A New Hybrid Model for Artificial Intelligence Assisted Tire Defect Detection: CTLDF+EnC

Araştırma Makalesi/Research Article

Özcan ASKAR[1] , Ramazan TEKİN*[2]

[1]Batman University, Graduate School of Education, Department of Information Technologies,
[2]Batman University, Faculty of Engineering and Architecture, Department of Computer Engineering,
ozcanaskar@gmail.com, ramazan.tekin@batman.edu.tr
(Geliş/Received:04.04.2024; Kabul/Accepted:08.07.2024)
DOI: 10.17671/gazibtd.1465294

***Abstract*—** This paper focuses on an artificial intelligence based worn tire detection system proposed to detect cracks in the tires of vehicle drivers. Although drivers are generally aware of the importance of tire tread depth and air pressure, they are not aware of the risks associated with tire oxidation. However, tire oxidation and cracks can cause significant problems affecting driving safety. In this paper, we propose a new hybrid architecture for tire crack detection, CTLDF+EnC (Cascaded Transfer Learning Deep Features + Ensemble Classifiers), which uses deep features from pre-trained transfer learning methods in combination with ensemble learning methods. The proposed hybrid model utilizes features from nine transfer learning methods and classifiers including Stacking, Soft and Hard voting ensemble learning methods. Unlike X-Ray image-based applications for industrial use, the model proposed in this study can work with images obtained from any digital imaging device. Among the models proposed in the study, the highest test accuracy value was obtained as 76.92% with the CTLDF+EnC (Stacking) hybrid model. With CTLDF+EnC (Soft) and CTLDF+EnC (Solid) models, 74.15% and 72.92% accuracy values were obtained respectively. The results of the study show that the proposed hybrid models are effective in detecting tire problems. In addition, a low-cost and feasible structure is presented.

***Keywords*—** transfer learning, deep features, ensemble learning, tire cracks.

Lastik Kusurlarının Tespiti için Yapay Zeka Destekli Yeni Bir Hibrit Model: CTLDF+EnC

***Özet*—** Bu çalışma, araç sürücülerinin lastiklerindeki çatlakları tespit etmek için önerilen yapay zeka tabanlı bir aşınmış lastik tespit sistemine odaklanmaktadır. Sürücüler genellikle lastik diş derinliği ve hava basıncının öneminin farkında olsalar da, lastik oksidasyonu ile ilişkili risklerin farkında değillerdir. Ancak, lastik oksidasyonu ve çatlakları sürüş güvenliğini etkileyen önemli sorunlara neden olabilir. Bu makalede, lastik çatlağı tespiti için, önceden eğitilmiş transfer öğrenme yöntemlerinden elde edilen derin özellikleri topluluk öğrenme yöntemleriyle birleştirerek kullanan yeni bir hibrit mimari olan CTLDF+EnC (Basamaklandırılmış Transfer Öğrenme Derin Özellikler + Ensemble Sınıflandırıcılar) önerilmektedir. Önerilen hibrit model, dokuz transfer öğrenme yönteminden gelen özellikleri ve İstifleme, Yumuşak ve Katı oylama topluluk öğrenme yöntemlerini içeren sınıflandırıcıları kullanmaktadır. Endüstriyel kullanıma yönelik X-Ray görüntü tabanlı uygulamalardan farklı olarak bu çalışmada önerilen model herhangi bir dijital görüntüleme cihazından elde edilen görüntülerle çalışabilmektedir. Çalışmada önerilen modeller arasında en yüksek test doğruluk değeri %76,92 olarak CTLDF+EnC (İstifleme) hibrit modeli ile elde edilmiştir. CTLDF+EnC (Yumuşak) ve CTLDF+EnC (Katı) modelleri ile sırasıyla %74,15 ve %72,92 doğruluk değerleri elde edilmiştir. Çalışmanın sonuçları, önerilen hibrit modellerin lastik sorunlarını tespit etmede etkili olduğunu göstermektedir. Ayrıca, düşük maliyetli ve uygulanabilir bir yapı sunulmuştur.

***Anahtar Kelimeler*—** transfer öğrenme, erin öznitelikler, topluluk öğrenme, lastik çatlakları.

1. INTRODUCTION

Tire problems are a difficult to detect issue that can cause serious risks when traveling at high speeds. The tire is one of the vehicle's key safety elements and even simple defects can affect safe driving and cause accidents. Therefore, tire quality is critical to the safe and smooth movement of the vehicle. Drivers often overlook problems such as tire aging, oxidation and cracking and do not carry out regular checks [1].

Contamination of tire components on the production line can lead to the production of defective tires, which can cause explosions at high speeds and risks to life. As traditional methods are not efficient, computer-aided systems are important to improve tire quality. In these systems, traditional visual detection [2,3] methods as well as deep learning [4,5] techniques can also be used. Tires may have cracks that can affect the safety of the vehicle while driving. Tire problems pose significant risks to drivers, especially at high speeds. However, while drivers are often aware of risks such as tire pressure and tread depth, they are not aware of problems related to tire life. Therefore, computer and AI-assisted applications using digital images are useful tools when manual checks are not safe [1].

Deep learning methods are used as a successful tool in many fields thanks to their ability to extract features [6–15]. However, they usually require large amounts of data and long training times. To overcome this problem, pre-trained Transfer Learning (TL) methods are used to speed up the training process of deep networks and provide effective learning with less data. In this study, the methods Xception [16], VGG16 [17], NASNet [18], ResNet50 [19], DenseNet [20], InceptionV3 [21] and MobileNet [22] advanced deep networks, such as the one used in this study.

In this study, a hybrid model is proposed to detect cracks on the tire surface and distinguish normal/cracked tires. The proposed model is a modified version of Özaydın and Tekin's DeepFeat-E model [9] In this model, deep features from pre-trained TL models are reduced and then cascaded and applied to stacking [23] and voting [24] ensemble classifiers. This approach enables more stable and accurate detection of oxidation-related problems in tires. Thus, it can enable drivers to detect tire problems in advance and minimize risks by taking preventive measures. Furthermore, the proposed system has the potential to save time and labor and increase efficiency for organizations with a large number of vehicles, such as fleets, logistics companies and public transport providers.

2. LITERATURE REVIEW

Defective tire detection methods are generally divided into two main categories based on the type of defect detection algorithm: traditional visual detection and deep learning-based methods. Traditional visual detection techniques can be classified as statistical, frequency and model-based methods. In the literature, there are several methods that use X-Ray images to detect tire defects during production. In a study by Zhao and Qin [3], tire texture images obtained by non-destructive X-Ray were used for the detection of internal structural tire defects on the production line. The researchers extracted the features of these images with the local inverse difference moment (LIDM) method, calculated the Hausdorff distance using the LIDM features to obtain a defect feature map (DFM) and proposed a pixel-level defect detection algorithm with high accuracy. Similarly, in another study by Guo et al. [2], an effective defect detection method that exploits the feature similarity of tire X-Ray images to detect tire surface abnormalities that may occur during production for the production of quality tires is proposed. In this method, the feature vectors of the tire images are obtained using a local kernel regression (LKR) descriptor to evaluate the feature dissimilarity of pixels and an abnormality map is generated. A simple thresholding process on the abnormality map was used to successfully locate the defects. It is reported that this method gives successful results for both cheek images and back images.

Another area of research is the studies in which convolutional neural network (CNN) models, which are deep learning methods, are included in the literature, as well as methods that directly use tire images. Wang et al. [5] successfully detected tire faults that occurred during production using X-Ray images of tires produced in an industrial area with a fully convolutional neural network (FCN) structure. The network architecture includes the first stage, where a traditional deep network method is applied to extract tire image features. In the second stage, a sampling layer is added to obtain outputs of the same size as the original images. In the last stage, the scaled feature map is added to the outputs and the correct defects are obtained as a result. In this study, the basic architecture of VGG16 was used using a total of 914 images and the results obtained were validated by comparing with AlexNet, VGG11, VGG13 and VGG16.

Another study by Zheng et al. [4] aimed to perform tire defect detection from X-Ray images to detect invisible defects in the internal structure of tires. In this research, a Concise Semantic Segmentation Network (Concise-SSN) model based on VGG16 is proposed for automatic tire visual inspection. This model performs purely pixel-based defect detection. The model combines the capabilities of an optimized semantic segmentation network and an integrated convolutional neural network. Experimental results show that the proposed model outperforms FCN, Mask R-CNN, Faster R-CNN, SegNet and U-net networks in segmentation and classification. The accuracy of the Concise-SSN model in detecting defective tires is reported to be 96.5%.

Another study by Zhang et al. [25] proposes a segmentation method based on wavelet transform using X-Ray images to detect tire surface defects. In this method, the edge information utilizes wavelet transform features that can be represented by larger coefficients in the high frequency band. The larger curve coefficients corresponding to the edge information are selected by thresholding. Comparisons show that the proposed method outperforms traditional edge detection methods (such as Canny, Sobel and LoG).

The TireNet method developed by Li et al. [26] uses X-Ray images of tires to reduce the high rate of tire returns due to manufacturing defects. In this method, a Siamese network is used for image feature extraction and then the model is built by adding the structure of the Siamese network to the Fast R-CNN classifier. Using 120,000 labeled tire images, the TireNet model outperformed expectations by reducing the error rate to 0.17% compared to Faster R-CNN, SSD and YOLO models.

Finally, Lin [1] proposed a ShuffleNet model with a deep learning architecture for the detection of tire defects caused by oxidation. The proposed ShuffleNet model is able to effectively detect defects on normal tire images and outperforms the GoogLeNet, traditional ShuffleNet, VGGNet and ResNet models.

3. MATERIAL and METHOD

In this paper, we propose a hybrid system, CTLDF+EnC, which aims to detect cracked and normal tires. This hybrid system combines deep features extracted from the outputs of a set of TL models, where traditional machine learning (ML) methods are evaluated in an ensemble classifier.

3.1. *General Structure of the System*

The hybrid architecture proposed in this study aims to detect cracked and normal tires. Figure 3.1 shows the general structure of the proposed system including four basic stages. The dataset used in the study contains 1028 images. These images include digital camera images of oxidized (cracked) and non- oxidized (normal) tires.

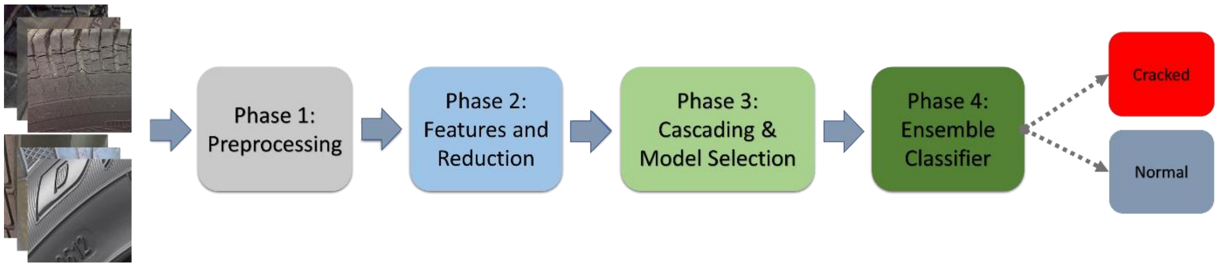

Figure 3.1. Stages of the proposed hybrid system.

Phase 1 (Preprocessing): This stage includes the data preprocessing step. All visual data are scaled to a size (224x224x3) suitable for the TL models. Then, image manipulation operations such as horizontal/vertical translation, size scaling, horizontal/vertical rotation, angular rotation, and brightness adjustment are applied for data enhancement. These operations are applied to make the model more robust to various conditions. Then, the dataset is divided into training and test subsets and analyzed, and the next step is taken.

Phase 2 (Features and Reduction): In this stage, the deep features obtained from each pre-trained TL model used are reduced. Each TL model is applied independently of the others to extract and then reduce the deep features of the training and test datasets.

Phase 3 (Cascading and Model Selection): In this stage, the deep features of the reduced training and test datasets obtained from each ML method are combined (cascading). Then, the five most successful ML methods are selected using the training set. The selection of the five best models is performed using 10 cross-validations and the AUC scale. The top five ML models are selected to be used in the next stage, the ensemble classifier stage.

Phase 4 (Ensemble Classifier): In this final stage, normal and cracked tires are predicted by stacking and voting based ensemble classifiers using the top five ML models identified in the previous stages and the resulting cascaded deep features.

All applications using the hybrid system proposed in this study and direct TL models are implemented in Python programming language using Tensorflow, Keras and scikit-learn libraries. The models and analyses were performed in a personal computer environment with an AMD Ryzen 7 (5800H) processor with 16 CPUs at 3.2GHz, 4GB GDDR6 memory / Nvidia GeForce RTX 3050 GPU at 1.5GHz and 16 GB RAM.

3.2. *Data Set*

In this study, we use a publicly available tire image dataset that includes cracked (oxidized) and normal (non-oxidized) tires. The dataset consists of a total of 1028 images, including tire sidewall and tread images [27]. The dataset used in this study has not undergone any augmentation or preprocessing, and the training and test sets are shared separately. The dataset is publicly available and can be downloaded from Harvard's dataverse webpage [27]. The current dataset contains 491 normal (non-oxidized) and 537 cracked (oxidized) tire images. Table 3.1 details the distribution of cracked and normal tire images between the training and test datasets.

Table 3.1. Data set image distributions.

	Normal	Cracked	Total
Train	376	327	703
Test	115	210	325
Total	491	537	1028

As seen in the table, since there are 703 samples in the training set and 325 samples in the test data set, the training-test data set ratios are approximately 68% and 32% respectively. The resolutions of the images in the dataset are not standardized and are composed of images with different resolutions. For this reason, the images were adjusted to standard sizes in the preprocessing stage and data augmentation methods such as horizontal/vertical shifting, horizontal rotation, scaling, rotation, etc. were applied to prevent overlearning. Examples of normal and cracked images are shown in Figure 3.2.

Figure 3.2. Sample tire images, a) Cracked and b) Normal.

3.3. Proposed Hybrid System

In this study, two main applications were carried out. In the first application group, pre-trained TL models were directly used to predict tire conditions. In the second application group, the proposed hybrid model named CTLDF+EnC (Cascaded Transfer Learning Deep Features + Ensemble Classifiers) is used. This model has a structure in which deep features obtained from TL models are cascaded and applied to ensemble classifiers.

Figure 3.3 shows the block diagram of the CTLDF+EnC hybrid system proposed in this study. In the figure, the Deep Features Generator (DFG) block extracts the deep features of a given TL model and the Reduced Features Set (RFS) is generated by the feature reduction process. This process is performed separately for the 9 TL models used in this study. Since the highest performance values were obtained with the Extra Tree Classifier (ETC), ETC was used as the feature reduction method and IDSs were created. The number of features before and after reduction for each TL is presented in Table 3.2.

Table 3.2. Number of features before and after reduction for each TL model.

TL Models	**Before Reduction**	**After Reduction**
Xception	2,048	527
NASNet	4,032	1,119
MobileNet	1,024	232
DenseNet169	1,664	360
DenseNet201	1,920	381
VGG16	512	139
InceptionV3	2,048	573
ResNet50V2	2,048	497
ResNet101V2	2,048	526
Total	35,714	4,354

In the Cascading and Ensemble Classification (CEC) block, the RFSs of each TL model are cascaded. The Model Selector evaluates 14 different ML methods after 10 cross-validations with these cascaded features, and the first k chooses the ML model (k = 5). Although the selected ML methods vary for each TL, RF (Random Forest), LDA (Linear Discriminant Analysis), LGBM (Light Gradient Boosting Machine), LR (Logistic Regression), ETC (Extra Trees Classifier), GB (Gradient Boosting) and KNN (K Neighbors Classifier) ML models were generally used. When the features obtained from TL models were cascaded, LDA, ETC, LR, GB and LGBM were selected as the best $k = 5$ ML models (ranked according to the top 5 AUC metric). Table 3.3 presents the performance metrics of the ML Models ranked according to the AUC value.

Table 3.3. Best ML models ranked by AUC value.

Model	Accuracy	AUC	Recall	Prec.	F1
Linear Discriminant Analysis	0.9474	**0.9863**	0.976	0.9311	0.9522
Extra Trees Classifier	0.9361	**0.9836**	0.9841	0.9074	0.9433
Logistic Regression	0.9517	**0.9828**	0.9814	0.9337	0.9564
Gradient Boosting Classifier	0.9531	**0.981**	0.9894	0.9293	0.9579
Light Gradient Boosting Machine	0.9503	**0.9805**	0.9841	0.9288	0.9552
Random Forest Classifier	0.9375	0.9798	0.9841	0.9092	0.9444
K Neighbors Classifier	0.9488	0.9705	0.9706	0.9369	0.953
Ada Boost Classifier	0.9317	0.9684	0.9496	0.9261	0.9372
Naive Bayes	0.9261	0.9372	0.9733	0.8998	0.934
Decision Tree Classifier	0.8734	0.8727	0.8859	0.8791	0.882
Quadratic Discriminant Analysis	0.5065	0.5087	0.4814	0.5466	0.5086
Dummy Classifier	0.5348	0.5	1.0	0.5348	0.6969
SVM - Linear Kernel	0.9531	0.0	0.9788	0.9377	0.9574
Ridge Classifier	0.9559	0.0	0.9787	0.9428	0.9599

The best k ML models selected in the Ensemble Learning Model (EnsLM) block are used in the Ensemble Classifier sub-block. This sub-block predicts tire state (Cracked and Normal) using three different ensemble classifiers, one based on Stacking and the other two based on voting, using selected ML models.

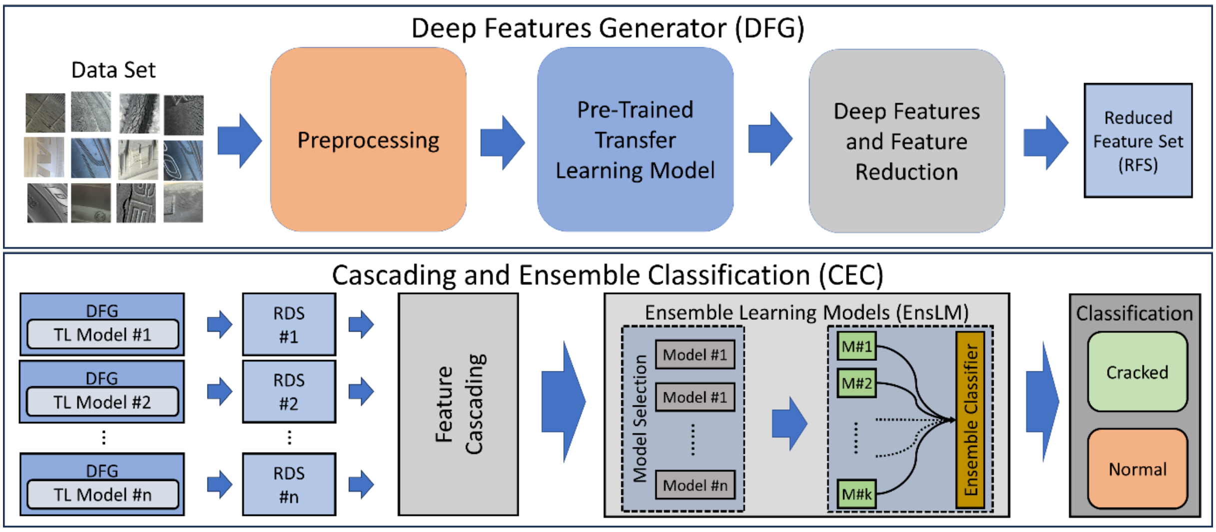

Figure 3.3. Schematic representation of the proposed hybrid system

3.4. Pre-Trained Transfer Learning Models

Transfer learning methods are a technique that enables the use of pre-trained models in a new problem or application and provides an effective solution in ML processes. This methodology is not considered as a different type of ML algorithm, but rather as a strategy or method used to train models. It involves applying the parameters and weights obtained from previous training to a new problem. The reused pre-trained model must have a high level of generality in order to be used in different problems. TL models can be applied to new or different problems without training, saving time and resources for training. Through reuse, problems such as resource shortages and long training times are minimized. With these advantages, TL offers more effective and efficient solutions in ML processes.

TL also offers a solution to the time-consuming process of accurately labeling large datasets [28]. This is a significant advantage, especially when considering the large datasets required to train a ML algorithm. Transfer methods are often preferred when large resources are required for the training phase in a system. Since the pre-trained structures of these models can be used directly, they can be used directly for similar problems related to the model. These aspects of transfer models make them general- purpose. TL allows for the development of more generalizable models rather than models being strictly bound to a training dataset. In this way, the models developed can be used under varying conditions and with different data sets.

3.5. Feature Selection with Extra Tree Classifier

Feature selection is an important step in ML models to improve prediction accuracy and reduce computational cost. For this purpose, the Extra Tree Classifier (ETC) method is used in this study to identify the most appropriate deep features obtained from ML models [29]. ETC is a decision-based method similar to the Random Forest classifier and provides a common framework between feature selection and classification. The method has the ability to generate many sub-trees and randomly select subsets. For feature selection, features are evaluated using the Gini metric and the most important features are identified. In this way, redundant or low-contributing features are eliminated and the complexity of the model is reduced, reducing computational cost and increasing success.

$$Gini = 1 - \sum_{i=0}^{c-1} p_i(t)^2 \tag{3.1}$$

where c is the number of unique classes at this node and $p_i(t)$ is the frequency of class i at node t.

3.6. Ensemble Classification Methods

Ensemble classification methods are approaches that use multiple decision makers instead of a single decision maker in order to reduce the number of incorrect predictions and to increase the achievement. In this study, the best five models are selected among traditional ML models and the best five models are selected among 14 models based on the AUC metric after 10 cross-validation procedures. In the applications, the features obtained from the pre-trained TL models were reduced and cascaded, and then applied to ensemble classifiers using these selected models. The common decision of the ensemble classifiers was determined by Stacking and Voting based ensemble learning methods. Voting strategies include two main approaches, Hard Voting and Soft Voting; Hard Voting is based on the majority decision, while Soft Voting uses the average of the prediction probabilities of the available classifiers. These strategies aim to achieve more efficient and reliable classification results.

3.6.1. Voting Based Ensemble Classifiers

Voting-based ensemble classifiers aim to obtain a common ensemble prediction by combining the classification predictions of various ML techniques. The most commonly used methods are called Hard Voting and Soft Voting. In Hard Voting, the prediction of the majority of classifiers in the ensemble is taken and this prediction is considered as the collective prediction of the ensemble. In Soft Voting, the ensemble prediction is formed by averaging the probabilistic weights of the classifiers' predictions. The contribution of each classifier is calculated with a specific weight and the final prediction is based on this weighted sum. These methods aim to achieve a more reliable and efficient classification result by combining the power of different classifiers in the ensemble. In this way, more accurate predictions can be made, avoiding the potential mispredictions that can arise from a single model. For example, these ensemble classifiers, when predicting the class of a sample x, allow to choose between k class $\{s_1, s_2, s_3, \dots, s_k\}$ predictions of n different classifiers $\{h_1, h_2, h_3, \dots, h_n\}$ according to certain criteria [24].

The Hard Voting method accepts the prediction of the majority of classifiers in the ensemble as the final prediction of the ensemble. In this method, the predictions of the classifiers in the ensemble are equally weighted and only the decision of the majority is taken into account. In this case, in the Hard Voting approach, the class of any instance x in the dataset is predicted as follows [24],

$$H(x) = s_{\sum_{i=1}^{n} h_i^j(x)} \tag{3.2}$$

The Soft Voting method makes class prediction by averaging the probabilistic weights of the predictions of the classifiers in the ensemble. This method aims to achieve a more accurate classification result by weighting the predictions of different classifiers in the ensemble according to their reliability. In this case, in the Soft Voting approach, the class of any instance x in the dataset is estimated as follows [24]:

$$H(x) = s_{\sum_{i=1}^{n} w_i h_i^j(x)} \tag{3.3}$$

The weight of the prediction of each classifier h_i among the other predictions in the ensemble is denoted by w_i.

3.6.2. Stacking Based Community Classifier

The Stacking Based Ensemble Classifier is a method proposed by Wolpert [23] and basically consists of two stages. In the first stage, the predictions of the different methods used in ensemble classifiers are obtained. Then, these predictions are processed by a meta-classifier to produce the final prediction of the ensemble. The aim of this approach is to improve the accuracy by balancing the incorrect predictions of a single classifier with the predictions of other classifiers. In this way, more reliable predictions are obtained through ensemble classifiers [23].

3.7. Performance Evaluation Metrics

In this study, a confusion matrix is used to analyze the performance of the proposed hybrid models and the TL methods used for comparison. The confusion matrix is presented in the form of a table as in Table 3.4, which shows the number of correct and incorrect classifications between the actual class labels of the examples in the dataset and the prediction labels of the models used.

Table 3.4. Confusion matrix

Actual/ Estimated	**Cracked**	**Normal**
Cracked	True Positive (TP)	False Positive (FP)
Normal	False Negative (FN)	True Negative (TN)

Where, TP (True Positive) refers to how many images with cracked true class labels are correctly predicted by the model used. TN (True Negative) refers to how many images with normal true class labels are correctly predicted by the model used. FP (False Positive) refers to how many images with real class labels cracked are incorrectly predicted as normal by the model used, while FN (False Negative) refers to how many images with real class labels normal are incorrectly predicted as cracked by the model used.

TP (True Positive): Refers to how many images with true class labels "Cracked" were correctly predicted by the model.

TN (True Negative): Refers to how many of the images with true class labels "Normal" were correctly predicted by the model.

FP (False Positive): Refers to how many of the images with true class labels "Normal" are predicted as "Cracked" by the model.

FN (False Negative): Refers to how many images with real class labels "Cracked" are predicted as "Normal" by the model.

The following four metrics were used to analyze and evaluate the performance of the models used in the study [30].

Accuracy: Refers to the ratio of correctly estimated samples to the total number of samples and is calculated by the formula below,

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{3.4}$$

Precision: Refers to the probability that samples predicted as Cracked are actually Cracked and is calculated by the following formula,

$$Precision = \frac{TP}{TP + FP} \tag{3.5}$$

Recall: Refers to the proportion of samples whose true class is Cracked that are correctly predicted to be Cracked and is calculated by the formula below,

$$Recall = \frac{TP}{TP + FN} \tag{3.6}$$

f1-score: Provides a balanced measure of performance by taking the harmonic mean of precision and recall and is calculated by the following formula,

$$f1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.7}$$

4. EXPERIMENTAL ANALYSIS

In order to compare the performance of the hybrid model proposed in this study, firstly, a set of applications were performed in which each of the Transfer Learning (TL) models was used as a direct classifier separately. In another set of applications, within the proposed method, after the features of all TC models are obtained and reduced, they are combined together and Stacking, Soft and Hard Voting ensemble classifiers are used separately. In all the applications, the classifiers used were aimed to classify oxidized (Cracked) and non-oxidized (Normal) tires. Table 3.5 presents the values of the performance metrics obtained on the training and test datasets for the first applications where the TL models were used directly. In these applications, the highest training and test accuracies were obtained with the ResNet50V2 TL model with 91.89% and 70.77%, respectively. The lowest accuracy values were observed for the VGG16 TL model with 72.69% and 56.31%, respectively.

Table 3.5. Training and test success criterion values of the TL models.

	Train (%)				Test (%)			
Model Name	**Acc.**	**Prec.**	**Rec.**	**f1-scr.**	**Acc.**	**Prec.**	**Rec.**	**f1-scr.**
Xception	89.05	90.41	88.43	88.79	61.23	71.65	69.21	61.04
NASNet	91.18	91.4	90.94	91.09	66.15	71.79	72.04	66.15
MobileNet	87.2	88.41	86.58	86.91	57.54	67.1	65.18	57.32
DenseNet169	86.91	89.11	86.09	86.49	59.38	71.45	67.98	59.01
DenseNet201	90.9	91.2	90.61	90.8	68.92	74.24	74.77	68.91
VGG16	72.69	74.03	73.41	72.61	56.31	54.79	55.18	54.49
InceptionV3	88.19	88.39	87.93	88.08	65.54	71.48	71.56	65.54
ResNet50V2	**91.89**	92.11	91.66	91.82	**70.77**	74.22	75.61	70.68
ResNet101V2	90.9	91.8	90.41	90.73	63.08	71.07	70.05	63.03

Figure 3.4 shows the training and test confusion matrices of the most successful TL model, ResNet50V2. In the Confusion matrix for the training dataset in Figure 3.4.(a), out of 327 cracked tire images, 289 were correctly classified and 38 were incorrectly predicted. Of the normal tire images, 357 were correctly classified and 19 were incorrectly predicted. In total, there are 376 normal tire images. In the complexity matrix for the test dataset in Figure 3.4.(b), 124 out of 210 cracked tire images are correctly classified and 86 are incorrectly predicted. Of the normal tire images, 106 were correctly classified and 9 were incorrectly predicted. In total, there are 115 normal tire images.

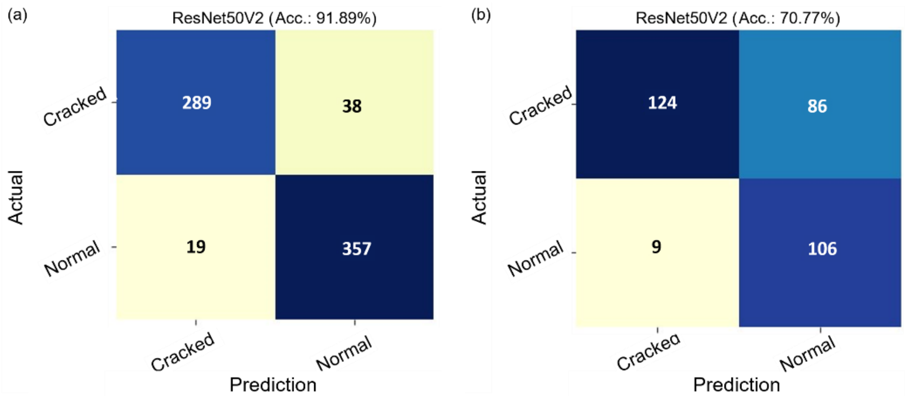

Figure 3.4 ResNet50V2 TL training and test confusion matrices.

Table 3.6 shows the performance of the CTLDF+EnC based ensemble classifiers on the test dataset. The table shows the performance of all TL models when deep features are applied to the ensemble classifiers. The highest success rate is obtained with the hybrid model with Stacking ensemble classifier (CTLDF+EnC(Stacking)), which has an accuracy of 76.92% and precision, sensitivity and f1-score of 79.10%, 81.36% and 76.75% respectively. The hybrid models of Soft CTLDF+EnC(Soft) and Hard CTLDF+EnC(Hard) voting methods have an accuracy of 74.15% and 72.92% respectively. The precision, sensitivity and f1-score values of these models are given in Table 3.6. It was observed that all ensemble classifiers achieved 100% success rate on the training dataset. The reason for this difference between the accuracy values for the training and test datasets is considered to be overfitting. In the case of overfitting, while the model learns each sample in the training set very well, it loses its generalization capability and classifies the samples in the test data set, which it has never seen, with low accuracy.

Table 3.6. CTLDF+EnC based ensemble classifier performance values.

	Train (%)				Test (%)			
Model	**Acc.**	**Prec.**	**Rec.**	**f1-scr.**	**Acc.**	**Prec.**	**Rec.**	**f1-scr.**
CTLDF+EnC(Stacking)	100	100	100	100	76,92	79,1	81,36	76,75
CTLDF+EnC(Soft)	100	100	100	100	74,15	77,26	79,02	74,05
CTLDF+EnC(Solid)	100	100	100	100	72,92	76,6	78,06	72,85

Figure 3.5 shows the test data set confusion matrices for each ensemble classifier. Since the training dataset achievements are 100%, they are not presented separately. In Figure 3.5(a), the CTLDF+EnC (Stacking) model predicts 135 correct and 75 incorrect images of cracked tires. For normal tire images, 111 were correctly predicted and 4 were incorrectly predicted. In Figure 3.5(b) and Figure 3.5(c), the correct/incorrect prediction values of the cracked tire images for the CTLDF+EnC(Soft) and CTLDF+EnC(Solid) models are 134/76 and 128/82, respectively. The true/false prediction values for normal tire images are 111/4 and 110/5, respectively. These matrices show the ability of each hybrid model to correctly and incorrectly predict cracked and normal tire images.

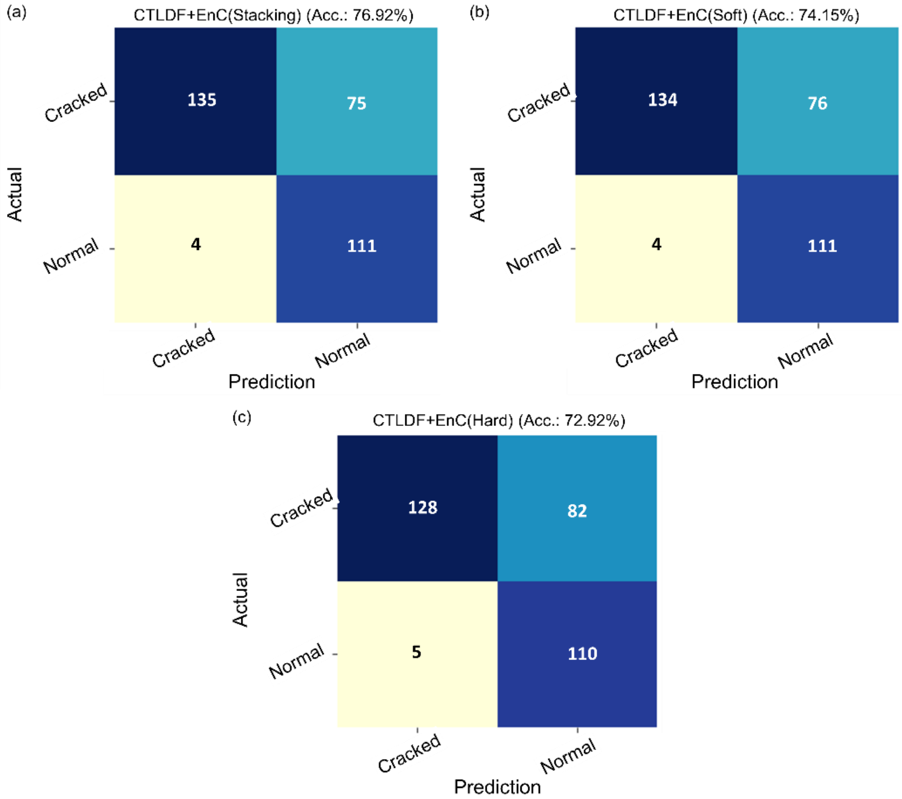

Figure 3.5. Ensemble classifier test confusion matrices

Within the scope of the study, Table 3.7 was created to determine the classification performance of hybrid models using the deep features of the proposed TL models and to more clearly demonstrate the advantages of deep feature-based models. In this table, the first column, TL Model (TLM), contains the success metrics obtained directly with the TL models, while the other columns show the test accuracies and improvement amounts obtained with ensemble classifiers such as Stacking Hybrid Model (SHM), Soft Voting Hybrid Model (SVHM) and Hard Voting Hybrid Model (HVHM), which are built using all the features obtained from the TL models. From Table 3.7, it can be seen that the proposed SHM, SVHM and HVHM ensemble classifiers all exceed the performance measures of the TL models and provide more accurate predictions. This table clearly shows that the proposed deep feature-based hybrid models provide higher classification performance compared to the direct TL models.

Table 3.7. Comparison of the quantities of improvement in the test performance of the hybrid models.

		Imp. Dif. (%)		
Models	**TLM**	**SHM**	**SVHM**	**HVHM**
Xception	61.23	15.69	12.92	11.69
NASNet	66.15	10.77	8	6.77

MobileNet	57.54	19.38	16.61	15.38
DenseNet169	59.38	17.54	14.77	13.54
DenseNet201	68.92	8	5.23	4
VGG16	56.31	**20.61**	**17.84**	**16.61**
InceptionV3	65.54	11.38	8.61	7.38
ResNet50V2	70.77	6.15	3.38	2.15
ResNet101V2	63.08	13.84	11.07	9.84

In Table 3.7, it is seen that the highest accuracy values obtained with the SHM, SVHM and HVHM are 76.92%, 74.15% and 72.92%, respectively. When these values are compared with the accuracy values of the TL models, it is determined that the highest improvement differences are 20.61%, 17.84% and 16.61% for the VGG16 TL model, respectively. On the other hand, the lowest improvement differences were 6.15%, 3.38% and 2.15% for the ResNet50V2 TL model, respectively. It is expected that the improvement differences would be low considering that ResNet50V2 was the model with the highest success in the first group of applications. It is clear from Table 3.7 that the proposed hybrid models of SHM, SVHM and HVHM significantly improve the accuracy of ResNet50V2 in particular.

5. DISCUSSION

Vehicle drivers are often aware of the importance of tire tread depth and tire air pressure, but overlook the risks of tire oxidation. Tire oxidation and related cracks pose potential hazards that seriously affect driving safety. In this context, this study aims to detect cracks in tires based on the use of pre-trained TL methods and ensemble classifiers.

According to the World Health Organization (WHO), a large number of fatal traffic accidents occur every year and the majority of these accidents are caused by tire defects [31]. While manual detection of such defects can be difficult and inaccurate, AI-enabled systems have significant potential in this area. Therefore, we propose an AI-assisted model for easy and fast detection of tire defects by vehicle users. In the proposed model, three different hybrid models based on the five best classical ML algorithms are proposed by combining the deep features of nine pre-trained ML models obtained from tire images with ensemble classifiers: CTLDF+EnC(Stacking), CTLDF+EnC(Soft) and CTLDF+EnC(Solid). This approach provides an effective solution for tire defect detection.

The use of non-destructive testing techniques for the detection of tire defects is widespread. These techniques include laser shearing [32], ultrasonic methods [33] and electromagnetic pulse [34]. However, these methods are often expensive and difficult to implement and are not widely used. Moreover, most of the studies based on traditional visual detection [2,3] and deep learning [4,35] methods use X-Ray imaging and these studies usually focus on the production line. In this study, we aim to detect tire defects that are worn or oxidized due to usage. Furthermore, a more efficient structure is proposed by replacing the expensive and complex X-Ray images with more cost-effective and easily available digital camera images. This approach provides a more accessible and practical solution for tire defect detection.

Although the use of digital images is almost non-existent in the literature, a similar approach was adopted in a study by Lin [1]. However, since the dataset of this study was not shared, the success of the hybrid models proposed here could not be tested on this dataset. This shows that each study has its own unique datasets and model success is dataset dependent. However, on the basis of the applications with TL models, it is seen that the hybrid models proposed in this study significantly increase the success. This shows that the study offers a new and effective approach.

In this study, the proposed CTLDF+EnC (Stacking) hybrid model has the highest test accuracy of 76.92%. The other hybrid models, CTLDF+EnC (Soft) and CTLDF+EnC (Hard) architectures have 74.15% and 72.92% accuracy respectively. According to these results, it is concluded that the proposed hybrid model performs at an acceptable level in general and can be an effective tool for tire defect detection.

When the performance of the proposed hybrid models is compared to the performance of the directly used TL models, it is seen that they are more successful. For example, in Table 3.5, the highest test accuracy rate is 70.77% with ResNet50V2 when the TL models are used directly. On the other hand, the test accuracy rate obtained with the proposed CTLDF+EnC (Stacking) method is 6.15% higher than ResNet50V2. As can be seen in Table 3.5 and Table 3.7, all of the ensemble TL models have lower accuracy values than the proposed hybrid classifiers. These results show that the proposed hybrid models are more effective than the TL models in detecting tire problems and have a reasonable level of success. In this context, the proposed hybrid models are considered to be a promising method for future studies in tire defect detection.

The original contributions of this work can be listed as follows: First, we propose a new architecture, CTLDF+EnC, which enables inspection from tire images. This architecture represents an important step forward in tire defect detection. Second, the architecture has the ability to combine deep features of TL models and classical ML methods in an ensemble classifier. This provides an efficient way to achieve more comprehensive and accurate results. Finally, the study emphasizes the use of images acquired with a digital camera, which is a cheaper and easy-to-use method to replace expensive and complex imaging techniques. This provides a more accessible and cost-effective solution for tire defect detection. Therefore, the proposed system can be considered as a useful tool to improve safety standards in the tire industry and solve tire-related problems more effectively.

6. CONCLUSION

This paper presents a novel approach, CTLDF+EnC, for tire defect detection. This architecture combines deep features derived from TL methods and classical ML methods to propose an effective solution for tire defect detection. The proposed architecture can work with cheap and easy-to-use digital images that can be acquired with regular digital cameras, which reduces the cost and increases the applicability. In addition, this work aims to detect problems that occur during the lifetime of tires, such as oxidation and aging, unlike the studies in the literature, which are usually aimed at detecting defects that occur on the production line. Nine different TL models were used to extract features, which were then combined and fed to ensemble classifiers containing classical ML models. Stacking, Soft and Hard voting methods are used as ensemble classifiers. The implementations show that the proposed architecture achieves satisfactory performance compared to other alternatives. The results show that the CTLDF+EnC architecture can successfully detect oxidized or worn tires and is advantageous in terms of cost and applicability. This study is considered to propose an effective system to improve safety standards in the tire industry and solve tire-related problems more effectively.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement

The data that support the findings of this study are openly available in “Oxidized and non-oxidized tire sidewall and tread images” at https://doi.org/10.7910/DVN/Z3ZYLI.

REFERENCES

[1] S.-L. Lin, Research on tire crack detection using image deep learning method, Sci Rep 13 (2023) 8027.

[2] Q. Guo, C. Zhang, H. Liu, X. Zhang, Defect detection in tire X-ray images using weighted texture dissimilarity, J Sens 2016 (2016).

[3] G. Zhao, S. Qin, High-precision detection of defects of tire texture through X-ray imaging based on local inverse difference moment features, Sensors 18 (2018) 2524.

[4] Z. Zheng, S. Zhang, B. Yu, Q. Li, Y. Zhang, Defect inspection in tire radiographic image using concise semantic segmentation, IEEE Access 8 (2020) 112674–112687.

[5] R. Wang, Q. Guo, S. Lu, C. Zhang, Tire defect detection using fully convolutional network, IEEE Access 7 (2019) 43502–43510.

[6] J. Huang, B. Kingsbury, Audio-visual deep learning for noise robust speech recognition, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, 2013: pp. 7596–7599.

[7] Y. Kaya, F. Kuncan, R. Tekin, A new approach for congestive heart failure and arrhythmia classification using angle transformation with LSTM, Arab J Sci Eng 47 (2022) 10497–10513.

[8] J. Liu, D. Capurro, A. Nguyen, K. Verspoor, “Note Bloat” impacts deep learning-based NLP models for clinical prediction tasks, J Biomed Inform 133 (2022) 104149.

[9] B. Özaydın, R. Tekin, *A Hybrid Model Based on Deep Features and Ensemble Learning for the Diagnosis of COVID-19: DeepFeat-E, Turkish Journal of Science and Technology* 18 (2023) 183–198.

[10] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, T. Sainath, Deep learning for audio signal processing, IEEE J Sel Top Signal Process 13 (2019) 206–219.

[11] B. Rim, N.-J. Sung, S. Min, M. Hong, Deep learning in physiological signal data: A survey, Sensors 20 (2020) 969.

[12] H. Tung, R. Tekin, New Feature Extraction Approaches Based on Spatial Points for Visual-Only Lip-Reading., Traitement Du Signal 39 (2022).

[13] D. Wang, J. Su, H. Yu, Feature extraction and analysis of natural language processing for deep learning English language, IEEE Access 8 (2020) 46335–46345.

[14] S.-H. Wang, S. Xie, X. Chen, D.S. Guttery, C. Tang, J. Sun, Y.-D. Zhang, Alcoholism identification based on an AlexNet transfer learning model, Front Psychiatry 10 (2019) 205.

[15] X. Yang, Y. Zhang, W. Lv, D. Wang, Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier, Renew Energy 163 (2021) 386–397.

[16] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: pp. 1251–1258.

[17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, ArXiv Preprint ArXiv:1409.1556 (2014).

[18] B. Zoph, V. Vasudevan, J. Shlens, Q. V Le, Learning transferable architectures for scalable image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: pp. 8697–8710.

[19] K. He, X. Zhang, S. Ren, Deep Residual Learning for Image Recognition, Microsoft Research 45 (2015) 1951–1954. https://doi.org/10.1002/chin.200650130.

[20] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-Janua (2017) 2261–2269. https://doi.org/10.1109/CVPR.2017.243.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: pp. 2818–2826.

[22] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, (2017). http://arxiv.org/abs/1704.04861.

[23] D.H. Wolpert, Stacked generalization, Neural Networks 5 (1992) 241–259.

[24] Z.-H. Zhou, Ensemble methods: foundations and algorithms, CRC press, 2012.

[25] Y. Zhang, D. Lefebvre, Q. Li, Automatic detection of defects in tire radiographic images, IEEE Transactions on Automation Science and Engineering 14 (2015) 1378–1386.

[26] Y. Li, B. Fan, W. Zhang, Z. Jiang, TireNet: A high recall rate method for practical application of tire defect type classification, Future Generation Computer Systems 125 (2021) 1–9.

[27] J. Siegel, Oxidized and non-oxidized tire sidewall and tread images, Harvard Dataverse (2021). https://doi.org/https://doi.org/10.7910/DVN/Z3ZYLI.

[28] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, G. Zhang, Transfer learning using computational intelligence: A survey, Knowl Based Syst 80 (2015) 14–23.

[29] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach Learn 63 (2006) 3–42.

[30] M.K. Bohmrah, H. Kaur, Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model, Global Transitions Proceedings 2 (2021) 476–483.

[31] S. Erdogan, Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey, J Safety Res 40 (2009) 341–351.

[32] C. Cai, L. He, Improved Mach–Zehnder interferometer-based shearography, Opt Lasers Eng 50 (2012) 1699–1705.

[33] H.L.M. Dos Reis, K.A. Warmann, Acousto-ultrasonic non-destructive evaluation of fatigue damage in steel-belted radial tires., Int J Fatigue 3 (1996) 216.

[34] L.E. Roemer, N. Ida, Location of wire position in tyre belting using Bayesian analysis, NDT & E International 24 (1991) 95–97.

[35] R. Wang, Q. Guo, S. Lu, C. Zhang, Tire defect detection using fully convolutional network, IEEE Access 7 (2019) 43502–43510.