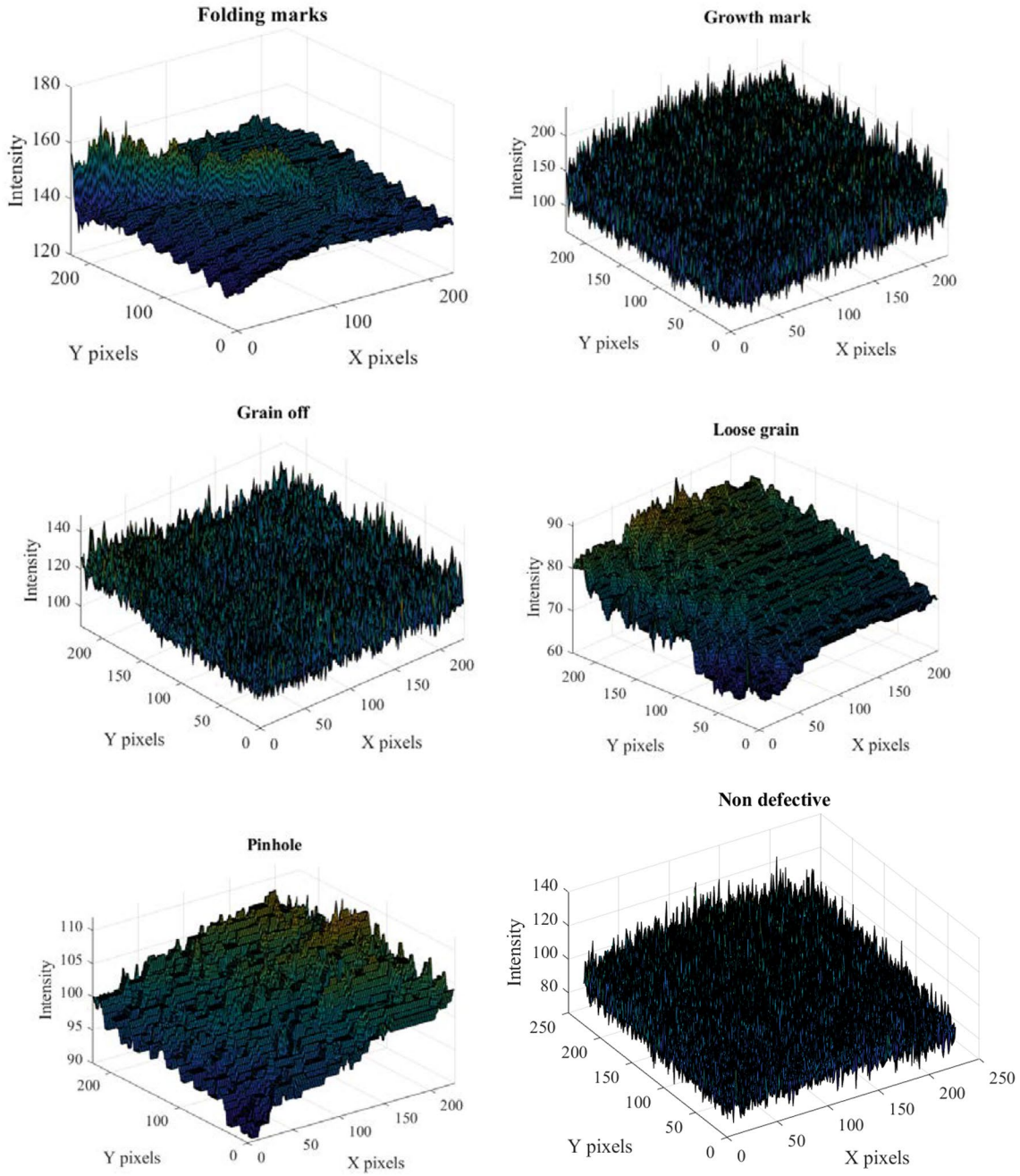


Fig. 3 Gray intensity map of different leather images with different leather defects

3 Deep learning neural network approach for classification and localization of leather defects

In order to reduce the error in detection and multi class classification of leather texture defects, this work presents deep learning convolutional neural network approach using state of the art convolutional neural network architectures like Alexnet, VGG-16, Google net, Squeeze Net, ResNet-50. Figure 4 shows the framework of the proposed approach for classifying and labeling leather images as non-defective leather, Loose grain, Grain off, Growth marks, Pinhole and Folding marks.

Using the developed deep learning neural network models, class activation map is generated for identifying the region of interest of the class of leather image.

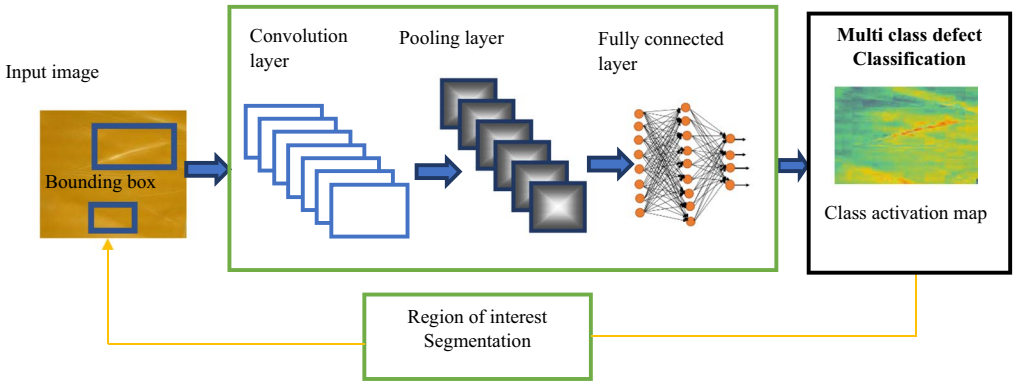


Fig. 4 Block diagram of deep learning methodology for multi defect classification and localization

The details of network architecture, layers are explained in the following subsections.

3.1 Leather image Data Set preparation and preprocessing

A data set of 3600 leather images consisting of non-defective leather, Loose grain, Grain off, Growth marks, Pinhole and Folding marks is used for training neural networks. All the leather images are preprocessed using histogram equalization for addressing the illumination variations during image acquisition and resized to 227×227. For evaluating performance of segmentation of defective regions in the leather image, ground truth leather images with hand labeled defective regions are kept in the data set. A fivefold cross validation approach is followed in the present work in which the data set of 3600 image samples are split into 5 mutually exclusive and exhaustive folds of 720 leather images. Repeatedly, a fold is selected and designated as testing set, all the other remaining leather images (80% of the data) are considered as the training set.

3.2 Deep learning convolutional neural network architectures

In the present Standard, pre trained convolutional neural network architectures such as Alexnet, VGG-16, ResNet, Google net, Inception Net and Squeeze net are considered for the multi class leather defect classification application as they are relatively established and proved their ability for object detection and multi class classification applications. Figure 5 shows the architectures of deep learning convolutional neural networks which are considered in the present work. A typical convolutional neural network models contains convolution layer, pooling layer and fully connected layer. It can be seen that the complexity network architecture increases with concatenation, parallel channels and feedback as it shown in Fig. 5c, d for inception net and Resnet respectively. The convolution layers are associated with different parameters such as weights, kernel size, stride, padding etc. More details of deep learning architectures can be studied in the literature [24].

3.2.1 Convolution layer

Convolution layer provides automated feature extraction from given images with the specific spatial locations using number filters of different sizes. A non-overlapping feature map is obtained as an output using convolution operation between weights of the filter and the output of the previous convolutional layer as given by Eq. (1).

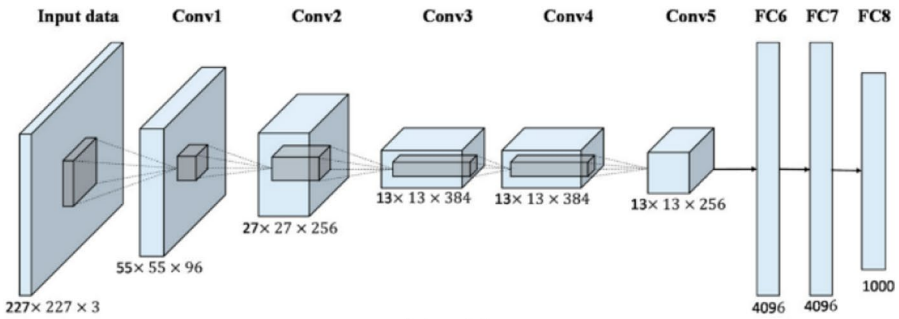
$$x^l = f\left(b^l + \sum_j^J \sum_i^I w_{i,j}^l x_{i,j}^{l-1}\right) \tag{1}$$

where (J, I) denotes the size of the filters, J is the height of the filters, and I is the width of the filters. b^l denotes the bias of the convolutional layer. x^{l-1} denotes the output of the previous convolutional layer. wl denotes the weight of convolutional layer. $f(\cdot)$ is the nonlinear activation function and ReLU activation function is selected and is shown as Eq. (2).

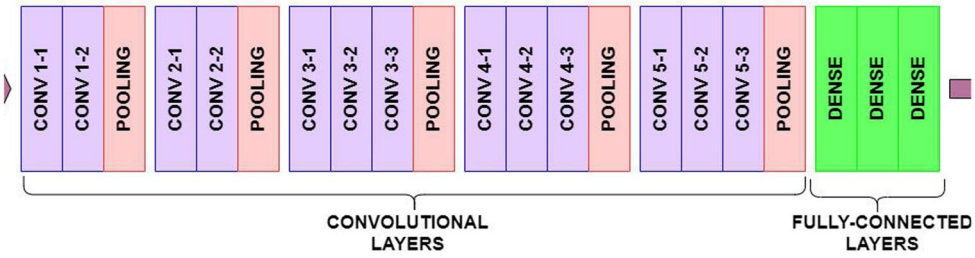
Size of the feature map depends on with several parameters including the input size, filter size, depth of the map stack, zero-padding and stride.

$$M_x = \frac{l_x - K_x}{S_x}, M_y = \frac{l_y - K_y}{S_y}, \tag{2}$$

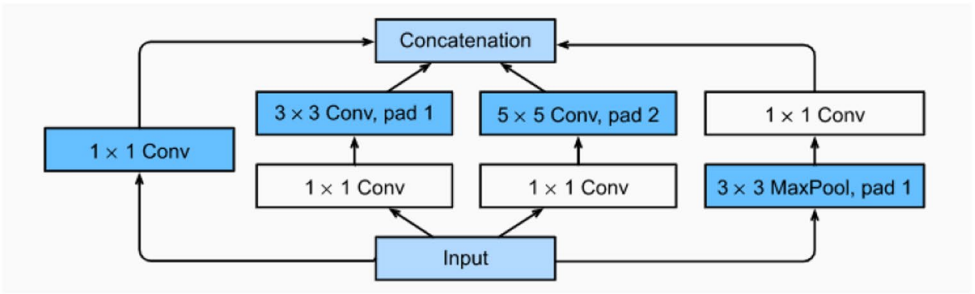
where (Mx, My), (Ix, Iy), (Kx, Ky) indicate the map size, input size, kernel size respectively and (Sx, Sy) indicate the stride in row & column. The number feature maps depend on with the number of filters and its size. In a typical deep learning convolution neural network, number of features with the increase in number of convolution layers and the associated filters.



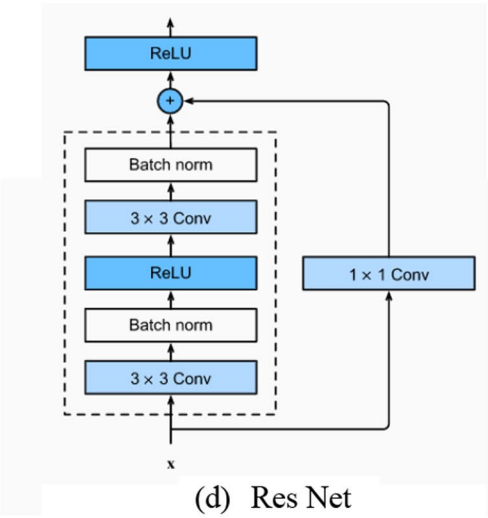
(a) Alexnet



(b) VGG 16



(c) Inception



(d) Res Net

Fig. 5 Block diagram of standard Deep Learning convolutional Neural Networks

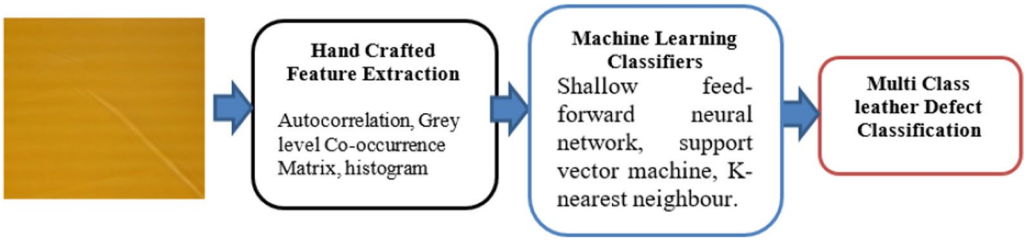


Fig. 6 Block diagram of Machine learning approach for multi class leather defect classification

3.2.2 Activation function

In order to learn the universal approximation of input values and output classes as continuous function in a Euclidean space, a suitable activation function is essential. As the sigmoid and Hyperbolic tangent functions suffer vanishing gradient problem, a rectilinear unit is used as activation function and it returns the same value (x) of the feature maps provided as output (x) if its magnitude is greater than zero.

3.2.3 Pooling layer

In order to prevent the overfitting and reduce the dimensionality of feature maps, pooling layer perform down-sampling of the input feature map using a window function u (n,n). Max pooling and average pooling strategies are often followed in the pooling layer.

$$xm = \underset{N\times N}{max}\,(xlu(n,n)) \tag{3}$$

where $u(n,n)$ is the window function, which is applied to calculate the maximum value of xl_j in the neighborhood.

3.2.4 Fully connected layer

Fully connected layer in the deep learning neural network receives a feature vector from the previous max pooling layer and it is trained for the multi class classification of given leather image using the associated weights and an activation function by reducing a loss function. In this work, a fully connected layer with six number of output neurons is configured for providing the categorical output such as non-defective leather, Loose grain, Grain off, Growth marks, Pinhole and Folding marks using an encoding technique.

More details of the training of the neural network for the multi class classification is described in Sect. 4.2. Typical output of a neuron in a fully connected layer for the feature vector of the max pooling layer x_{m-1} is given by Eq. (4).

$$v^m = f(w^m x^{m-1} + b^m) \tag{4}$$

where bm denotes the bias of the fully connected layer. w_m denotes the weights of the fully connected layer. x_{m-1} denotes the output of the previous max-pooling layer. $f(.)$ is the activation function.

3.3 Visualization of region of interest for defect localization

In this work, Gradient-weighted class activation mapping is followed which is the weighted sum of each channel of the feature map to identify the specific discriminative regions of the given leather image non-defective leather, Loose grain, Grain off, Growth marks, Pinhole and Folding marks. Here, class activation map with values of scores for the class of the leather 'c' at the spatial location (x,y) of the image is generated using the kth channel of the feature map and corresponding weight w_{ck} as given by

$$Mc(i,j) = \sum w_k^c, f_k(x,y) \tag{5}$$

where M_c is the class activation map of class c and w_{ck} represents the k_{th} weight of the fully connected layer of class c. As the part of the image with larger score influences the corresponding class, a thresholding approach is followed for the selecting the region of interests in the original image.

$$T(x,y) = \begin{cases} 0 & M(x,y) > threshold \\ 1 & M(x,y) < threshold \end{cases} \tag{6}$$

Subsequently, the region of interest ROI(x,y) is obtained with threshold T(x,y) for indicating the discriminative region in the image $I(x,y)$:

$$ROI(x,y) = I(x,y) \cdot T(x,y) \tag{7}$$

A bounding box is generated from the region of interest of the given image and it is compared with t ground truth bounding boxes in the ground truth leather images.

Table 2 Description of GLCM texture features used for leather defect identification [10]

GLCM texture parameters	Description
$Entropy = \sum_{i=0}^{ng-1}\sum_{j=0}^{ng-1} -p_{ij} * \log p_{ij}$	Measure of statistical randomness of the leather surface
$Correlation = \frac{\sum_{i=0}^{ng-1}\sum_{j=0}^{ng-1} -(i,j)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$	Measure of linear dependency of gray levels of neighbouring pixels
$Contrast = \sum_{i,j=0}^{ng-1} p_{ij}(i-j)^2$	Measure of the intensity between a pixel and its neighbour
$Energy = \sum_{i,j=0}^{ng-1} p_{ij}^2$	Measure of orderliness of pixels
$Homogeneity = \sum_{i,j=0}^{ng-1} \frac{p_{ij}}{1+(i-j)^2}$	Measure of smoothness of the gray level distribution
$Dissimilarity = \sum_{i,j=0}^{ng-1} p_{ij}\lvert i-j\rvert\rvert$	Measure of distance between pairs of pixels
$Mean\ \mu_i = \sum_{i,j=0}^{ng-1} i(P_{ij})$	Measure of the average intensity of all pixels
$Variance\ \sigma_i^2 = \sum_{i,j=0}^{ng-1} p_{ij}(i-\mu_i)^2$	Measure of dispersion of gray-level distribution of pixels

4 Machine learning based approaches for multi class defect classification of leather defects

Proposed deep learning neural network based multi class leather defect classification and localization are compared with the machine learning approaches like shallow feed-forward neural network (SFFNN), support vector machine (SVM), K-nearest neighbour (KNN). These machine learning approaches require manual feature extraction techniques from the leather images for the classification leather defects. Typical steps followed in machine learning based multi class defect classification of leather defects is shown in Fig. 6.

The details of feature extraction and machine learning classifiers are given in the subsequent sections.

4.1 Hand crafted Feature extraction from leather images

In this work, color and texture features are extracted from 3600 leather images for each class such as non-defective leather, Loose grain, Grain off, Growth marks, Pinhole, Folding marks using color histogram, Autocorrelation and GLCM.

4.1.1 Color histogram

As the leather has different colours and it has colour variations in the defective locations of the leather image, it is analysed using histogram in RGB colour space for understanding the intensity variations in red, green and blue channels as given by Eq. (8)

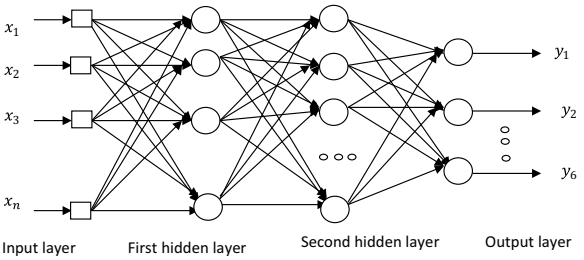


Fig. 7 Typical architecture of shallow feed-forward neural network

$$f(i,j) = \{f_R(x,y), f_G(x,y), f_B(x,y)\} \tag{8}$$

where i = 1,2, …, M/m and j = 1,2, …, N/n Further, a colour histogram is applied for the given image block and dominant intensity value of the red, green and blue channel $(R_{\max}, G_{\max}, B_{\max})$ is extracted as the colour feature of the leather image as given below:

$$C\max(i,j) = \{\max_{\forall x,y} f_R(x,y), \max_{\forall x,y} f_G(x,y), \max_{\forall x,y} f_B(x,y)\} \tag{9}$$

Here C_{max} refers to the magnitude of counts of red, blue, and blue channels of the leather image. These color features will help in quantifying the color changes of the leather image with the leather defects.

4.1.2 Autocorrelation

As the magnitude of the autocorrelation function is useful in describing the disturbance in the regular texture

pattern due to the presence of leather defects in the leather surface, the autocorrelation function is calculated for the leather image to measure its coarseness due to leather defects. Here, the color leather image $f(x,y)$ is converted into grey scale image and a two-dimensional autocorrelation function of the given leather image $f_g(x,y)$ is calculated using the following Eq. (3).

$$G(a,b) = \sum_{x}^{m}\sum_{y}^{n} f_g(x,y) * f_g(x-a, y-b) \tag{10}$$

where G (a,b) is the autocorrelation function for the grey scale and a and b represent the typical lag from the corresponding x and y position.

4.1.3 Grey level co-occurrence matrix

Grey level co-occurrence matrix provides important information for understanding the variation in texture pattern due to the type of leather defects on folding marks, grain off, pinhole, growth marks, and loose grain. Grey Level Co-occurrence Matrix (GLCM) for the given leather image is constructed by counting all pairs of a reference and neighbouring pixel separated by an offset (d) having the gray levels i and j at the specified relative orientation (θ) as given below:

$$p[i,j|d,\theta] = n_{ij} \tag{11}$$

where n_{ij} is the number of occurrences of reference and neighbouring pixels (i,j) lying at offset (d) in the image. $p[i,j]$is gray level co-occurrence matrix and it is calculated for the given grayscale image at four different orientations ($\theta = 0°$, 45°, 90° and 135°) and offsets ($d = -3, -2, -1$, 0, 1, 2, 3). The number of rows and columns of co-occurrence matrix p $[i,j]$ is equal to the number of distinct gray levels (n).To reduce the computational burden of calculating GLCM for the given image, the gray level was set to 64.Further, the elements of $P[i,j]$ are normalized by dividing each entry by the total number of pixel pairs.

Table 2 lists the formulae for calculating the different texture features from the GLCM and the corresponding descriptions. In this work, statistical texture features such as contrast, correlation, dissimilarity, energy, entropy, homogeneity, mean, and variance are calculated as the texture features of the given leather image.

Using the extracted features using colour histogram, autocorrelation and GLCM, a labelled data set is developed for training the state of the machine learning classifiers such as shallow feed-forward neural network (SFFNN), support vector machine (SVM), K-nearest neighbour (KNN) for multi class classification of leather defects.

4.2 Shallow feed-forward neural network-based machine learning classifier

In this work, a shallow feed-forward neural network (SFNN) is trained to classify leather defects such as folding marks, grain off, pinhole, growth marks, and loose grain. Figure 7 shows the typical architecture of the proposed SFFNN with two hidden layers. Here, the color features and texture feature is used as the input vector (x_i). As the magnitude of the extracted color and texture features are different, unity-based normalization is followed to ensure the proper fusion of extracted features for reducing the bias and gross influences. It also ensures the values of the input vector into the range [0, 1].

A shallow feed-forward neural network can be considered as a nonlinear model with nonlinear basis functions ϕ_j (x) as given by the Eq. (12).

$$y(x,w) = f\left(\sum_{j=1}^{H} w_j \phi_j(x)\right) \tag{12}$$

Here the weights W_j can be adjusted during training and ϕ_j (x) is a nonlinear function of a linear combination of inputs. x refers to the extracted color and texture feature vector of the given leather image. The output of the feed-forward neural network (y) can be expressed as series of functional transformations as given by Eq. (13).

$$zk = g\left(\sum_{j=0}^{H} w_{kj}^{(2)} h\left(\sum_{i=0}^{D} w_{ji}^{(1)} x_i\right)\right) \tag{13}$$

Here the superscripts (1) and (2) indicates the parameters of the respective hidden layers, x_i indicates the input feature vector. j refers to the 'H' number of hidden nodes. K refers to the number of output neurons. g, h is the nonlinear activation function of the hidden layer and it used the sigmoid activation function.

In order to achieve the multi class classification of leather type such as folding marks, growth marks, grain off, loose grain, pin hole and non-defective leather, softmax function is applied as given below:

$$P(y = j|z^{(i)}) = \varphi_{soft\,\max}^{(z(i))} = \frac{e^{z(i)}}{\sum_{j=0}^{k} e^{z_k^{(i)}}} \tag{14}$$

This softmax function computes the probability of the given training sample $x^{(i)}$ belongs to class j given the weight and net input $z^{(i)}$. Hence, we compute the probability $p(y=j|x^{(i)}; w_j)$ for each class label in $j=1$, …, k. Here, the normalization term in the denominator causes these class probabilities to sum up to one.

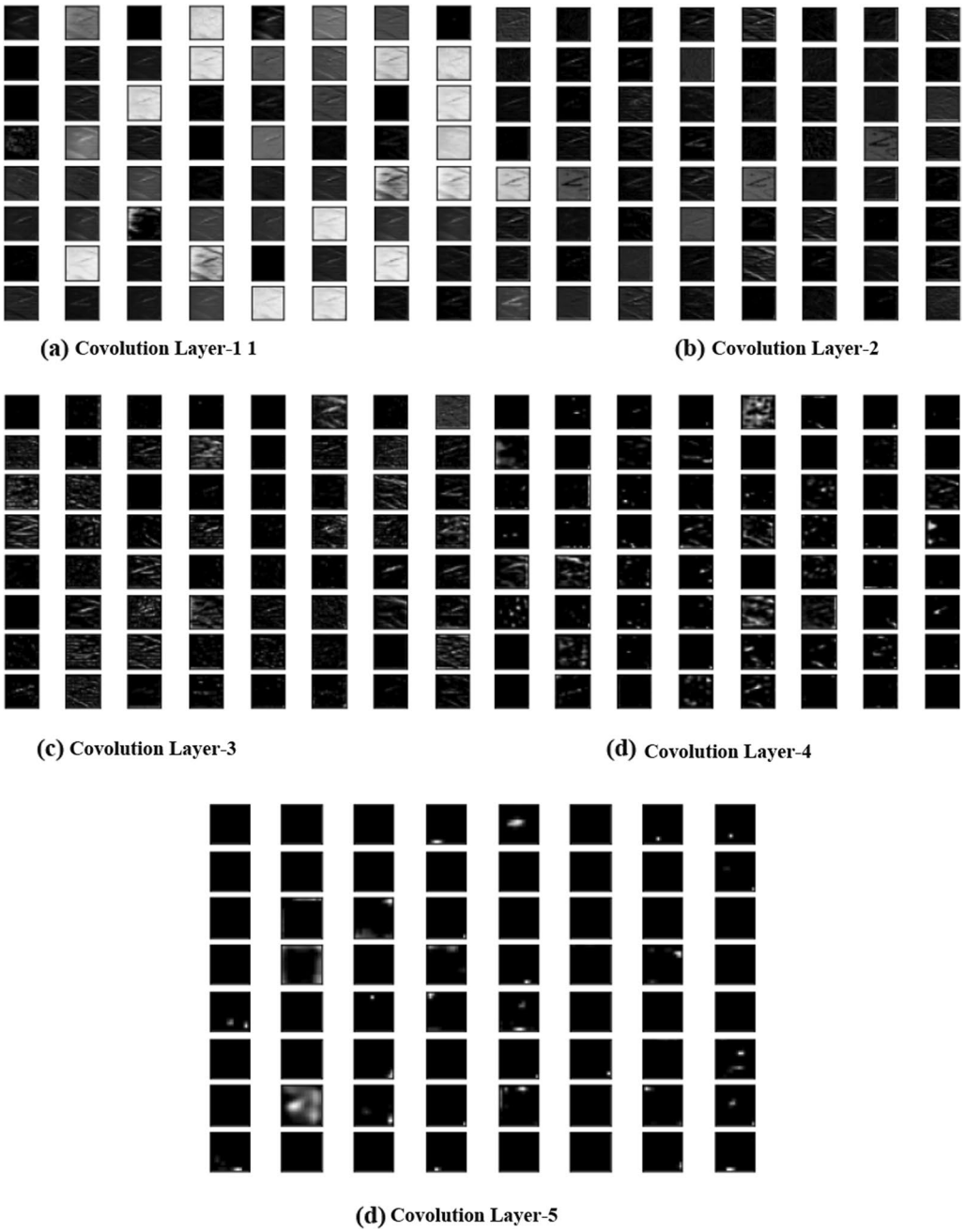


Fig. 8 Typical feature maps of Alexnet Deep Learning Convolutional Neural Network

Further, the shallow neural network is trained by adjusting the weights by defining and minimizing a cost function J. which is the average of all cross-entropies over the training data set as given below:

$$J(W) = \frac{1}{n}\sum_{i=0}^{n} H(T_i, O_i) \tag{15}$$

Here, the function (H) refers the cross-entropy function as defined below

$$H(T_i, O_i) = -\sum_{m} T_i.\log(O_i) \tag{16}$$

Here the T corresponds to the "target" labels and the O stands for computed probability from the SoftMax function. The cross-entropy based cost function is

minimized for the given training data set using the stochastic gradient descent method by iteratively updating the weight matrix until the specified number of epochs or desired cost threshold is reached.

5 Performance metrics of deep learning and machine learning classifiers

In this work, confusion matrix is constructed for evaluating the classification performance of the deep learning neural networks and machine learning classifiers. A Confusion matrix is an N × N matrix where N is the number of target classes and it compares the number of predicted classes by the classifier. Here the diagonal elements represent the correct classifications whereas all the other entries show misclassifications. Based on the confusion matrix, performance metrics like Precision, Sensitivity, F1-score and accuracy are calculated using the formula as given below.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{F1 Score} = \frac{2\,Sensitivity \times Precision}{Sensitivity + Precision} \tag{19}$$

$$\text{Overall accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

where TP (true positive) is the correctly classified positive leather samples, TN (true negative) is the correctly classified negative leather samples FP (false positives) is the incorrectly classified negative samples and FN (false negative) is the incorrectly classified positive leather samples.

6 Results and discussion

In this work, a deep learning computing system involving 64-bit Windows 10 system with i5 CPU, 16 GB memory, and 2.30 GHz basic frequency is used for developing the deep learning neural networks and machine learning models for multi class classification of leather textures. Proposed deep learning neural network models and machine learning models using the programming tools with MATLAB software (Version.2021b). Training and testing performance of popular pretrained deep learning neural networks such as Alexnet, VGG-16, Google net, Squeeze Net, ResNet-50 are evaluated for the classification of folding marks, grain off, pinhole, growth marks, loose grain, and non-defective leather defects. Further, the accuracy of classification of deep learning convolutional neural networks is compared with the existing machine learning approaches like shallow feed forward neural network, support vector machine and K-nearest neighbor. Classification performance metrics such as precision, sensitivity, f1-score, and accuracy are calculated using confusion matrix and the results are presented in this section.

6.1 Feature maps of convolution neural networks

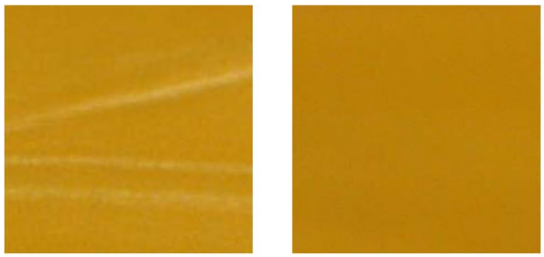
In order to understand and interpret the feature maps in the convolution layers of the Alexnet, VGG-16, Google net, Squeeze Net, ResNet-5, it is extracted for few convolution layers. Figure 8 shows feature maps of a leather image in Alexnet for the 5 convolution layers. It can be seen in Fig. 8a that, simple features like edges are filtered by kernels in the first convolution layer and high-order features are extracted in the subsequent layers using the learned weights of the kernels as shown in Fig. 8b–e.

As layer depth increases the feature maps does not show much details due to the finer size of the filter in the same receptive field. Though it is difficult to interpret the extracted feature maps, it provides low level and high-level features of leather texture variations in the leather image at the same receptive field which is used for classification of leather images.

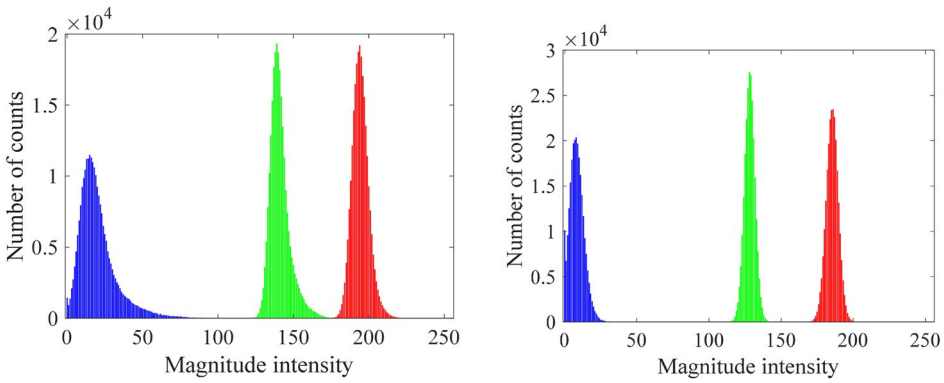
6.2 Feature extraction using GLCM, autocorrelation

Colour histogram, GLCM, autocorrelation functions are applied extract the colors, texture features of the defective, non-defective leather images for evaluating the performance of the machine learning approaches such as for multi class leather defect classification. Figure 9 shows feature extraction results of typical defective leather with a fold mark and a non-defective leather image of yellow colour. As the presence of defects in the leather results in intensity variations, there is a variation in the number of counts of red, green, blue channel as shown in Fig. 9b. From Fig. 9c, it can be noted that hence the autocorrelation function of defective leather with folding marks decays slowly due to the coarser texture as compared to the non-defective leather with a finer texture as shown in Fig. 9c.

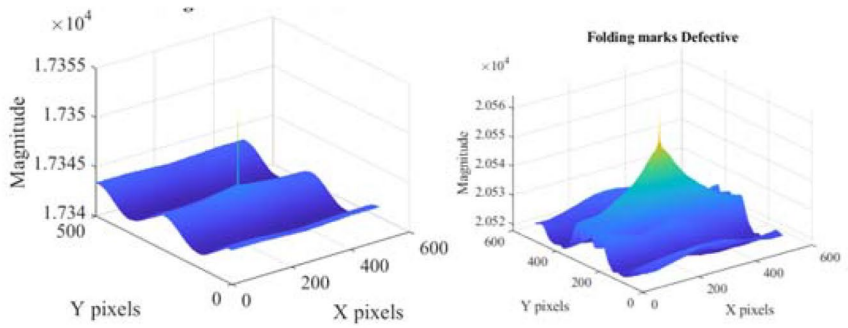
As there is a differing texture pattern in defective and non-defective leather, it results in variations in intensity variations and grey levels of neighbouring pixels which leads to change in magnitude of GLCM as shown in Fig. 9d. From these results, it is noted that the extracted color and texture features are varying due to the differing texture pattern of leather defects and the corresponding intensity variations.



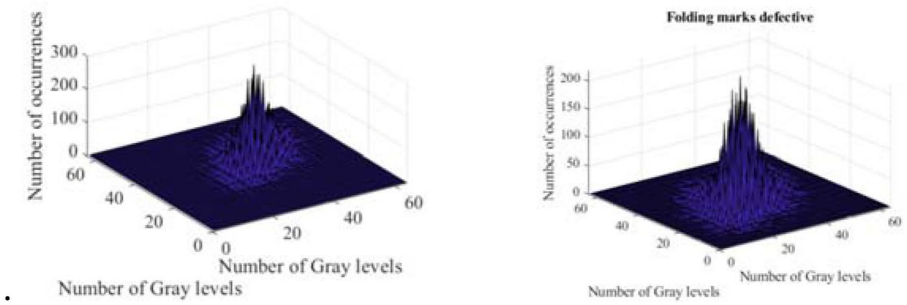
(a) Defective leather with folding mark and non-defective leather image



(b) Color histogram



(c) Autocorrelation



(d) GLCM

Fig. 9 Color and texture feature extraction for defective and non defective leather

Table 3 Parameters of the stochastic gradient descent with momentum (SGDM)

Hyper-parameters	Value
Optimization algorithm	SGDM
Initial learning rate	0.0001
Epochs	220
Batch rate	30

6.3 Training and testing performance of deep learning neural networks

In this work, Stochastic Gradient Descent with momentum (SGD) is used to optimize the model hyper-parameters, particularly the initial rate, stride, filter size of deep learning neural networks. Table 3 shows the training details of the neural network.

Figure 10 shows the training and testing performance curves for the different epochs and iterations of AlexNet. Bottom plot shows the cross-entropy loss function for different epochs of the training (blue) and testing (black) dataset, Top plot shows the trend and variation of the classification accuracy of AlexNet over epochs.

From the plots, it is noted that the training process converged well to reach the classification accuracy of 99.4%. It can be seen that the magnitude of loss reduces for each epoch with the enhance in accuracy of classification. Accuracy and loss function shows bumps as weights of the neural networks are learnt from the given examples of training and testing leather images for multi class classification. The elapsed

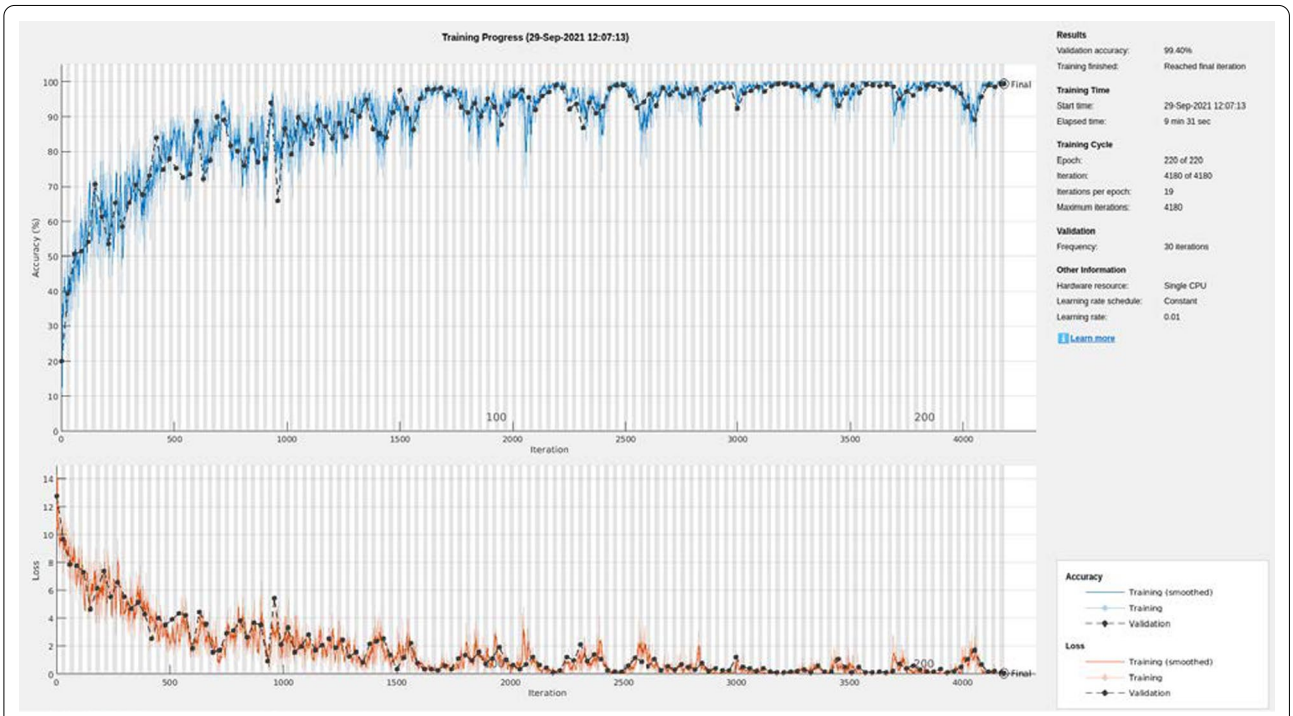
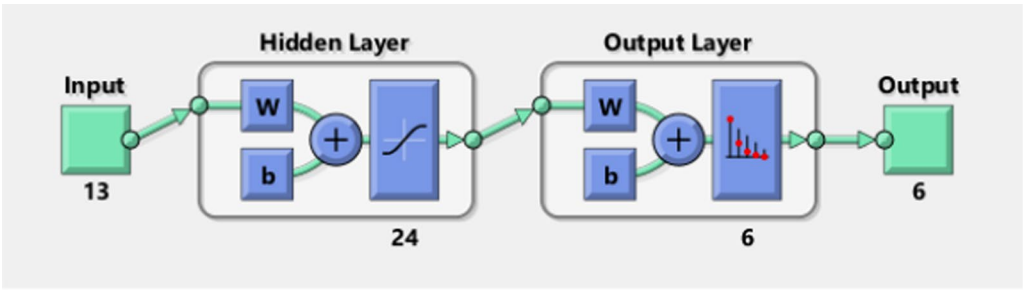


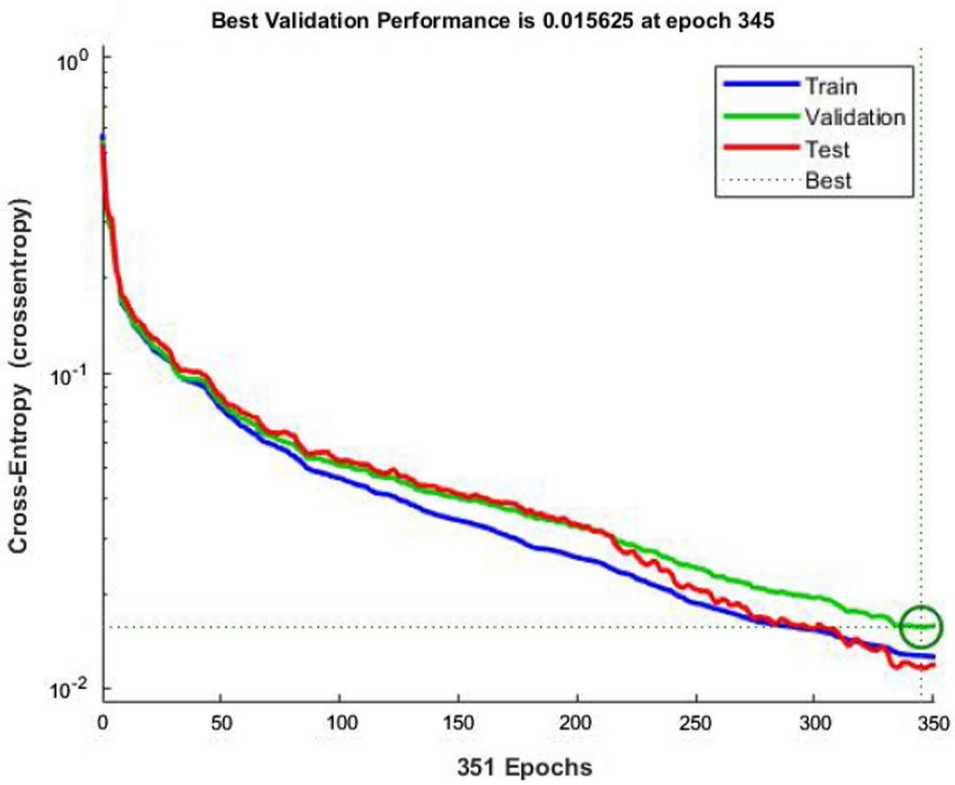
Fig. 10 Training and testing performance of AlexNet architecture

Table 4 Computational training time of the Alexnet deep learning model

Epoch	Iteration	Time elapsed (hh:mm:ss)	Mini-batch accuracy (%)	Validation accuracy (%)	Mini-batch loss	Validation loss	Base learning rate
1	1	00:00:03	25.78	21.15	10.2304	12.5705	1.000e−04
12	210	00:00:31	60.16	72.98	6.2358	4.3037	1.000e−04
21	390	00:00:55	82.03	73.17	2.7172	4.2595	1.000e−04
124	2340	00:05:18	96.88	94.60	0.4982	0.8561	1.000e−04
220	4180	00:09:31	99.40	97.49	0.336	0.2404	1.000e−04



(a) Neural network architecture of leather defect classification



(b) Training and validation performance

Fig. 11 Configuration and training of SFFNN

computational time of the training process and the accuracy improvement, loss reduction is noted for different number of epochs and it is shown in Table 4.

It can be noted that mini batch loss reduces as the weight values are learnt for the correct classification with the increase in number of iterations and the classification accuracy reaches 99.40% at 220 Epoch.

6.4 Training performance of shallow feed forward neural network classifier

In this work, a shallow neural network architecture with 13 neurons in input layer, 24 neurons in the hidden layer are developed and trained in MATLAB environment Fig. 11 shows the architecture of the proposed neural network and the output layer is configured with 6 neurons for classification of type of leather image with folding marks, grain off, pinhole, growth marks, loose grain and non-defective leather surface. Training and development of proposed shallow feed-forward neural network are carried out, a cross-entropy function and gradient descent method are used for adjusting the weight values of the neural network.

The training, testing and validation performance plot of the proposed SFFNN and the number of epochs is

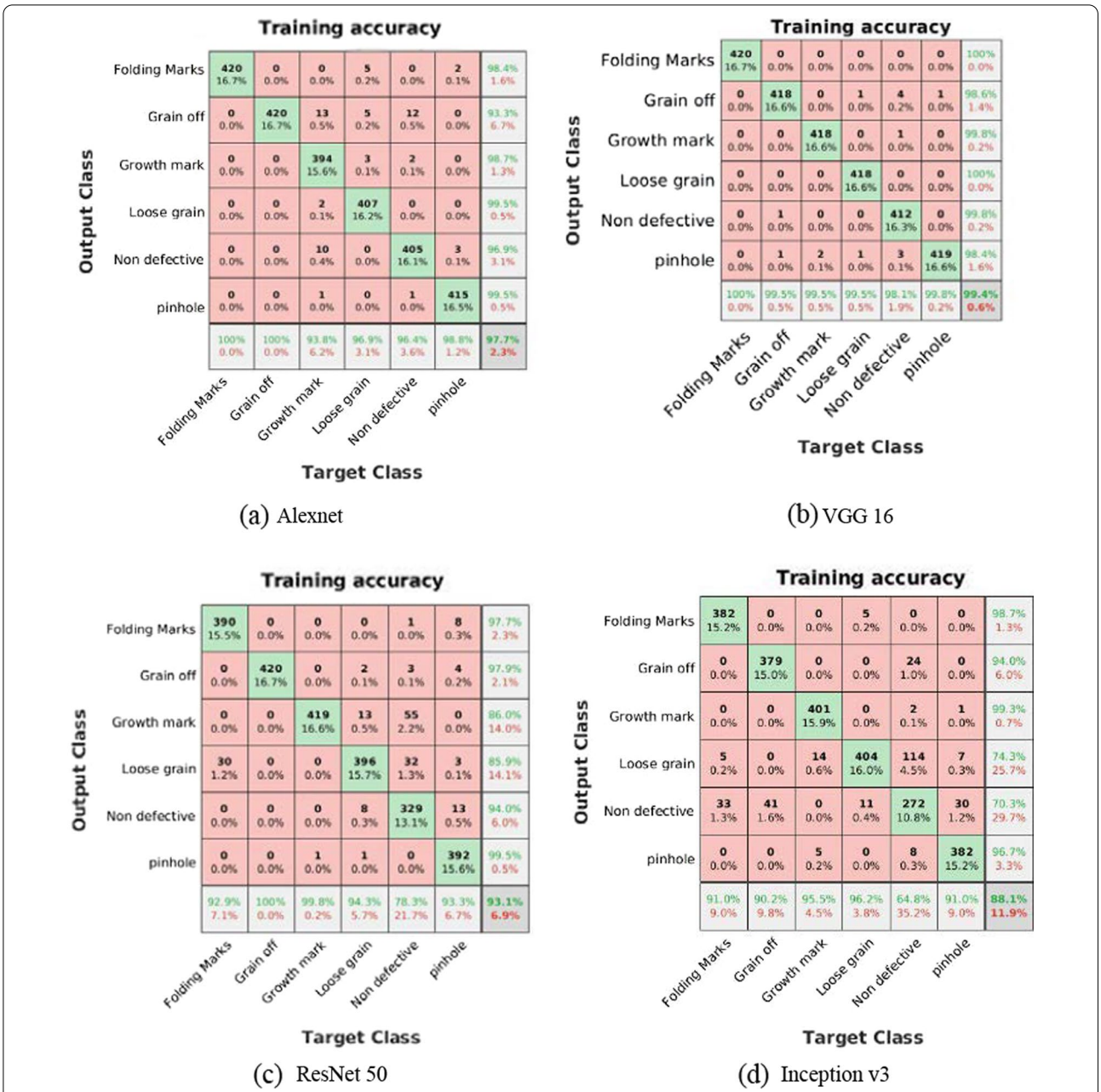


Fig. 12 Comparision of classification accuracy of the deep learning convolution neural network classifiers

shown in Fig. 11b. It is found that the classification accuracy of 97.6% for the minimum cross-entropy value of 0.015625 and best validation performance is achieved at 345 epochs.

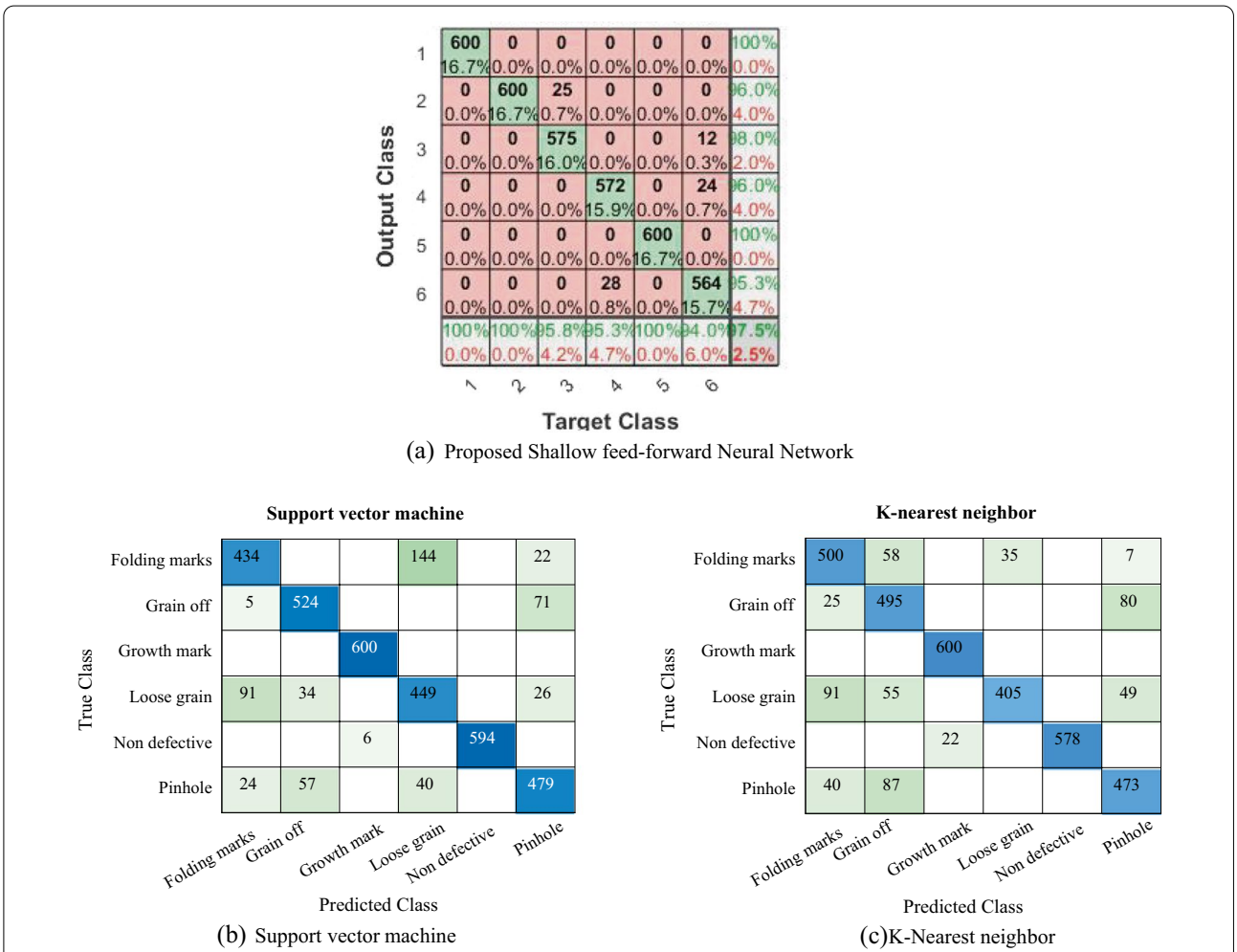
6.5 Classification performance of deep learning neural networks

In order to quantify the classification performance of the deep learning convolution neural networks for the multi class classification of leather texture, confusion matrix is calculated based on the number of output classes given by the classifier and the given target classes of leather images. Figure 12 shows the confusion matrix for deep learning neural networks and machine learning classifiers which provided top three highest classification accuracies during training and testing.

In Fig. 12, high numbers in green cells represent correct responses and the low numbers in red cells correspond

Table 5 Comparison of classification accuracy of deep learning neural networks

Leather classes and accuracy	AlexNet		VGG16		ResNet 50		Inception v3	
	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Folding marks	1	1	1	0.98	0.92	0.97	0.90	0.98
Grain off	0.99	0.97	1	0.93	0.99	0.97	0.90	0.94
Growth mark	0.99	1	0.93	0.98	0.94	0.86	0.95	0.99
Loose grain	0.91	1	0.96	0.99	0.78	0.85	0.96	0.74
Non defective	0.99	0.94	0.96	0.96	0.93	0.94	0.64	0.70
Pinhole	0.99	0.97	0.98	0.99	0.91	0.99	0.90	0.96
Training accuracy	99.40%		97.7%		93.10%		88.1%	
Testing accuracy	87.60%		85.80%		83.60%		80.0%	



(a) Proposed Shallow feed-forward Neural Network

(b) Support vector machine

(c)K-Nearest neighbor

Fig. 13 Comparision of classification accuracy of the machine learning classifiers

to incorrect responses. The percentage values in the far-right column of the confusion matrix shows the precision (or positive predictive value) and false discovery rate of the proposed neural network for the classification of each class of leather. Further, the bottom row of the confusion matrix shows the sensitivity and false-positive rate. Due to the superior feature extraction capability with the kernel structure, Alexnet performed better with

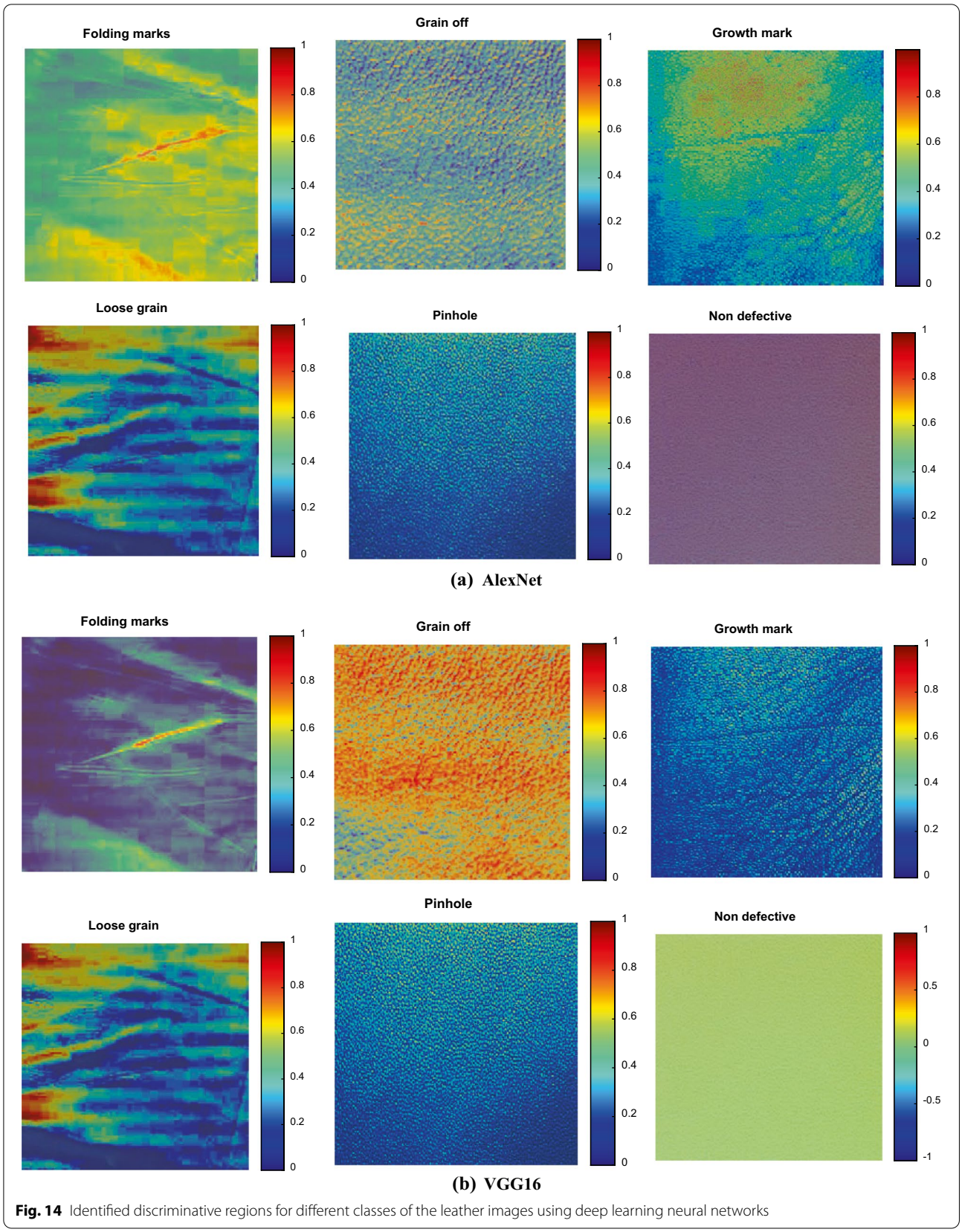


Fig. 14 Identified discriminative regions for different classes of the leather images using deep learning neural networks

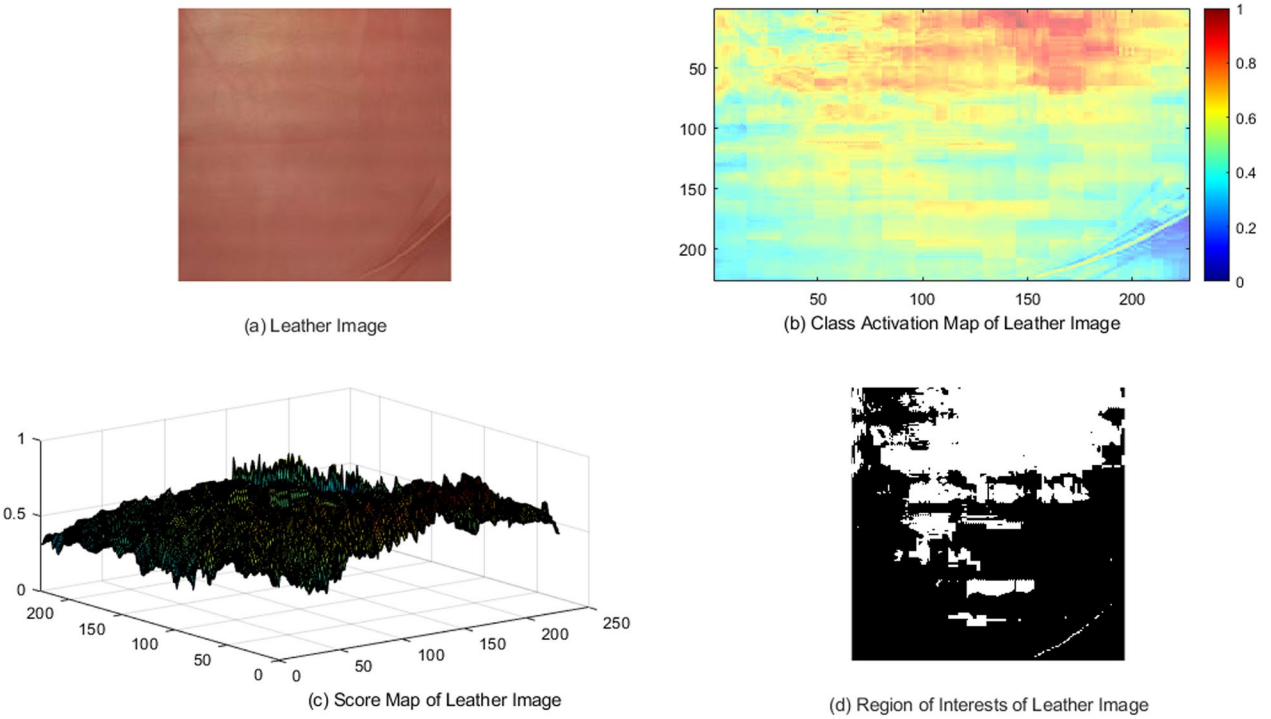


Fig. 15 Identified discriminative regions for different classes of the leather images using AlexNet

the classification accuracy of 99.40% and 87.60% for the training and testing data sets than the other deep learning convolution neural network classifiers.

Using the number of correct and incorrect classifications of the target class of the leather images, the precision, recall values are calculated and listed in Table 5 for different classifiers.

With the better testing accuracy of 87.60%, Alexnet performed better for multi class classification of leather textures of unseen leather images. These results proved the capability of deep learning neural network classifiers for the application of multi class leather texture classification.

6.6 Classification performance of machine learning approaches

State-of-the-art machine learning algorithms like shallow feed forward neural networks, support vector machine, K nearest neighbour, Decision tree, Naïve Bayes is applied using the hand-crafted colour and texture features for multi class leather texture classification. Confusion matrix is constructed for summarizing the performance of the different machine learning classifiers and it is shown in Fig. 13. As indicated by the higher magnitude of diagonal elements in the confusion matrix for the correct classification of each class, shallow feed forward neural network showed better performance in classification of different leather texture images. The overall classification accuracy of shallow feed forward neural network for multi class leather image classification is found to be 97.5% which is lesser than the deep learning convolution neural network models as the hand-crafted feature extraction limits the important discriminative features in the leather images.

6.7 Class activation maps for selection of region of interest in leather images

As the leather defects are localized in the specific regions of the leather image, class activation maps are generated using the trained deep learning neural network models for the given leather images and the sample results for different leather texture classes are shown in Fig. 14a, b for Alexnet and VGG-16 respectively.

It can be seen that the regions with the red color is identified as the discriminative regions and it can be identified from the pixels with highest magnitude using the score map. In this work, maximum value of the class activation map is chosen and applied as the threshold for segmenting the region of interests in the leather images. Figure 15a sample leather image with ground truth bounding box in the defective areas and the corresponding class activation map in Fig. 15b. It can be seen that score map shows peaks and highest values in the

respective regions of the image as identified in red color in Fig. 15b.

A threshold value of 0.5983 is selected for segmenting the region of interest and the results are shown in Fig. 15c. It is found that the area of discriminative regions in the leather as identified by the class activation map is higher than the ground truth bounding box. Hence, it requires suitable algorithms for the precise detection and localization of leather defects in the leather images.

7 Conclusions

Leather texture plays is an important role in deciding the quality of the leather products. This work presented deep learning convolutional neural networks and machine learning classifiers for the multi class classification leather images. A comprehensive data set of 3600 leather images with different defects such as folding marks, grain off, pinhole, growth marks, loose grain, and non-defective leather surfaces are classified using pretrained deep learning neural networks such as Alexnet, VGG-16, Google net, Squeeze Net, ResNet-50. Performance of classification of deep learning convolutional neural networks is compared with the existing machine learning approaches like shallow feed forward neural network, support vector machine and K-nearest neighbour. From the results obtained from the confusion matrix, it is found that the deep learning convolution neural network like Alexnet performed better with the classification accuracy of 99.4% than the shallow feed forward neural network machine learning technique due to the superior feature extraction capability. Further, the use of class activation maps of the trained deep learning neural network for segmenting the regions of interest in the leather images is demonstrated for the localization of the defective regions. Proposed method can be suitably implemented for automated multi class classification of leather samples in an industrial environment.

Acknowledgements
Authors thank the VIT management for providing necessary experimental facilities to carry out research in leather inspection. Authors also thank K.H Leather Company, Vellore for providing leather samples for inspections.

Authors' contributions
PKM carried out the image processing and algorithm development for leather defect classification. DASS provided the frame work for neural network model for training and testing application of leather defect classification. Both authors read and approved the final manuscript.

Funding
Not applicable.

Availability of data and materials
A leather image data set containing different defects such as folding marks, grain off, pinhole, growth marks, loose grain, and non-defective leather surfaces is available at Kaggle: https://www.kaggle.com/praveen2084/leather-defect-classification.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 15 March 2021 Accepted: 21 January 2022
Published online: 16 March 2022

References

1. Lovergine FP, Branca A, Attolico G, Distance A. Leather inspection by oriented texture analysis with a morphological approach. In: Proceedings of international conference image processing, vol 2. 1997. p. 669–671.
2. Yeh C, Perng D-B. A reference standard of defect compensation for leather transactions. Int J Adv Manuf Technol. 2005;25(11–12):1197–204.
3. Newman TS, Jain AK. Survey of automated visual inspection. Comput Vis Image Underst. 1995;61:231–62.
4. Georgieva L, Krastev K, Angelov N. Identification of surface leather defects. In: Proceedings of the 4th international conference on computer systems and technologies e-learning. 2003. p. 303–307.
5. Kavitha C, Karthika K, Hamsaveni VG, Supriya B, Manasa BMR. Identification and classification of defects at different leather processing stages using vision based system. Int J Adv Sci Technol. 2019;28(19):123–8.
6. Mihran T, Anil J. Texture analysis. In: Chen CH, Pau LF, Wang PSP, editors. The handbook of pattern recognition and computer vision. Singapore: World Scientific Publishing Co; 1993. p. 235–276
7. Jawahar M, Chandra Babu NK, Vani K. Machine vision inspection system for detection of leather surface defects. J Am Leather Chem Assoc. 2019;114(1):10–19.
8. Sobral JL. Leather inspection based on wavelets. In: Marques JS, de la Blanca NP, Pina P, editors. Pattern recognition and image analysis. Berlin: Springer; 2005. p. 682–8.
9. Branca A, Lovergine FP, Attolico G, Distante A. Defect detection on leather by oriented singularities. In: International Conference on Computer Analysis of Images and Patterns. Berlin: Springer; 1997. p. 223–230.
10. Aslam M, Khan TM, Naqvi SS, Holmes G, Naffa R. On the application of automated machine vision for leather defect inspection and grading: a survey. IEEE Access. 2019;7:176065–86.
11. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;6:610–21.
12. Jawahar M, Babu NKC, Vani K. Leather texture classification using wavelet feature extraction technique. In:Proceedings of IEEE international conference on computational intelligence and computing research. 2014. p. 1–4.
13. Hemerson P, Willian PA, Priscila SM, Mauro CP, Pereira MA, Jacinto MAC. Defect detection in raw hide and wet blue leather. In: CompIMAGE, computational modelling of objects represented in images: fundamentals methods and applications, Coimbra. 2007. p. 355–360.
14. Viana R, Rodrigues RB, Alvarez MA, Pistori H. SVM with stochastic parameter selection for bovine leather defect classification. In: Mery D, Rueda L, editors. Advances in image and video technology. PSIVT 2007. Lecture notes in computer science, vol. 4872. Berlin: Springer; 2007.
15. Kwak C, Ventura JA, Tofang-Sazi K. Automated defect inspection and classification of leather fabric. Intell Data Anal. 2001;5(4):355–70.
16. Jawahar M, Babu NKC, Vani K, et al. Vision based inspection system for leather surface defect detection using fast convergence particle swarm optimization ensemble classifier approach. Multimed Tools Appl. 2021;80:4203–35.

17. Villar P, Mora M, Gonzalez P. A new approach for wet blue leather defect segmentation. In: San Martin C, Kim S-W, editors. CIARP 2011, LNCS 7042. Berlin: Springer; 2011. p. 591–8.
18. Amorim WP, Pistori H, Pereira MC, Jacinto MAC. Attributes reduction applied to leather defects classification. In: Proceedings of 23rd SIBGRAPI conference on graphics, patterns and images. 2010. pp. 353–359.
19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2015. arXiv: 0706.1234.
20. Ren R, Hung T, Tan KC. A generic deep-learning-based approach for automated surface inspection. IEEE Trans Cybern. 2018;48(3):929–40.
21. Liong S, Gan YS, Huang Y, Liu K, Yau W. Integrated neural network and machine vision approach for leather defect classification. 2019. arXiv: 1905.11731.
22. Liong S-T, Zheng D, Huang Y-C, Gan Y. Leather defect classification and segmentation using deep learning architecture. Int J Comput Integr Manuf. 2020;33(10):1–13.
23. Xu D, Wang C-S, Wu C-F. The recognition of color and texture features based on the senses of vision and touch. In: Proceedings of 13th international conference on language, education, humanities and innovation & 2nd international conference on open learning and education technologies. 5th & 6th April, 2019. 2019. p. 22–29.
24. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8:53. https://doi.org/10.1186/s40537-021-00444-8.

Publisher's Note