AUTOMATIC DEFECT DETECTION SYSTEM FOR FINISHED LEATHERS USING DEEP LEARNING

P.Anandan[1], K.S.Murali[2], M.Naveen[3], V.Akshay[4],
[1]Associate Professor, [2,3,4]UG scholars,
Department of Electronics and Communications Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology Chennai, India
[1]anand.phd.dip@gmail.com, [2]muralishankar556@gmail.com, [3]naveenreddy12343@gmail.com
[4]sunny65184@gmail.com

ABSTRACT

Purpose - This Purpose of this paper is to focus on the fully automatic defect detection system that uses a convolutional neural network on the finished leather to provide defect-free leather.

Design/methodology/approach - The camera then accumulates a sample of the leather, trains and tests the image using a deep learning architecture. The histogram gradient preprocessing has been introduced to improve the quality of the leather image to maintain the display of important features from a perceptibility viewpoint. After that convolutional neural network classifiers are used to distinguish Between a defective leather area and a non- defective leather area.

Findings - Leather is a natural and strong material created by tanning animal skins and skins. Leather pricing is subjective because it is highly dependent on quality. of the leather and the condition of surface imperfections. Before the mass manufacture of certain goods, there was little work to detect leather defects. But it's annoying because it's labor-intensive, time-consuming, eye strain, and often prone to human error. A manual error checking process is essential in the leather manufacturing industry as a quality control step.

Originality/value - The proposed method can generate a significant accuracy rate from the 1100 leather patch sample collection.

Keywords - Convolutional Neural Network (CNN), Defect detection, Histogram Gradient, Leather patch.

Paper type - Research paper

1.INTRODUCTION

Indian leather and leather products have a standing place in India's economy. India's revolutionary community occupies about 13% of worldwide leather production and handles about 3 billion square leather production. It is important to ensure high quality of leather to improve customer satisfaction due to the importance of leather in the industrial business. Raw substantial excellence is usually negotiated by ante-mortem defects (scratches, rubs, horn rakes, yoke marks, scabies, smallpox, burn marks, etc.) or post-mortem defects (flame cuts, meat cuts, grain cracks). Leather quality plays an important role in the amount of the leather manufactured, accounting for about 55-75% of the manufacturing cost, so raw materials are the most valuable and important factor of production.

In short, the basic steps for transforming the leather skins are: (1) Immersion: Dirt and pickle salt from soaking rawhide in water, for example Hours to days; (2) Emerald: Cuticle, hair and subcutaneous substances; (3) Browning: Creates protein crosslinks in collagen by penetrating Chemicals in the cowhide; (4) Drying: To remove excess water; (5) Coloring: Choose the appropriate convention colour. Few of nature's flaws were previously unnoticed. They appear to be progressively becoming evident under leather finishing throughout the tanning process. In addition, fault areas with minor damage are fixed and roughened with a filler to produce a smooth and consistent surface. Finally, the completed portion of leather is pre-classified prior to shipping to consumers One of the most crucial and challenging procedures is grading, which entails a manual examination to physically evaluate problematic sections of the leather.

The most significant parts of quality control are defect type, defect size, and defect severity. Examiners are required to accomplish numerous detailed manual assessments of to assure authenticity and integrity, the identical piece of leather was photographed from several perspectives, distances, and lighting conditions. However, keep in mind that each decision is subjective, as it differs significantly from individual to individual. As a result, human testing is expensive, time consuming, wasteful, and unreliable. Because of this dull and laborious duty, or when the operator is anxious and completes the job in a rush, human error can be more expected to occur. Therefore, it is critical to build an automatic leather defect inspection system in order to enhance the grading and inspection process while saving money. The ultimate purpose of this paper is to identify the defects in finished leather samples.

The ultimate goal of this paper is to classify leather Samples into either a defective class or a non-defective class. The four main contributions to this research work

In summary:

(1) Preparing the image dataset by image acquisition.

(2) Then proposed preprocessing step and Convolutional Neural Network Classifier for classifying defective leather patches

(3) Wide-ranging experimental evaluation and relative analysis have been conducted on more than 1100 leather samples.

(4) Demonstration of capable classification outcomes

By reporting all performance metrics. The following sections are organized in the rest of the paper: Section 2 contains an overview of linked publications. Section 3 then describes the suggested technique. Contains client intuition as well as a description of the image processing technique employed. In Section 4 Evidence about critical performance indicators, the dataset utilized the experiment's parameter setup is as follows, which has been made available. The classifier's performance is existing, discussed in Section 5, and further examined. Finally, the conclusions may be found in Section 6.

2.LITERATURE SURVEY

Most flaw detection and annotation operations in the industry are still carried out by highly skilled examiners. Visual inspection includes two automated tasks: categorization and instance segmentation. The preceding categorizes the many types of Rawhide flaws such as scratches, tick bites, folds, scabies, and so on. Others; the latter identifies defective areas and within the time being, listens to the defect's appeal Despite the fact that there has been a lot of previous research on leather fault categorization, only a few researchers anticipate this. The exact position of the flaw This is frequently worth highlighting. Many of the experiments described in the paper were put to the test. Image contains or is identified with only one sort of defect Single defect present in sample image.

Kwon et al. [1] Proposed a framework for identifying multiple defect types (holes, pinholes, cuts, wrinkles, etc.) reinforced with a histogram of pixel intensity values. She discovered that the right leather image pixel composition has to represent a standard statistical circulation. For hole faults, their normal distribution is z. Image pixels are usually concentrated in bright areas (That is, near the pixel value of 255). In contrast, pinholes Much darker pixels (that is, near the pixels) Value 0). Problems like scratches and wrinkles are normal. Shows a transparent pattern in comparison to the traditional distribution The quality of the leather presented will then be decided. To support the referenced analysis results of solidity and decimal of extricate defects.

On the other hand, **fuzzy logic is employed** [2] to research feature sets of leather images for performing surface defect detection. The leather picture is initially encumbered in grayscale and it's represented by a histogram. in particular, the choice of the histogram, the position of the histogram, the median, the mean, the variance, the energy, the entropy, the contrast, etc. are calculated as maximum, minimum, average value. However, the sample size is little (that is, the image). The procedure of the experiment is ambiguous. for B. Explanation of experimental equipment Distribution of coaching sets and test sets doesn't include evaluating the proposed algorithm paper.

Recent study by **Winiarti et al. [3]** uses both to classify five varieties of leather. Handmade feature descriptors and deep learning architecture are two feature extractors. Type of leather Monitor lizards, crocodiles, sheep, goats, and cows are among them. In the instance of craft expression, statistical fusion Color attributes (i.e., mean, standard deviation, skewness). And kurtosis) as well as statistical textural characteristics (i.e, contrast, energy, correlation, uniformity, and entropy). Classification performance suggests that it was accepted. As a deep learning structure, a pre-trained Alex Net is employed. learning methods are better at capturing Overall leather characteristics with 99.97% accuracy. Please note that this document does not include error classification as all data are error free Picture.

Similar defect classification work is performed by **Pistori et al. [4].** In particular, they tend to discriminate 4 types of defects: Ticks, cuts, itching, and markings blue leather traces of both raw and wet skin with a hot iron The former looks to be more refined. Various types of surfaces (textures, colors, shapes, the hindmost is a most usual type of cowhide that has been treated (even if it is thick and has serious imperfections). The browning process that seems to be more obvious for visual inspection of both humans and machines. Image features are extracted using a common texture analysis technique, that is, grayscale co-occurrence. Matrix (GLCM). As a result, a significant classification result is achieved using the Support Vector Machine classifier.

Pereira et al. [5] aim to decrease differences amongst leather classifications and the goat skin expert assessments is to increase quality productivity Classification by Machine learning classifier and feature extractor The proposed approach covers picture capture, image preprocessing, and image enhancement, and functional processes. Extraction and classification of machine learning. In short, A new approach has been introduced to extract features called pixels The Intensity Analyzer (PIA) has proven to be the most cost-effective method when used in combination with this problem. Extreme Learning Machine (ELM) classifier. However, the details of setting up the experiment are missing. Duplicating and comparing frameworks can be difficult for other works.

For debugging and segmentation tasks **Lovergine et al. [6]** conduct one of the frontier studies. They notice and identify problematic regions. Using a CCD camera in black and white the gathered pictures are then subjected to the morphological segmentation technique [16, 17] to obtain texture orientation characteristics. Leather. A sample defective leather image is shown in Figure 1. Paper, but quantitative methods and numbers There is not enough data to evaluate the proposed algorithm.

The goal of this research is to develop an automated defect detection system for segmenting defect areas of leather. Certain types of defects, namely tick bites. Such defects appear as small surface damage on animal skin.

Fig 1 Example of a corrupted image

Figure 1 shows an example of a corrupted image. Instance segmentation Using a deep learning model, a convolutional neural network (CNN), to develop a robust architecture for evaluating test datasets.

3.PROPOSED METHOD

The proposed automatic defect detection system included four steps.

3.1) Image acquisition.

3.2) Preprocessing procedure.

3.3) Classification using Convolutional Neural Network Architecture.

3.4) Calculating Performance metrics.

Figure 4 shows an overview of the process. In short, the image is first sent to a preprocessing step which is Histogram gradient. The classification task, on the other hand, uses cutting-edge supervised classifiers such as Convolutional Neural Network (CNN). The details of the preprocessing method and the mathematical derivation of the classifier above are described in detail in Section 3.2 and Section 3.3, respectively.

3.1) Image acquisition: We have collected 1100 defective and non-defective leather samples and prepared a dataset.

3.2. Preprocessing procedure: Figure 3 shows the preprocessing technique used in the project. Each step is described as follows. An image is shown in Figure 3 to illustrate the effect of the Preprocessing step.

Histogram gradient:

Histograms of oriented gradients, also referred to as HOGs, are feature descriptors like Canny Edge Detector, SIFT (Scale Invariant and have Transformation). it's employed in computer vision and image processing for beholding purposes. This system counts the occurrence of gradient directions within the localized part of the image. This method is extremely similar to the sting Histogram and Scale Invariant Feature Transformation (SIFT). The HOG signifier emphases on the construction or outline of the object and it is superior to any edge descriptor because it uses both gradient magnitude and angle to calculate features. For areas of the image, use the magnitude and direction of the gradient to get a histogram Procedure for calculating HOG features.

1.Gets the input image that calculates the HOG feature.

Resize the image to an image of 4096*256 pixels.

$$G_x(r,c) = I(r,c+1) - I(r,c-1)$$

$$G_y(r,c) = I(r-1,c) - I(r+1,c)$$

2. The gradient of the image is calculated. Gradients are obtained by combining sizes and angles from the image. Considering a block of 3x3 pixels, Gx and Gy are first calculated for each pixel. First, calculate Gx and Gy for each pixel value using the following formula.

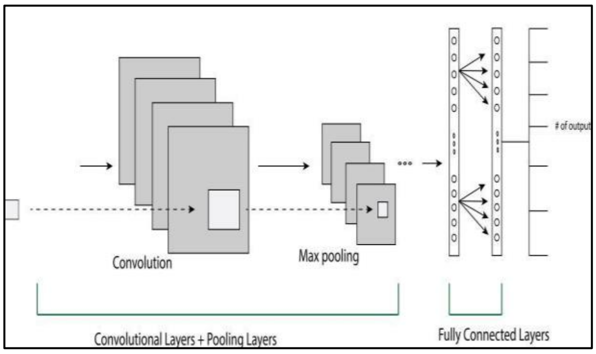
$$Magnitude(\mu) = \sqrt{G_x^2 + G_y^2} \quad Angle(\theta) = |tan^{-1}|(\frac{Gy}{Gx})$$

After calculating with Gx, the size and angle of each pixel is calculated using the following formula.

3.3 CNN Architecture

This section describes the proprietary CNN architecture utilized in this defect detection system. Because this system must be able to swiftly scan images collected from High quality camera in order to inspect the roll as quickly as possible, a simpler architecture was selected instead of other more complex state-of-the-art deep learning models. To boost system speed even more, only flaw detection is conducted at this step, leaving defect classification at a subsequent stage. Figure 2 depicts a high-level overview of the CNN layers and their size. This architecture is made up of four convolutional and max-pool layers,

Fig 2 CNN Architecture



which are followed by two fully-connected layers. All of the activities made use of ReLU. Table 1 lists all of the layers and their hyperparameters, where F denotes the number of feature mappings, K denotes the kernel size, and S is the stride parameter. As previously stated, this CNN was trained and validated using the Leather-Dataset. The network's activations at each layer were also displayed to ensure that it could recognize the leather defect features.

Table 1 The proposed custom CNN architecture hyperparameters

Layer	**Output**	**Hyperparameters**
INPUT	150*150*1	
CONV_1	150*150*64	F=64; K=5*5; S=1;
MAX-POOL_1	75*75*64	Pool=2*2; S=2;
CONV_2	75*75*64	F=64; K=5*5; S=1;
MAX_POOL_2	38*38*64	Pool=2*2; s=2;
CONV_3	38*38*124	F=128; k=3*3; S=1;
MAX_POOL_3	13*13*128	Pool=2*2; s=3;
CONV_4	13*13*128	F=128; K=3*3; S=1;
MAX_POOL_4	5*5*128	Pool=2*2; s=3;
Flatten	3200	
FC-2	256	Neurons=256
FC-2	128	Neurons=128
CNN OUTPUT	1	Optm=RMSPROP Loss=binary_crossentropy

Figure 3: The example of after applying the step of preprocessing method: Histogram gradient.

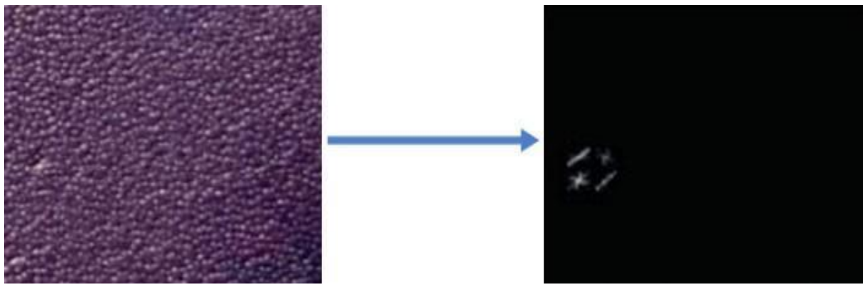


Figure 4: The proposed automatic defect detection system.

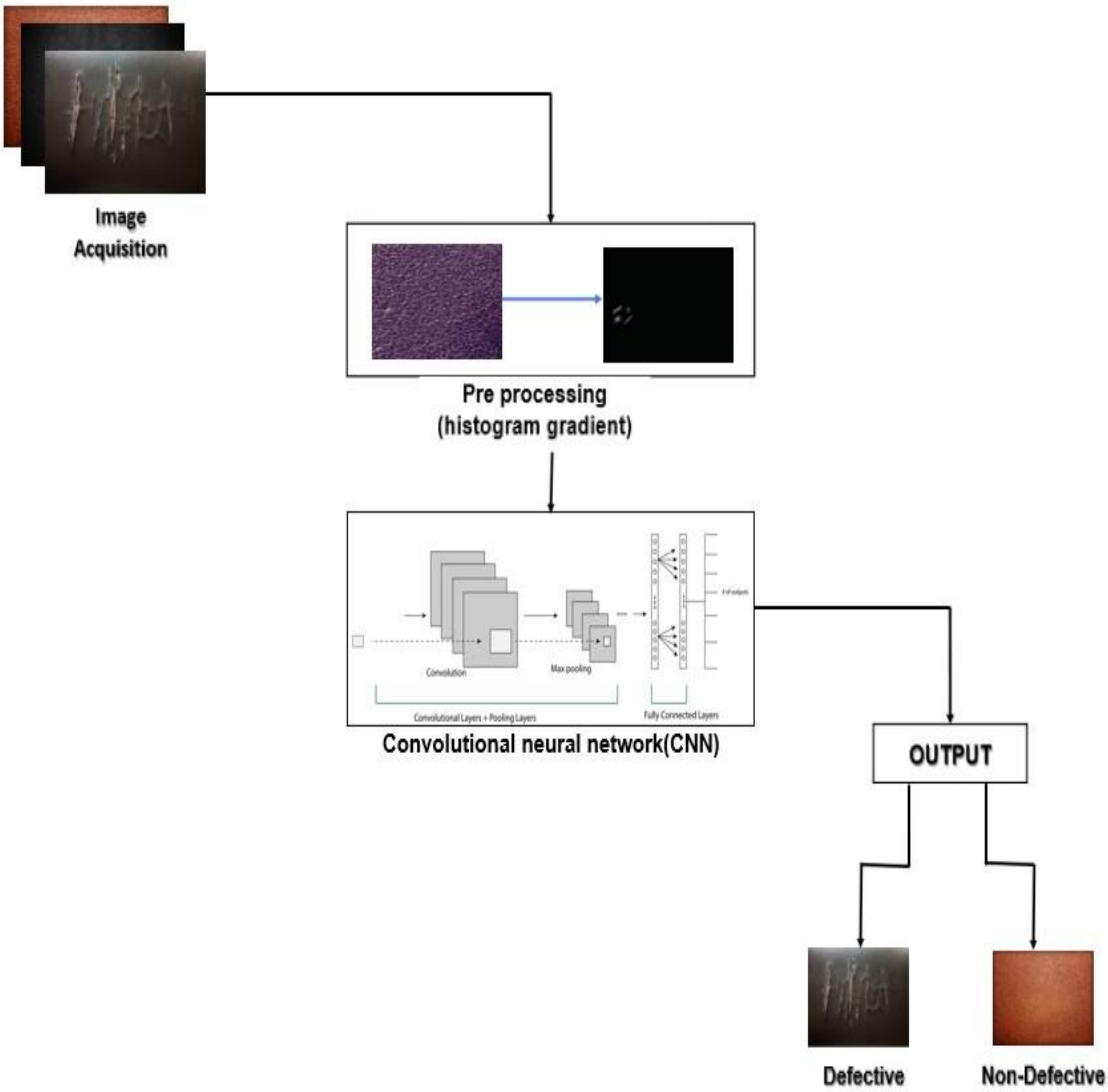


Figure 5 Sample leather images that contain (A) defective and (B) non defective.

3.4 Performance metrics

A) Confusion Matrix

Before going into the specifics of each statistic, let's first review the basic terminology utilized in categorization issues. The confusion matrix, which is a tabular depiction of the model predictions vs the ground-truth labels, is a crucial concept in classification performance. Each row of the confusion matrix represents occurrences from a predicted class, while each column represents instances from an actual class. Let's have a look at an example. Assume we're developing a binary classifier to distinguish between faulty and non-defective photos. Assume our data set contains 1100 leather sample pictures (1000 non-defective images and 100 defective images) and the confusion matrix shown below.

Table 2 Confusion matrix.

		Actual	**class**
		Defective	**Non-defective**
Predicted class	**Defective**	90	60
	Non defective	10	940

The program accurately predicted 90 of the 100 Defective pictures and misclassified ten of them. If we consider the “defective" class to be positive and the non-defective class to be negative, then 90 samples projected as cat are genuine positives and 10 samples predicted as non-cat are false negatives. The model successfully categorized 940 of 1000 non-defective photos and incorrectly classified 60. The 940 correctly classified samples are referred to as true-negative, while the 60 incorrectly classified samples are referred to as false-positive. As we can see, the diagonal parts of this matrix represent the right prediction for distinct classes, but the off-diagonal elements represent the samples that were misclassified. We have a better knowledge of the now that we have a better comprehension of the confusion matrix Let's move on from the confusion matrix and into the real metrics.

B) Accuracy

This is a binary classification issue, with the result labeled "defective" or "non-defective." As a result of the 2*2 confusion matrix, the four metrics listed below may be produced.

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

Accuracy = (90+940)/ (1000+100) = 1030/1100= 93.6%

where TP is the projected pixel that properly detects the faulty pixel; TN is the correctly predicted non-defective pixel; FP is the pixel that is forecasted incorrectly as a defective pixel; and FN is the untraversed defective pixel.

C)Precision

In many circumstances, classification accuracy is not a reliable predictor of model performance. When your class distribution is skewed, this is one of these circumstances (one class is more frequent than others). In this situation, even forecasting all samples as the most common class would result in a high accuracy rate, which makes no sense. For example, in our defective vs non-defective classification above, if the model forecasts all samples as non-defective, the model would score 1000/1100= 90.9 percent correct.

As a result, we must consider class-specific performance indicators as well. Precision is one of these measures, and it is defined as follows:

Precision = (True Positive + False Positive)/ (True Positive + False Positive).

In the preceding example, the accuracy of the Defective and Non-Defective classes may be computed as follows:

Precision Defective = #samples properly predicted as Defective/#samples predicted as Defective = 90/ (90+60) = 60%

Precision Non-Defective= 940/950= 98.9%

As we can see, the model predicts non-defective data significantly more accurately than defective. This is hardly surprising given that the model has seen more non-defective photos during training, making it stronger at categorizing that class.

D)Recall

Another significant metric is recall, which is defined as the proportion of samples from a class that the model correctly predicts. In more formal terms:

Recall = True Positive / (True Positive + False Negative)

As a result of our preceding example, the recall rate of defective and non-defective classes may be calculated as follows:

Recall defective = 90/100 = 90%

Recall Non-defective= 940/1000 = 94 %

E) F1-Score

Depending on the application, you may wish to prioritize recall or accuracy. However, there are numerous situations where both recall and accuracy are critical. As a result, it is logical to consider how to integrate these two into a single measure. The F1-score, which is the harmonic mean of accuracy and recall, is a common statistic that combines precision and recall.

F1-score = 2*Precision*Recall / (Precision + Recall).

So, for our classification example with the confusion matrix in Figure 1, the F1-score is

F1 Defective= 2*0.6*0.9 / (0.6+0.9) = 72 %.

The generalized definition of F-score is as follows. As we can see, F1-score is a subset of F_B when B = 1.

It is important to note that there is always a trade-off between accuracy and recall in a model; if you increase the precision too much, you will observe a decrease in the recall rate, and vice versa.

4.RESULT AND DISCUSSION

All tests were run on Python 3.6 with an Intel Core i5-9300H 2.40 GHz processor, 8.0 GB RAM, and an NVIDIA GeForce GTX 1080 Ti GPU. Table 3 displays the performance of the fault instance segmentation. To forecast the output, The same learnt CNN architecture is used in both the train and test datasets. The train and test datasets each include 100 and 1000 sample images. It is revealed that the train dataset has higher accuracy than the test dataset. The model, in particular, has accuracies of 95 percent for training data and 75.35 percent for testing data.

A receiver operating characteristic (ROC) can be used to indicate the efficacy of the classifier's statistical model in assessing the effects of the HoG during the preprocessing step. Figure 8 depicts the ROC curve when HoG is not employed as one of the preprocessing operations. The micro-average ROC is shown to be 69 percent accurate, whereas the macro-average ROC is 50 percent accurate. When HoG is included in the recommended approach, the ROC accuracies improve by up to 95%. The model's test accuracy is lower than its train accuracy. This is due to the trained model not seeing the test data, and the train data being the data used to train the model. The specificity and F1-score for the train dataset are both zero, as shown in Table 2. The train dataset is zero. This is because there will be no TN case in assessing the training data because the training images are confined to those with a defective region. In other words, images that are not faulty cannot be utilized as training data. Furthermore, the specificity in Table 2's test dataset is 75.35 percent, demonstrating that the majority of the testing images are not faulty and that the model is capable of correctly predicting them Tables 4 and 5 exhibit confusion matrices for the train and test datasets, respectively, to further evaluate performance. In summary, a confusion matrix is a common statistic used to depict the categorization rate for each faulty or non-defective example. Table 4 shows that 97 of the faulty areas were appropriately identified. Although there are 100 photos in total, it should be noted that each image may have more than one flaw. There are 104 occasions where the model fails to recognize the problematic regions appropriately. However, because of the high TN value the total testing accuracy is reasonable.

Table 3 Performance parameters for our proposed method

Performance metrics	Obtained result	
	Train dataset(%)	Test dataset(%)
precision	0.57	1.00
recall	1.00	0.62
F1-score	0.73	0.77
Accuracy	**0.75**	

Fig 7 Confusion matrix for train accuracy

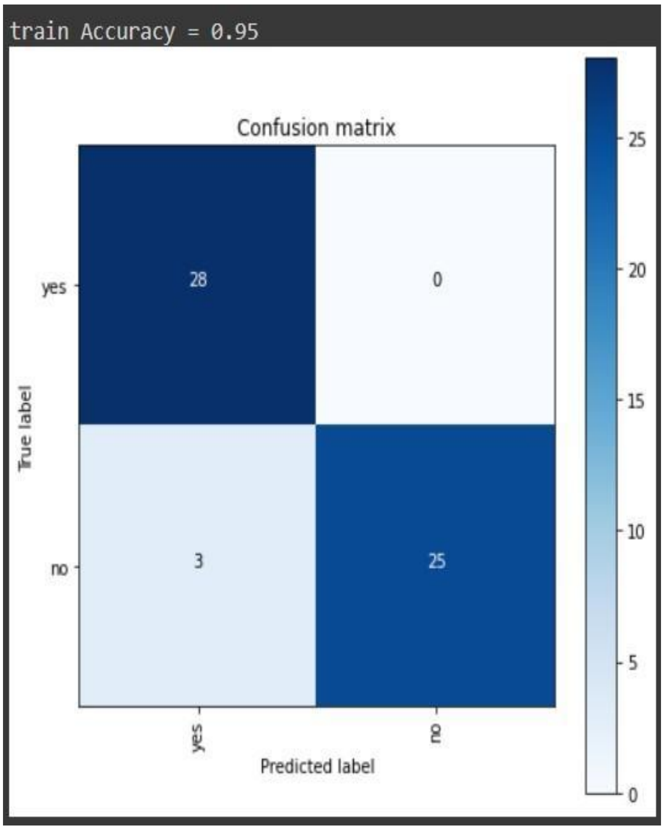


Fig 9 Plot for loss

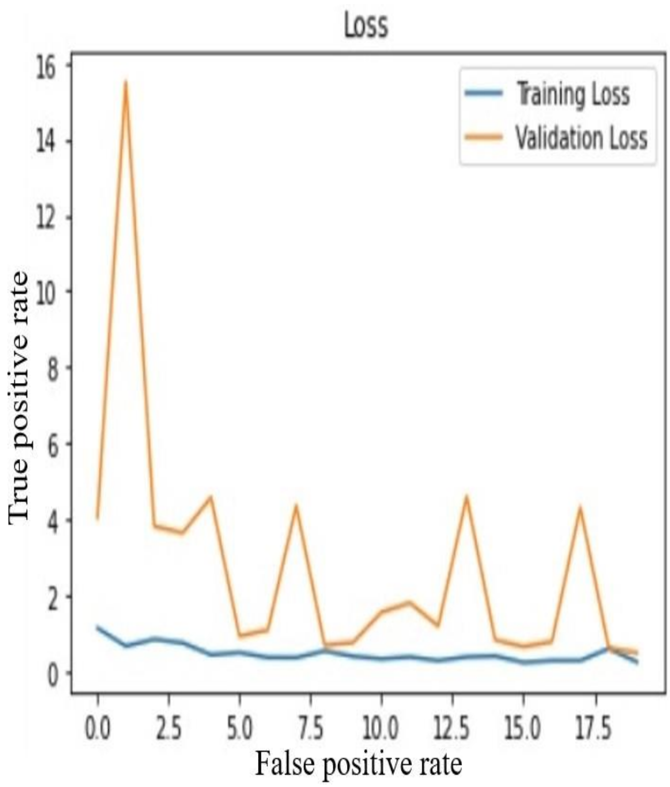


Fig 8 Confusion matrix for test accuracy

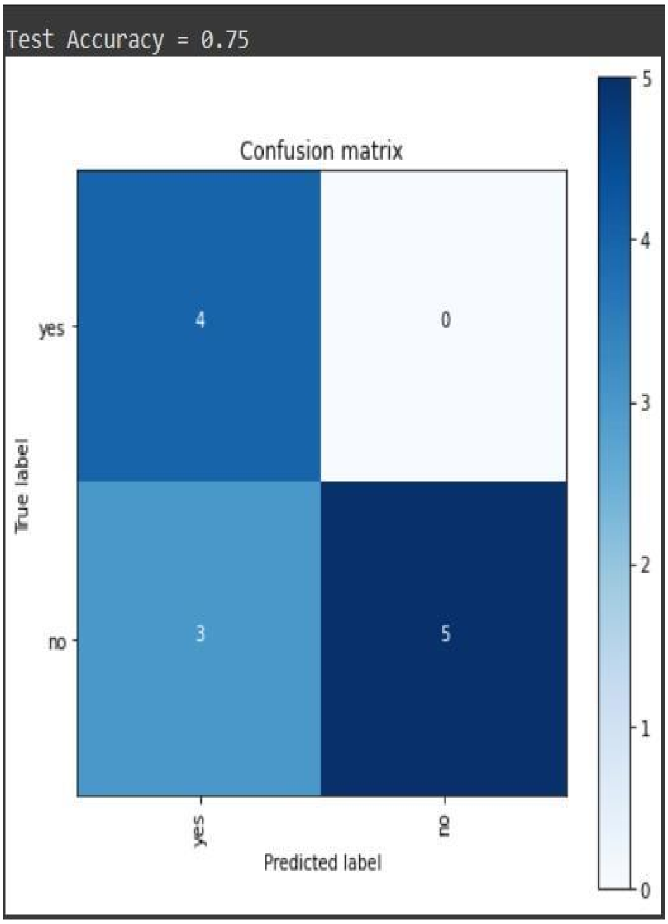
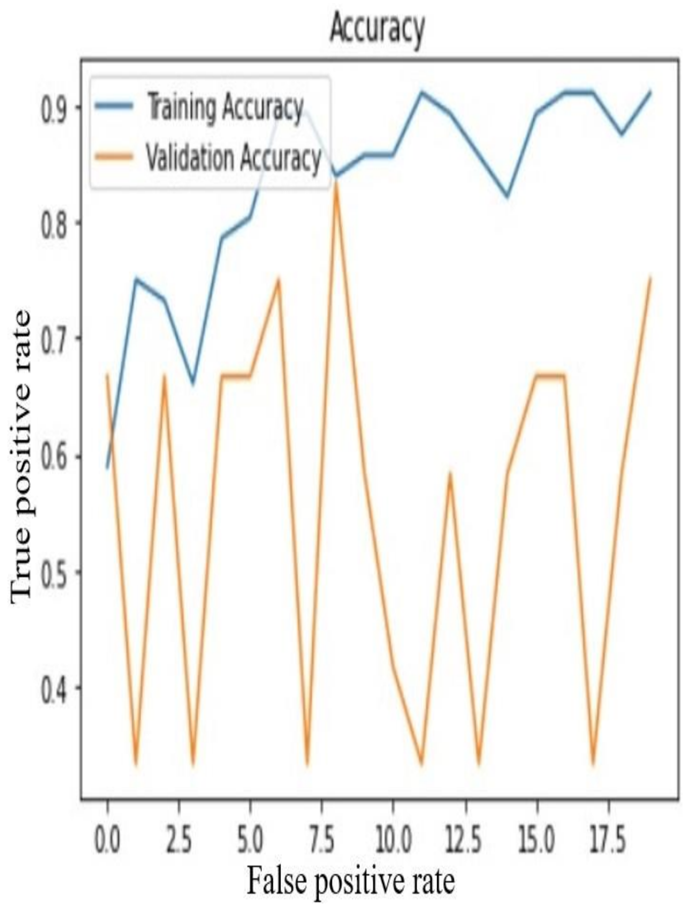


Fig 10 plot for accuracy



5.CONCLUSION

This work provided a binary classification method to determine whether a leather picture contains a defect or not. Extensive tests and analysis were performed to validate the resilience of the suggested technique. Overall, when preprocessing approaches and neural network classifiers are used, good results are obtained. As a consequence, when CNN is used as a classifier, the best classification accuracy attained is 95 percent. Because this experiment was confined to the defect type, further study in this area might involve the construction of a classification or segmentation method that is used to detect fault types such as open cuts, closed cuts, wrinkles, holes, and scabies. In addition to evaluating the fault type, the experiment may be expanded to test different leather types such as lamb, crocodile, and snake hides. A totally automated hardware and software setup consisting of the functions of gathering leather image patches, detecting troublesome locations, and laser cutting of the leather can be constructed in the future.

6.REFERENCES

1. Y.S. Gan, Wei-Chuen Yau, Sze-Teng Liong, and Chih-Cheng Chen, *Automated Classification System for Tick-Bite Defect on Leather*, Mathematical Problems in Engineering, Volume 22, Article ID 5549879 (2022).
2. M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, *On the application of automatic machine vision for leather defect detection and grading: a survey, IEEE Access*, vol. 7, Article ID 176065, (2019)
3. M. Jawahar, N. K. Babu, K. L. J. A. Vani, L.J. Anbarasi, and Geetha, *Vision Based Automatic Inspection System for Leather Surface Defect Detection Using Fast Convergence Particle Swarm Optimization Ensemble Classifier Approach, Multimedia Tools and Applications*, vol. 80, no. 3, pp. 4203–4235, (2020).
4. S. T. Liong, Y. S. Gan, Y. C. Huang, C. A. Yuan, and H. C. Chang, *Automatic Leather Defect Segmentation on Leather with Deep Learning*, March 2019, https://arxiv.org/abs/1903. 12139, (2019)
5. Y. Gan, S.-S. Chee, Y.-C. Huang, S.-T. Liong, and W.-C. Yau, *Automatic leather defect detection using statistical approach on image intensity, Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 1–17, (2020).
6. Jiehang deng, jiaxin liu, changzheng wu1, tao zhong1, guosheng gu 1, and bingo wing-kuen ling 3, (Senior Member, IEEE), *A Novel Framework for Classifying Leather Surface Defects Based on a Paramete r Optimized Residual Network*, volume 8. IEEE access (2020)
7. Masood Aslam, Tariq m. Khan, (member, IEEE), Syed Saud Naqvi, Geoff Holmes, and raffia naffa, *Ensemble convolutional neural networks with knowledge transfer for leather defect classification in industrial settings*, volume 8 IEEE access, (2020).
8. Prahar M. Bhatt, Rishi K. Malhan, Pradeep Rajendran, Shantanu Thakur, *Image-Based Surface Defect Detection Using Deep Learning, Journal of Computing and Information Science in Engineering*, volume 21, (2021)
9. Edmilson Q. Santos Filho,Pedro Henrique Feijó de Sousa,Pedro P. Rebouças Filho2 · Guilherme A. Barreto,Victor Hugo C. de Albuquerque, *Estimation of Goat Leather Quality Based on Computational Visualization Techniques*, Springer (2019).
10. Sze-Teng LIONG, Y.S. GAN, Kun-Hong LIU, Tran Quang BINH, Cong Tue LE, Chien A WU, Cheng-Yan YANG, Yen-Chang HUANG. *Efficient Neural Network Approaches for Leather Defect Classification*, ARXIV (2019).
11. SHIH-YU CHEN, (Member, IEEE), YU-CHIH CHENG, WEN-LONG YANG, AND MEI-YUN WANG, *Surface Defect Detection of Wet-Blue Leather Using Hyperspectral Imaging*, IEEE,2021.
12. Renato F. Pereira,Claudio M. S. Medeiros,Pedro P. Rebouc¸as Filho, *Goat Leather Quality Classification Using Computer Vision and Machine Learning*,IEEE,2018
13. Murinto Murinto, Adhi Prahara,Sri Winiari,Dewi Ismi, *Pre-Trained Convolutional Neural Network for Classification of Tanning Leather Image*, RESEARCH GATE,2018.
14. Qingsheng Jiang, Dapeng Tan, Yanbiao Li, Shiming Ji, Chaopeng Cai and Qiming Zheng, *Object Detection and Classification of Metal Polishing Shaft Surface Defects Based on Convolutional Neural Network Deep Learning*, MDPI,2019.
15. Kaiming He,Xiangyu Zhang, Shaoqing Ren, Jian Sun, *Deep Residual Learning for Image Recognition*, international journal of computer integrated manufacturing ,2016