Engineering Research Express

INCLUSIVE PUBLISHING TRUSTED SCIENCE

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

Bubble defect detection for tire shearography images with transfer learning based deep CNN models

To cite this article: Nagmy A A Saleh *et al* 2024 *Eng. Res. Express* **6** 045423

View the article online for updates and enhancements.

You may also like

- Structural Design and Economic Evaluation of Roller Compacted Concrete Pavement with Recycled Aggregates
Yavuz Abut and Salih Taner Yildirim
- Spatial Character Analysis of Streets as Public Spaces: The Case of Izmit Hurriyet and Cumhuriyet Street, Turkey
Mehtap Özbayraktar, Merve Pekdemir and Gumru Mrzaliyeva
- An *in vitro* 3D diabetic human skin model from diabetic primary cells
Candan Yilmaz Ozdogan, Halime Kenar, Kivanc Emre Davun et al.

This content was downloaded from IP address 92.44.150.127 on 21/11/2024 at 23:33

Engineering Research Express

PAPER

Bubble defect detection for tire shearography images with transfer learning based deep CNN models

Nagmy A A Saleh[1], **Farid Al-Areqi**[2], **Mehmet Zeki Konyar**[3], **Kaplan Kaplan**[3], **Semih Öngir**[4] and **H Metin Ertunç**[1]

[1] Mechatronics Engineering Dept, Kocaeli University, Kocaeli, Türkiye
[2] Biomedical Engineering Dept, Kocaeli University, Kocaeli, Türkiye
[3] Software Engineering Dept, Kocaeli University, Kocaeli, Türkiye
[4] Pirelli Motorsports Dept, Pirelli Automobile Tyres, Kocaeli, Türkiye

E-mail: mzeki.konyar@kocaeli.edu.tr

OPEN ACCESS

RECEIVED
31 August 2024

REVISED
27 October 2024

ACCEPTED FOR PUBLICATION
12 November 2024

PUBLISHED
20 November 2024

Keywords: bubble detection, tire shearography images, tire defect detection, transfer learning

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Abstract

Safety of drivers and passengers is an important quality assurance aim of the tire manufacturers. Tires that have air bubbles in them are a frequent cause of accidents. Recently, manufacturers have placed a high value on employing laborers with tire diagnostic experience. Along with the expenditure of time and money, this process also necessitates a sizable workforce of experienced laborers. The detection of tire defects can be automated utilizing a number of different ways, greatly reducing the margin for human error. One such method is digital shearography, which can be used to detect air bubbles in images. In this study, transfer learning based models are proposed to classify air bubble defects in tire shearography images. Proposed method employed in transfer learning with several deep learning CNN-based models, including VGG16, VGG19, ResNet50, Xception, MobileNetV2 and DenseNet201. Owing to the assistance of the experts at the Pirelli Automobile Tires Izmit factory, we were able to collect and label the dataset used in this study. The dataset contains 1392 tire shearography images. 811 of these images belong to the 'good' class, while the remaining 581 images belong to the 'bad' class. The proposed method in this study achieves remarkable bubble detection results over 95% in terms of accuracy and other metrics with the DenseNet201 model. The results indicate that the proposed model has comparable accuracy and recall with the existing studies. This study is valuable for the tire industry, as the existing literature offers limited research on the classification of tire bubble defects, the differentiation between defective and defect-free tires, and the categorization of various defect types. In addition, the proposed model offers several advantages over existing studies, including higher robustness against overfitting, suitability for real-time industrial applications, and improved efficiency in processing, making a practical solution for defect detection in industrial settings.

1. Introduction

Depending on the driving conditions, air bubbles in tires can cause road accidents and even take lives. Therefore, tire manufacturers must consider the diagnosis of air bubbles in order to provide safe transportation for passengers and drivers. Tire manufacturers use non-destructive testing (NDT) methods to detect air bubbles in tires. They use digital imaging techniques such as digital shearography and digital radiography to capture images for the whole tire to classify the faulty tire from the intact one. Currently, factories need experienced and qualified operators to perform the classification process by eyes. This process costs time, and the classification accuracy by the operator is subjected to human error. However, with technological progress and the development of automation techniques, many industrial enterprises have replaced labor with automated systems which have largely replaced labor-intensive tasks in traditional manufacturing units [1]. In addition, it has been noted that technologies that enable intelligent diagnosis of tire defects have gained great importance in

© 2024 The Author(s). Published by IOP Publishing Ltd

achieving mass production of tires [2]. Many tire defect detection algorithms have been proposed in literature [3]. However, the vast majority of them are not for automatic tire classification. Our research focuses on automating the tire classification process by detecting air bubbles in the tire shearography images using transfer learning models.

Tire manufacturers carry out tire quality inspections to meet safety standards and prevent loss of life. For that, they use digital imaging techniques to take images of the tires through the production process. There are two common imaging techniques used by tire manufacturers: digital radiography and shearography. There are different types of manufacturing defects in tires such as carcass bubbles, delamination, impurities, carcass steel ply defects, and belt defects. However, our paper does not focus on all defect types due to the specification of the shearography quality inspection device. This study uses images obtained from a shearography device. The shearography devices are only used to check the absence or presence of bubble defects of tires for this study. The other defect types are detected with another radiography device that produces x-ray images.

The digital radiography technique is a non-destructive testing technique that uses x-ray machines to capture images for the whole tire. Since x-rays can penetrate the raw materials of the tire, the images can show a difference between the density of the tire's raw material and any foreign body or even air bubbles.

Digital shearography is a nondestructive testing (NDT) method that can be utilized to inspect materials. Shearography, also known as speckle pattern shearing interferometry, was introduced in the 1970s as an electronic alternative to speckle pattern interferometry (ESPI). When compared to other optical NDT technologies, shearography has some advantages. Similar to holography and ESPI, the main difference is that strain information and the deformation gradient are measured directly. Holography and ESPI, on the other hand, only measure the deformation gradient. Defects introduced during the manufacturing process often lead to internal stresses within objects. When a load is applied during shearography, these internal stresses can manifest as strain, which the technique is designed to detect. Additionally, because the movement of a rigid body does not produce strain, shearography has the advantage of being insensitive to even the smallest rigid-body motions. The finest rigid-body movements' refers to minor, the non-deforming shifts of the tire structure that do not produce strain. Shearography effectively ignores these movements, allowing for accurate detection of strain-related defects. Digital shearography has emerged as one of the most sensitive and accurate methods for nondestructive testing (NDT), particularly in the detection of strain concentrations and defects in composite structures, due to its simple setup and robustness against environmental disturbances [2].

Photographic shearography, electronic shearography, and digital shearography [4, 5] are the three stages of shearography development. Digital shearography is currently the most cutting-edge and widely utilized technique available. The images are captured and analyzed by means of charge-coupled devices (CCDs) and complementary metal-oxide semiconductors (CMOS). Phase measurement has also been implemented to improve the accuracy of digital shearography measurements and the quality of resulting phase maps [6]. Image shear, a laser-based optical interferometry technology, is introduced into digital shearography via a shearing device put in front of the camera.

One of the most important uses of sherography is in the rubber industry to evaluate aircraft and vehicle tire quality. The key point that the laser shearography technique focuses on is the detection of air bubbles between the rubber layers [7]. In this study, images obtained with the Intact® 1200 model shearography tire inspection device produced by Steinbichler Optotechnik GmbH. All images gathered with default settings for race cars. The scope of the study is not how the images are created, but how the images obtained from this device can be evaluated more easily with artificial intelligence techniques. Because this study aims to strengthen the weaknesses of quality control operators and eliminate their deficiencies. Figure 1 shows the inside and outside of the shearography imaging system for tire quality control. Steinbichler Optotechnik Intact® shearography tire testing systems were developed for non-destructive testing of new tires for quality control. The technical basis of the systems is the phase shearography system. With the phase, the shearography technique detects tire bubble deformations in the range of microns. The tire test systems have been successfully used for car, truck, and aircraft industries, as well as in the racing tire industry. By phase shearography, a CCD camera records the interferogram. The original tire image at normal pressure is stored, and real-time subtraction is done with phase images of the deformed tire at reduced pressure. Fault-specific defects appear as phase differences and interference fringes on the system monitor. The inner side of the tire tread is usually measured in 8 sectors. The tire is loaded horizontally into the vacuum chamber, where the measuring head rotates inside. The flexible system accommodates tires of various sizes without adjustments, and the tilting axis allows inspection of the bead and sidewall. In the bubble defect detection, shearography control is performed for the tread and sidewall of the tire. The machine used has the ability to do both. However, mostly the factory works on tread images. Although it varies according to the tire size (tires with 13-inch to 22-inch rim range), it should be considered as the total of the rim and sidewall thickness. The measurement distance varies between approximately 20–30 cm. In order to take each image, the cameras scan at an angle close to 60-degree vertically, make a 45-degree rotation movement horizontally, and complete 360-degrees in 8 images. While the image is being taken by the camera,



Figure 1. Inside and outside of the Intact® 1200 shearography imaging system.

the surface is illuminated with a linear moving laser light. The cameras move with the lasers. If the sidewall control is also performed, the vertical angle can be increased even more, the cameras have the ability to move up and down. The device uses an out-of-plane configuration and the surface being viewed is inclined. As is usually the case, no crack control is performed. Only air bubble control is performed. There is a vacuum pressure defined as a load in the device, the value of which generated about 40 mbar. The first image is taken without vacuum, then the second image is taken after the vacuum is applied. The air bubble shape is obtained from the difference between the two images (vacuum increases the volume of the bubble if there is air). The system takes eight images without vacuum and eight with vacuum in one minute with two cameras. If the same process is done with a single camera, the same process takes two minutes. Since there is no thermal in the machine, there is no heating or cooling system. Therefore, there is only a vacuum or non-vacuum structure. What is called power is the 40-mbar vacuum applied. During the tire production process, air bubbles are formed between layers of rubber fabric as they are added on top of each other. Air bubbles between the tire layers can be found in the tread or sidewalls of the tire. Detection of air bubbles is made after the tyre's baking process. For a healthy control, the tire that has just come out of the cooking process needs to be cooled down to room temperature and should not be subjected to impacts before entering the control machine. In this case, the tire must be kept waiting for certain periods and the impact should be expected to pass.

The shearography is an important technique for detecting surface defects in tire quality inspections [8]. However, it is limited in its ability to identify only certain types of defects, making it applicable only in specific scenarios. Accurate tire inspection and defect classification, followed by proper product categorization, are essential for preventing serious damage and accidents. Currently, the most commonly used method for detecting defects, aside from air bubbles, is x-ray imaging [9]. Nevertheless, automated defect classification and product grading remain challenging tasks that heavily rely on the expertise of human inspectors.

In this study, we propose models for the detection of air bubble defects in shearography images of tires by utilizing transfer learning models such as VGG16, VGG19, ResNet50, Xception, MobileNetV2 and DenseNet201. The models could be implemented in automation systems for the detection of air bubble defects in a tire manufacturing facility. Automation in the detection and classification process has the potential to supplement or even replace human expertise in manufacturing facilities, reducing the potential for human error and increasing productivity. Our contribution is summarized as follows:

- The main problem of artificial intelligence-based tire shearography studies is the absence of a dataset. This study built a dataset by collecting and labeling images for a tire manufacturing facility.
- This paper proposed transfer learning-based models to detect and classify the air bubbles defects in the tire shearography images with high accuracy and recall.
- To the best of our knowledge, there is only one existing tire shearography study in literature and many normal tires are classified as faulty in that study. However, the results of the proposed method in this paper have better accuracy than the existing paper. Besides, it has superior performance over most of the tire defect detection studies in the literature.

2. Literature review

The artificial intelligence methods have contributed to tire automation by classifying tire problems to automate the tire manufacturing process and improve classification accuracy. Most of these researchers didn't achieve the anticipated results due to a lack of data. This means that the most promising findings concern the classification

of easily detectable tire defects. Only a few research have aimed at detecting faults with x-ray images. As far as is known, only a few studies have employed laser shearography images to detect tire bubbles. Image processing, machine learning (ML), and deep learning (DL) have all been used to detect tire bubble defects by many researchers.

Chang *et al* [10] investigated air bubble defects in tire shearography images, developing and testing an approach for classifying these defects using high-performance Convolutional Neural Networks (CNNs). Their study utilized a training dataset consisting of 325 shearography tire images with air bubble defects and 2,954 images without defects. The testing dataset included 541 shearography tire images without air bubble defects and 256 images with defects. The experimental results demonstrated that the proposed model outperformed traditional machine learning techniques. Specifically, the model achieved an accuracy of 86.87% and a recall of 100% using images of tire sidewalls, and an accuracy of 89.16% with a recall of 100% using images of tire treads.

Zhang *et al* [11] also focused on air bubble defects in tire shearography images, proposing the use of laser shearography combined with an edge detector based on the curvelet transform. This technique was applied to detect defects such as air bubbles and delamination within the tire carcass. Their experimental results indicated that the proposed method for detecting the edges of interference fringes outperformed traditional edge detection techniques, including Canny, Sobel, and LoG. However, laser shearography has limitations, such as requiring a vacuum environment to obtain interference images and its reduced detection performance for non-metallic foreign objects, like plastics. In shearography, interference images are typically obtained in a vacuum environment to reduce the effects of external disturbances such as air currents and vibrations, which can distort the interference pattern and compromise accuracy. Furthermore, non-metallic foreign objects, including plastics, demonstrate poor detection performance as they often lack the necessary optical and mechanical response under the applied stress conditions, making it challenging to produce a clear interference image.

Zhang *et al* [12] also proposed a local regularity analysis and scale features to look at radiographic x-ray images of tires in an effort to identify any defects that might be existent. Their study of 400 images revealed that there were several kinds of defects existing in a wide range of configurations and densities. These defects included bubbles, foreign bodies, and a number of different cable faults and crack flaws. They were able to determine the right scale and threshold parameter values for framing the defective edges with the assistance of a fault edge measuring model.

The study conducted by Zhu and Ai [3] suggests that a faster RCNN network of DL algorithms must be used to detect defects in tire x-ray images. There are many types of tire defects like carcass bubbles, delamination, impurities, carcass steel ply defects, and belt defects that they were able to detect in this study. However, the results of their proposed model were ineffective in detecting air tire bubbles due to the lack of data they had. An accuracy of 96.8% has been obtained in this study.

A novel approach that was developed by Cui *et al* [13] based on Independent Component Analysis (ICA) and Topographic ICA algorithms were used to extract tire defect features and enable automatic tire defect detection. This approach was implemented in a prototype. They classified the test samples using a nonlinear support vector machine, and they had a success rate of 95.76% with their classification.

In Wang *et al* study [14], it has been suggested that a fully convolutional network (FCN) model might be used to detect defects in x-ray images of tires. They used a dataset containing 914 images, including images of the sidewall and tread. 700 images have been selected at random to be used as training groups, and another 214 images have been selected at random to be used as testing groups. These training and testing groups include metal impurities, bubbles, and overlaps. For the purpose of extracting features from tire images, VGG16 is created. The proposed model had a 78.91% accuracy rate in detecting tire defects.

Yang *et al* [15] used a simple technique for tire bubble defects called line laser. This technique used a laser to detect bubble defects on the surface of the tire. The laser line is straight when no bubble exists and changes to the bubble shape when a bubble exists on the surface. A CCD camera takes an image of the laser line and image processing techniques are used to remove noise and enhance the detection process. They used 200 samples and compared the proposed method with different tire defect detection algorithms. The accuracy rate is 93% for the single bubble detection and 91.5% for two bubble detection. Table 1 summarizes the state-of-the-art studies of tire defect detection.

3. Material & methods

3.1. Dataset

The dataset has been collected from the Pirelli Automobile Tires İzmit Factory and labeled with the assistance of their expert to be used in this study. The dataset consists of 1392 shearography images of 174 tire treads which is an extended version of our previous dataset [7]. Each tire tread is captured with 8 images, in which the tire is rotated 45° to capture each image. Thus, we ended up with 8 shearography images for each tire. Figure 2 shows

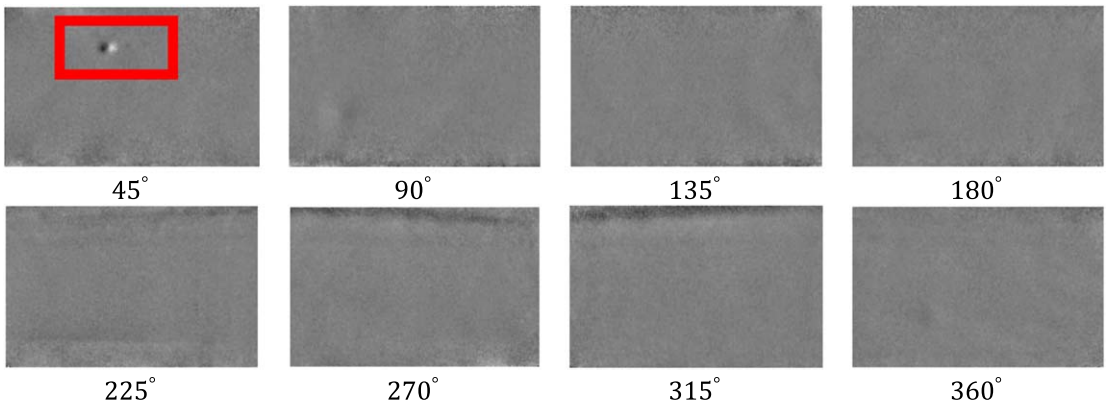


Figure 2. Sample shearography images of a tire for different angles.

Table 1. Summary of the studies in the literature.

Author	Dataset size	Classifiers	Accuracy
Chang *et al* [10]	4076 images	Support Vector Machine	50.13%
		RF	44.61%
		Faster Region-based Convolutional Neural Networks	89.16%
Zhu and Ai [3]	—	Faster RCNN	96.8%
Zhang *et al* [12]	400 images	Support Vector Machine	95.76%
Wang *et al* [14]	914 images	VGG16	78.91%
Yang *et al* [15]	200 images	Image processing techniques	91.5%

an example of the 8 images for one tire. To train the models, each image of the tire is labeled separately. That is because the bubble may exist in one or more parts of the tire but not in the other parts. In figure 2, each set contains eight images, and if any one of these images shows the presence of an air bubble, the entire set is labeled as 'bad' to maintain consistency in defect classification. Additionally, in the 45-degree image, the air bubble is highlighted with a red square for clear identification, ensuring the defect is easily visible. Thus, 811 images are labeled as good for images without air bubbles, and 581 are labeled as bad for images with air bubbles. Some samples of the shearography images for the good class are shown in figure 3, and for the bad class in figure 4.

3.2. Convolutional neural networks (CNNs)

CNNs are comparable to steady Neural Networks in that they are composed of neurons with learnable weights and biases, similar to how steady Neural Networks are constructed. Each neuron may take many inputs, does a dot product, and then executes a nonlinearity. The entire network, from the raw picture pixels on one end to the class scores on the other, is still represented by a single differentiable scoring function [16]. ConvNet topologies are used to create images, which feature certain properties. After that it is possible to use the function that appears appropriate to design using several network variables, resulting in a reduction in an outstanding manner. A *ConvNet* is a neural network that positions neurons in three dimensions. It is represented as a single layer with depth, breadth, and height. In each layer of the *ConvNet*, neuron activations are transformed from the 3D input volume to the 3D output volume of the network. In this case, the input layer contains the image, therefore the height and width correspond to the image's dimensions, however, the penetration can be three times more. These three images are assigned as the blue channels, red channels, and green channels [14].

The selection of models for this study focused on those that are widely used, well-suited for transfer learning, and easily accessible through reputable public libraries like Keras and TensorFlow. The chosen models—VGG16, VGG19, ResNet50, Xception, MobileNetV2 and DenseNet201—are known for their documented success in image classification tasks and their extensive application in transfer learning. Their support within popular deep learning libraries ensures accessibility and ease of implementation. Additionally, these models offer varying architectures and complexities, providing a robust basis for comparing performance across models with different depths and structures.

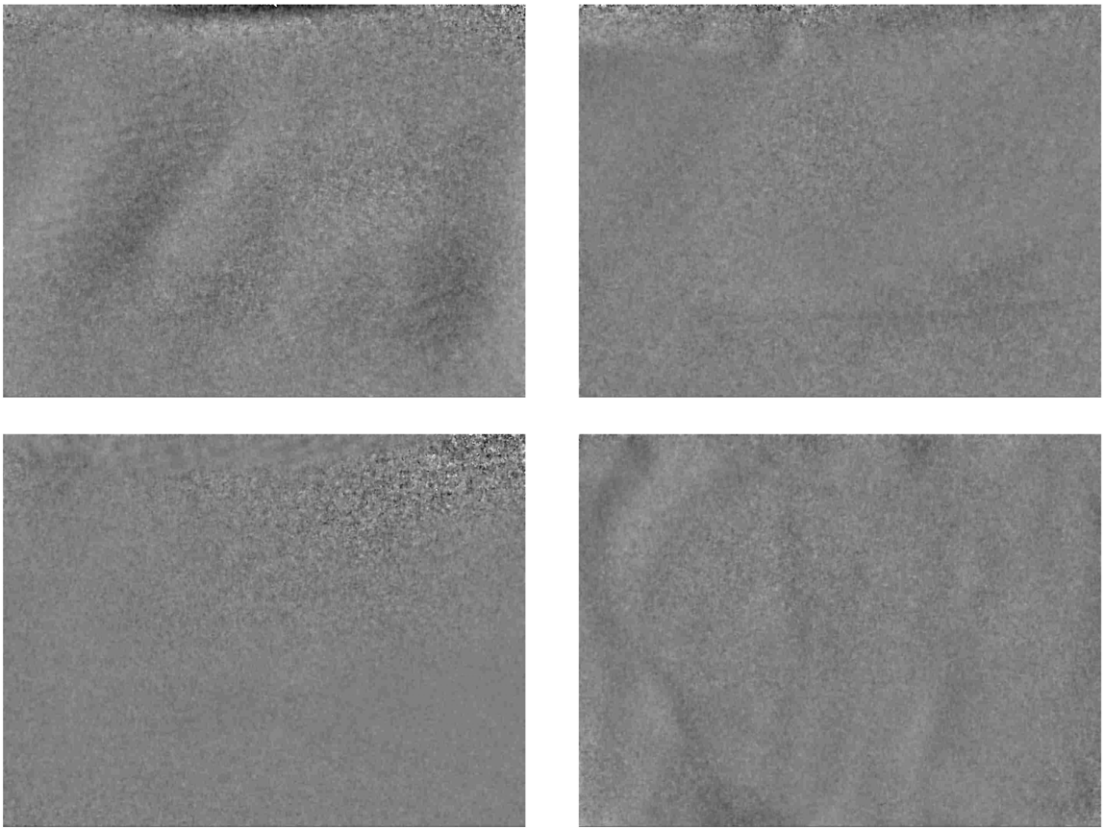


Figure 3. Shearography images examples of the good class.

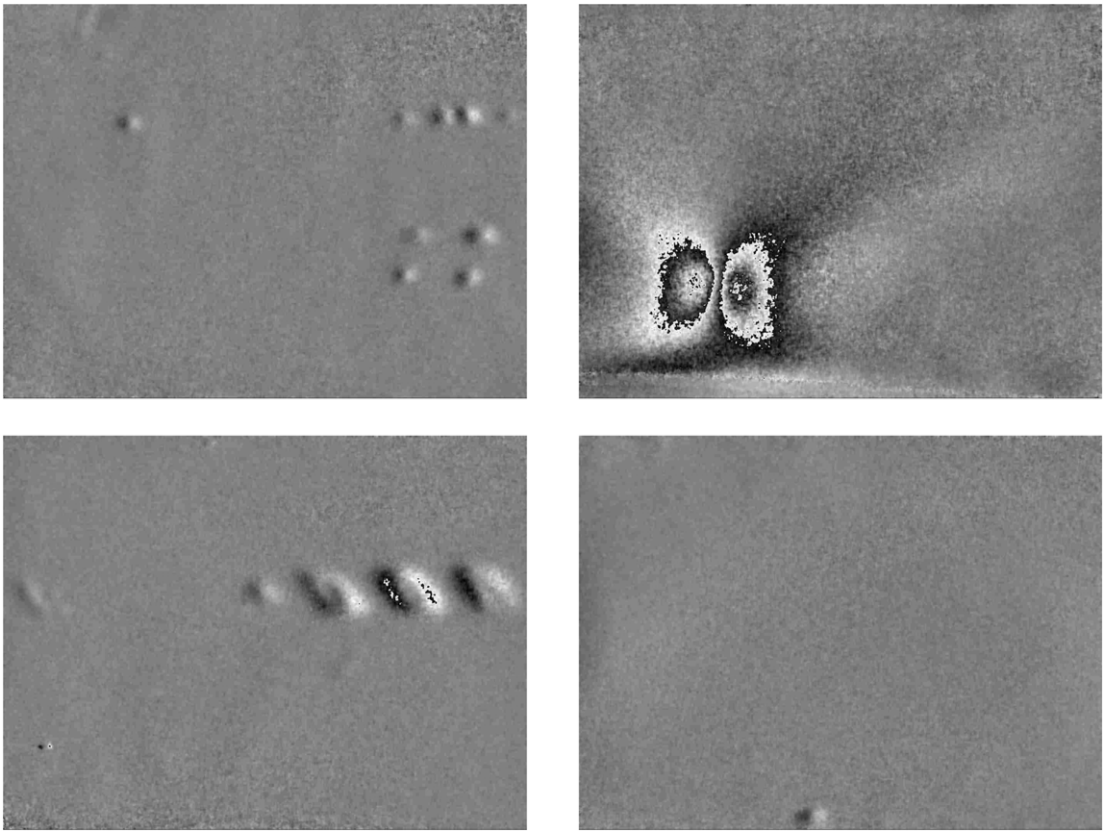


Figure 4. Shearography image examples of the bad class.

3.3. Transfer learning

In traditional machine learning applications, once the model has been trained with the data that is already available, it is tested with data that it has never seen before. As more data is used in the training process, the model starts to learn more, and its performance increases. When the training data or test data changes, or when new situations add to an existing dataset, the system needs to be retrained. Transfer learning tries to solve a complex problem by using a model that has already been trained (pre-trained) on a large dataset to solve a simple problem with a small dataset.

Transfer learning is a lot like how humans learn from their past experiences. When transfer learning is used to solve a new problem, the model does not have to be trained from scratch. Instead, the gained knowledge is used to solve the new problem. So, the process of learning is sped up, and knowledge gained before is added to the model to be used for the new problem. Transfer learning is also better when there isn't enough data to use for training [16]. There are a lot of pre-trained transfer-learning models, but in this study, only VGG16, VGG19, ResNet50, Xception, MobileNetV2 and DenseNet201 were used. Images from different classes are easy to put into different groups. Changes are made to certain layers of the pre-trained models to update weights and biases for different applications [17].

3.3.1. VGG16 and VGG19 models

This model, which was released in 2014, is largely considered to be one of the most effective image recognition available models. Additionally, this model offered several essential qualities. It accepts 224×224 RGB images. The VGG16 model has 16 weight layers, making it complex. There are thirteen convolutional layers with a 3×3 filter size and fully connected layers in the VGG16 model [18, 19], which is far more sophisticated than any other model. The VGG19 model has 19 layers. Also VGG19 model has sixteen convolutional layers with a 3×3 filter size and fully connected layers [20].

3.3.2. ResNets model

Residual Networks, often known as ResNets are a type of deep convolutional network [21]. These networks are based on the idea of bypassing entire convolutional layer blocks by making use of shortcut connections. The building blocks, also known as bottleneck blocks, are constrained by two principles. First, if the output feature map's size is constant, so is its filter number. Second, if the map size of the feature is reduced by one-half, then the number of filters is increasing by one. Also, batch normalization is done immediately in stride 2 by convolutional layers after each convolution and before activating the ReLU layer.

The identity shortcut is used if the input and output are inside identical dimensions. Whereas, if increasing dimensions, then to match dimensions through 1×1 convolutions is used the projection shortcut. Shortcuts through feature maps of different sizes are carried out with a stride of 2. The last layer of the network is a fully connected one that has SoftMax activation for 1000 nodes. Total trainable parameters are 23534592, and the model has 50 weighted layers [22].

3.3.3. Xception model

Another example of a transfer learning model is the Xception model which was developed [18]. The idea of depth-wise separable convolutions (DSC) is utilized in the construction of this model. DSC is a type of convolution that has been modified to be more efficient in terms of calculation time. DSC uses depth-wise and point-wise convolution.

Xception's input image is RGB, 299×299, and 126 depth. It has 36 convolutional layers to extract characteristics from images By changing the fully-connected layer with a global average pooling layer, the SoftMax function is used to make the prediction. They were able to reduce the number of parameters and the complexity of the model [19]. Except for the first and last layers, all of the intermediate layers contain linear skip connections. After going through the input layer, the data are processed 8 times through the intermediate layer before the output layer.

3.3.4. MobileNetV2 model

MobileNetv2 was proposed by Sandler *et al* [23] to improve to reduce computation and enhance accuracy with an inverted residual and linear bottleneck module. Unlike traditional residual blocks, this module expands low-dimensional inputs, applies depthwise convolution, and then projects back to lower dimensions. MobileNetv2 has a convolutional layer, 19 residual bottleneck modules, and two final convolution layers. A shortcut is used only when the stride is 1.

3.3.5. Dense convolutional network (DenseNet201) Model

Huang *et al* [24] proposed DenseNet201, a deep CNN. DenseNet201's CNNs perform astonishingly. The model's success can be attributed in large part to its ability to reuse features for better parameter estimates and

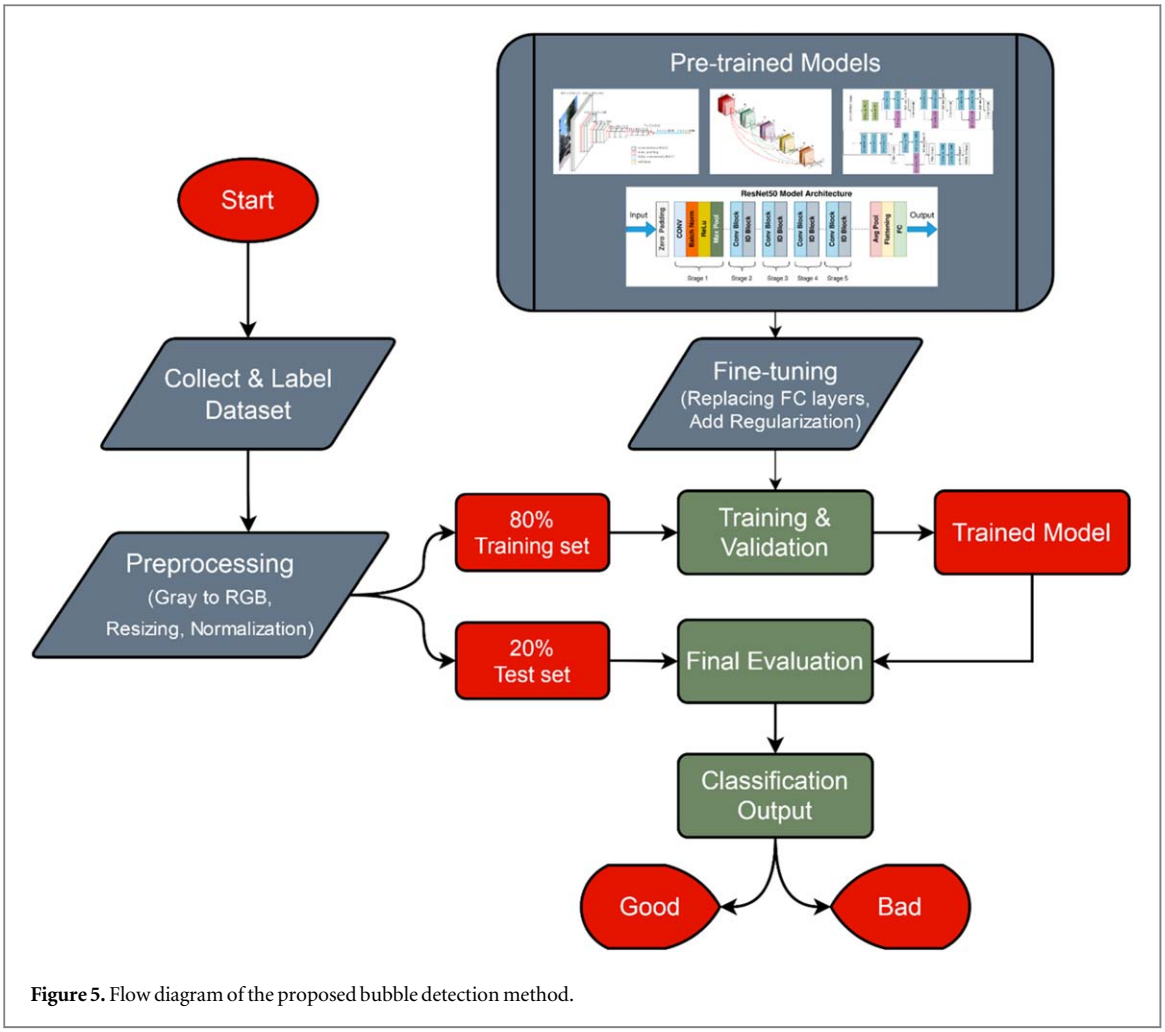


Figure 5. Flow diagram of the proposed bubble detection method.

more efficient training. When creating a new layer, the feature maps from the previous layer and the maps from the next layer are combined to create the new layer. The previous layers' feature maps may now be accessed easily by the network deep layers, allowing the features to be utilized. Each of DenseNet201's 305 layers contains at least one kernel [24].

3.4. The proposed method

To classify air bubble defects in tire shearography images, the VGG16, VGG19, ResNet50, Xception, MobileNetV2 and DenseNet201 models have been utilized. The models utilized in this study aimed to detect whether or not a given image of a tire shearography shows the presence of an air bubble by classifying the image into two classes; good class for images without air bubbles and bad class for images with an air bubble. The flow diagram of the proposed classification model is shown in figure 5.

At first, the dataset was collected from the Pirelli Automobile Tires factory-Izmit and labeled by an expert into good and bad classes. Since the images in our dataset were gray images and the Transfer Learning models were trained with RGB images, they were converted into RGB images. After that, the images were resized to match the pre-trained models' input size. Normalizing the image was also an important step to reduce the computational cost. Next, the dataset was split into 80% training set and 20% test set. The models used in this study were fine-tuned by freezing the pre-trained weights in the convolutional layers and replacing the fully connected layers with a new fully connected layer to be able to use it for the classification of our specific problem and achieve high accuracy and recall scores. The number of neurons in the new fully connected layer was chosen to be (100, 128, 256, and 500). Then, the 80% training set was used to train the new fully connected layer of the models. The best outcomes were obtained with 500 neurons. The dropout rate has been estimated to range from 0.1 to 0.6. Additionally, 200 epoch numbers were used. The accuracy and recall of each model were optimized by modifying the model's hyperparameters. These hyperparameters are summarized in table 2.

Initially, the dataset was acquired from the Pirelli Automobile Tires factory in İzmit and was meticulously labeled by an expert into two categories: good and defective classes. Given that the images were originally in

Table 2. The hyperparameters values of models.

Models	Value
Numbers of Epochs	200
Numbers of Neurons	500
Batch Size	64
Dropout Rate	0.6
Optimizer	Adam
Regularization	L1 = 1e-4, L2 = 1e-5

grayscale format, and the transfer learning models were designed to process RGB images, all images were converted into RGB format to ensure compatibility.

Subsequently, the images were resized according to the input size requirements of the pre-trained models used in this study—224 × 224 pixels for VGG16, VGG19, ResNet50, MobileNetV2 and DenseNet201 models, and 299 × 299 pixels for the Xception model. This resizing step was essential for maintaining consistency with the models' architecture. Furthermore, image normalization was performed to standardize the pixel values, which is crucial for reducing computational complexity and enhancing training efficiency. The dataset was then split into an 80% training set and a 20% test set, ensuring a balanced distribution that allows for robust model evaluation. The transfer learning models were fine-tuned by freezing the convolutional layers' pre-trained weights and replacing the fully connected layers with new ones tailored to the specific classification task. The number of neurons in the new fully connected layer was varied (100, 128, 256, and 500) to identify the configuration yielding the highest performance. The optimal results, in terms of both accuracy and recall, were achieved with 500 neurons. A range of dropout rates (0.1 to 0.6) was explored to mitigate overfitting, as shown in table 2. The dropout rate of 0.6 was selected as the most effective value, particularly given the relatively small dataset size. The training process was conducted over 200 epochs, a duration sufficient to achieve convergence while minimizing the risk of overfitting, as demonstrated through the early stopping criteria. To optimize performance, the Adam optimizer was employed due to its adaptive learning rate mechanism, which facilitates efficient convergence. Additionally, L1 and L2 regularization techniques (L1 = 1e-4, L2 = 1e-5) were applied, as indicated in table 2, to penalize large weights and improve the generalization capability of the models. These hyperparameter configurations were carefully selected and fine-tuned to maximize the classification models' performance, focusing on achieving high accuracy and recall rates.

Initially, the classification models were trained using standard configurations, and baseline results were obtained. To limit the impacts of overfitting, we applied L1 regularization for each model, with the parameter set to 1e-4. After completing the training process, the final performance of the models was evaluated using a 20% test set to ensure reliable assessment of generalization capabilities. Subsequently, to further validate and enhance the reliability of the results, we incorporated L2 regularization technique alongside L1 (L1 = 1e-4, L2 = 1e-5) technique. This combined approach aims to improve the models' generalization by penalizing large weights and minimizing overfitting effects. Additionally, the ReduceLROnPlateau technique from the Keras library was utilized. This method dynamically reduces the learning rate when the recall metric does not show improvement over a specified number of epochs. The parameters were configured as follows: monitor='recall', factor = 0.3, patience = 6, and min_lr = 1e-5. This approach was employed after observing the initial results, with the goal of stabilizing and optimizing the training process further when performance gains plateaued. Following these modifications, the models were retrained, and the new results were compared against the baseline performance to evaluate the effectiveness of the regularization techniques and adaptive learning rate strategy.

The study experiments were implemented on a workstation with an Intel Core i7 32002 2.00 GHz CPU, Nvidia GTX 1080 Ti 8 GB GPU, 32 GB RAM, and Windows 10 OS. The implementation of the models was in Python 3.7 using the Keras library [25] on the Tensor Flow back-end [26].

4. Experimental results and discussion

The metrics of recall, accuracy, precision, and *F*1-score are calculated to evaluate the performance of this model. These formulas for the metrics are presented as follows in equations (1)–(4).

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

Table 3. Tire bubble defect classification results for both the bad and good classes.

Model	Accuracy	Precision	Recall	F1_score	ROC AUC score
VGG16	91.85%	92.55%	85%	91.81%	96.35%
VGG19	93.13%	93.96%	86%	93.1%	96.7%
ResNet50	71.67%	78.78%	47%	71.67%	87.63%
Xception	93.56%	93.59%	92%	93.56%	97.65%
MobileNetV2	90.68%	83.76%	93.33%	88.29%	96.1%
DenseNet201	**94.85%**	**94.97%**	**92%**	**94.85%**	**98.32%**

Table 4. Tire bubble defect classification results with reduced overfitting.

Model	Accuracy	Precision	Recall	F1_score	ROC AUC score
VGG16	92.27%	93.08%	85%	92.24%	96.43%
VGG19	93.56%	94.29%	87%	93.53%	96.73%
ResNet50	72.96%	78.92%	50%	71.46%	87.67%
Xception	93.13%	93.15%	92%	93.13%	97.63%
MobileNetV2	91.35%	83.96%	93.78	89.11	96.2%
DenseNet201	**95.71%**	**95.72%**	**95%**	**95.71%**	**98.52%**

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1Score = \frac{2*Precision*Recall}{Precision + Recall} \tag{4}$$

It is important to note that the objective of the proposed model is bubble detection; thus, the positive class corresponds to the defective (bad) class (presence of bubbles), while the negative class corresponds to the non-defective (good) class (absence of bubbles). When it comes to the detection of air bubbles, with accuracy the second most essential metric to focus on is recall, which refers to the accurate detection rate of the bubble class [10]. Then there are additional metrics that are taken into consideration as well. Table 3 shows the classification results of the recall, precision, and F1-score for both the bad and good classes of the VGG16, VGG19, ResNet50, Xception, MobileNetV2 and DenseNet201 models.

As shown in table 3 the results of the proposed models with 200 epochs are shown to be superior to the results. In addition, the DenseNet201 model achieved the best results for all the metrics, including accuracy, precision, recall, F1 score, and ROC AUC score, with respective values of 94.85%, 94.97%, 92%, 94.85%, and 98.32%. While the results are promising, to further ensure the reliability of the outcomes, L2 regularization was incorporated alongside L1 (L1 = 1e-4, L2 = 1e-5) to mitigate overfitting. We have also utilized a technique in Keras so-called ReduceLROnPlateau to reduce the learning rate if Recall metric stopped improving for a number epoch. In our case, we set the parameters as follows: ReduceLROnPlateau (monitor='recall', factor = 0.3, patience = 6, min_lr = 0.00001).

Following the implementation of overfitting prevention techniques, the updated bubble classification results are presented in table 4. As indicated in table 4, the highest performance is achieved using the DenseNet201 model, with improvements observed across all evaluation metrics: accuracy (95.71%), precision (95.72%), recall (95%), F1 score (95.71%), and ROC AUC (98.52%). These results demonstrate that the strategies employed to mitigate overfitting have effectively enhanced classification performance. Moreover, all models exhibit superior outcomes compared to the results reported in table 3.

The comparison between tables 3 and 4 demonstrates the significant impact of applying L2 regularization and the ReduceLROnPlateau strategy across all models. DenseNet201 shows the most pronounced improvement, with accuracy increasing from 94.85% to 95.71%, recall rising from 92% to 95%, and other metrics such as precision and F1 score also reaching higher levels. This highlights DenseNet201's responsiveness to the applied strategies. The VGG models (VGG16 and VGG19) also exhibit moderate gains; VGG16 shows improvements in accuracy and precision while maintaining its recall, whereas VGG19 improves in all metrics, particularly recall, which increases from 86% to 87%. The Xception and MobileNetV2 models maintains stable performance, with recall and precision remaining high, indicating its robustness even after adjustments. However, the ResNet50 model, which initially had the lowest metrics, shows slight enhancements, particularly in recall, moving from 47% to 50%. Overall, while the strategies effectively minimize overfitting and enhance performance, the degree of improvement varies, with DenseNet201 benefiting the most.

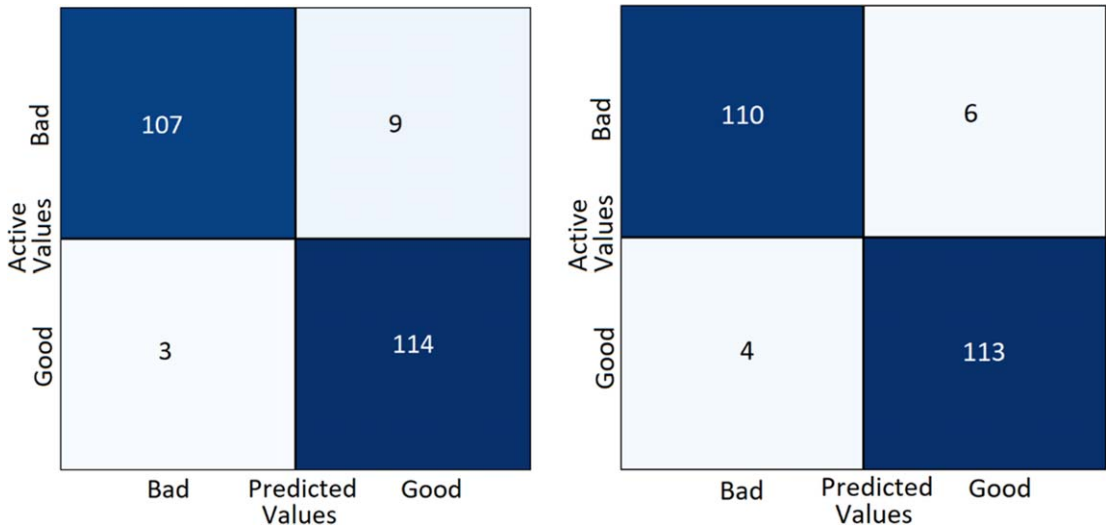


Figure 6. Confusion matrix of the DenseNet201 model for proposed classification method. Normal results on the left hand side and for reduced overfitting results on the right hand side.

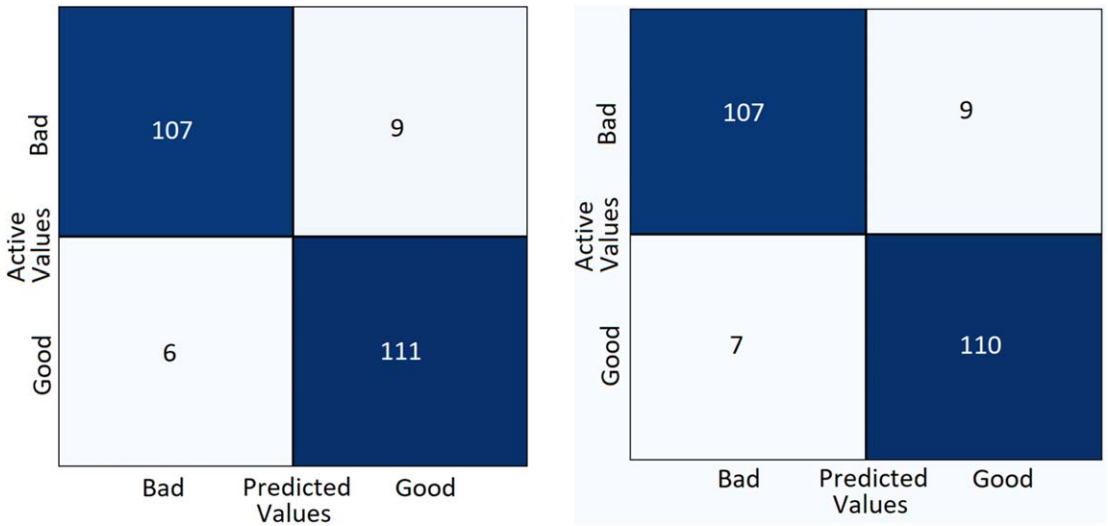


Figure 7. Confusion matrix of the Xception model for proposed classification method. Normal results on the left hand side and for reduced overfitting results on the right hand side.

A confusion matrix representing test data for a DenseNet201 model is presented in figure 6. The size of the test dataset is 233 that is 20% of the dataset with 116 bad (positive/with bubble) and 117 good (negative/without bubble). The DenseNet201 model initially recorded 107 True Positives (TP) and 114 True Negatives (TN) under normal conditions. After implementing overfitting reduction techniques, the model's TP value increased to 110. The proposed architecture classified most of the bad tires into the true class. As mentioned before, as well as accuracy, recall value is a very important metric for defect detection methods. Because one of the most important cases is the detection of a faulty class of tires. In the proposed method, the second best recall value is obtained with the Xception model. So, the confusion matrices of the second-best classifier model are given in figure 7.

The convergence plots of the training and validation results with regard to epochs for the two best models for recall metric are depicted in the figures 8, 9 when the number of epochs is 200. The convergence of losses is represented by the figures on the right side of the graphs, whereas the convergence of accuracies is represented by the figures on the left side of the graphs.

In spite of the fact that our models include regularization and dropout techniques, it is abundantly obvious that overfitting occurred as a result of the insufficient amount of dataset. Nevertheless, the overfitting problem with the DenseNet201 model is not as significant as it is in other models, which highlights the higher

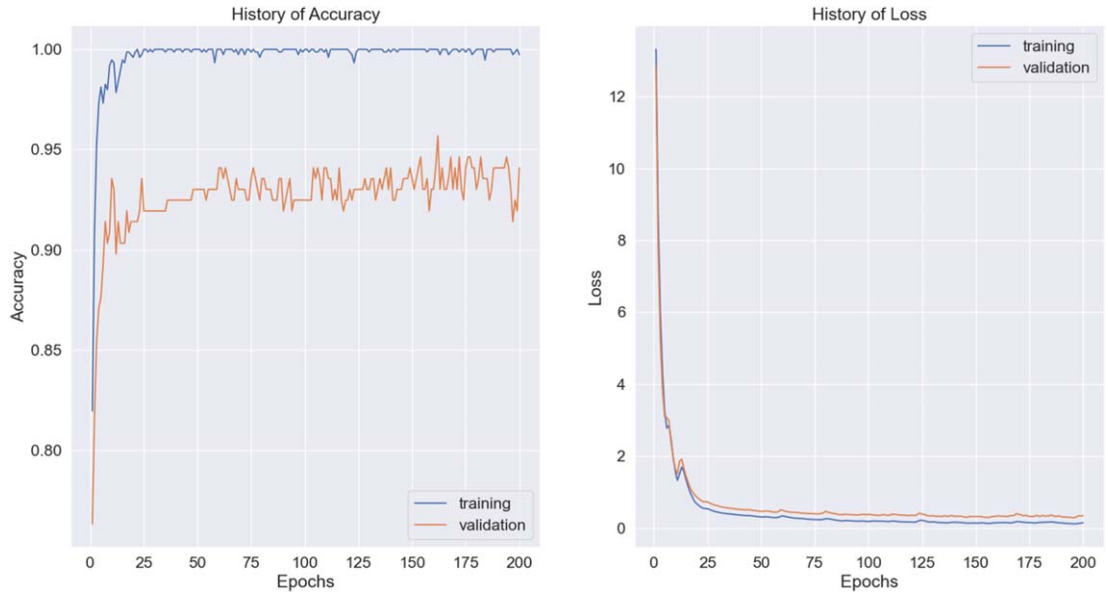


Figure 8. The training accuracy and loss of the DenseNet201 model for 200 epochs.

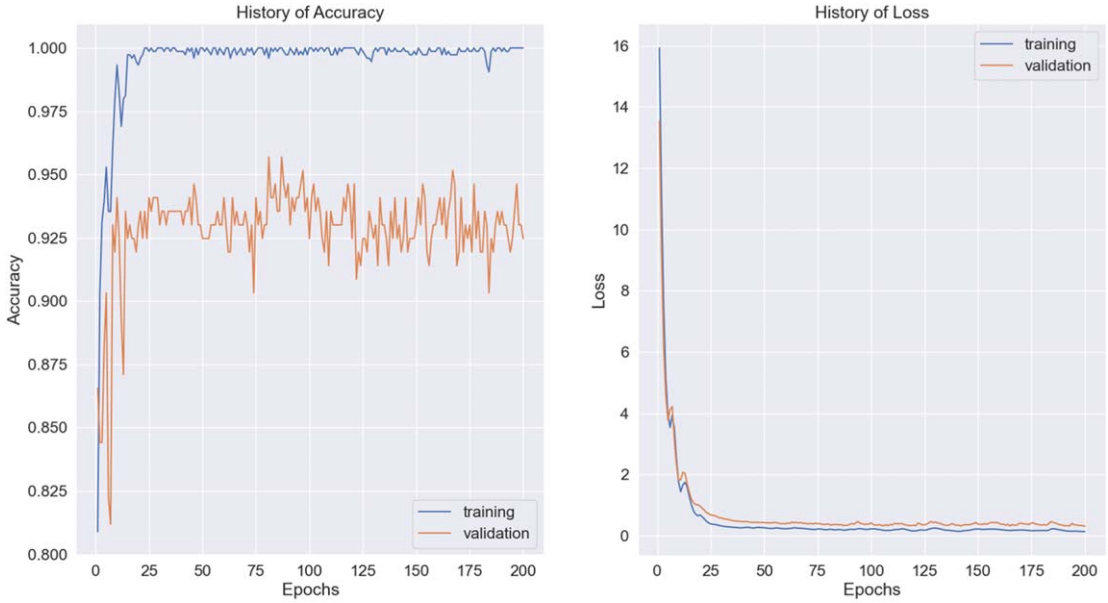


Figure 9. The training accuracy and loss of the Xception model for 200 epochs.

outperformance of the DenseNet201 model in comparison to other models. The ROC-AUC curves of the DenseNet201 and Xception models are shown in figures 10 and 11, respectively.

In figure 12, we present visual examples of the DenseNet201 model's erroneous predictions; namely false positives and false negatives. These examples highlight the model's limitations in challenging cases. The false positives often occur in regions with noise or artifacts that the model mistakenly identifies as bubbles, highlighting the need for better discrimination in these cases. Conversely, the false negatives tend to arise in challenging scenarios where bubbles are faint, overlapping, or partially obscured, indicating areas where the model struggles to detect more subtle features. These visualizations provide a clearer understanding of the model's performance and areas for potential enhancement.

The detection of air bubbles in the shearography images using transfer learning models is an emerging challenge in the field of research. As a result, there are not sufficient studies in the literature. So, the results of this study are firstly compared with few existing tire defect detection methods. Table 5 shows the result compared

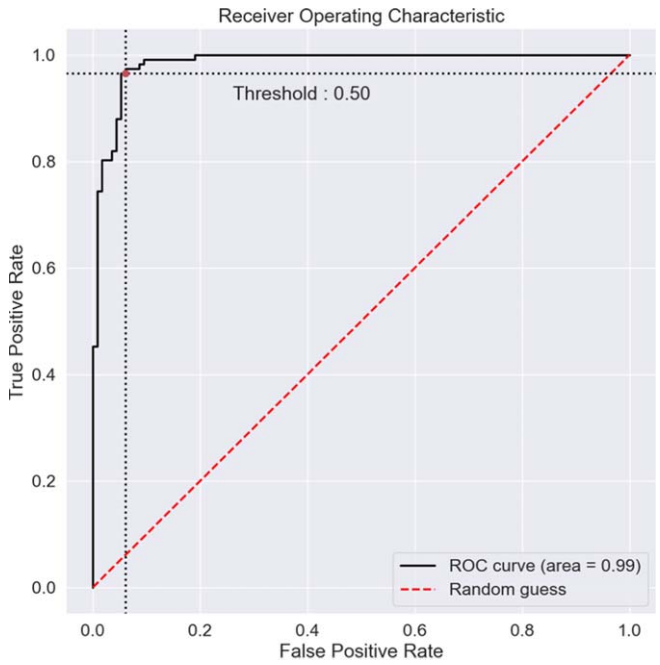


Figure 10. The ROC AUC curve of the DenseNet201 model.

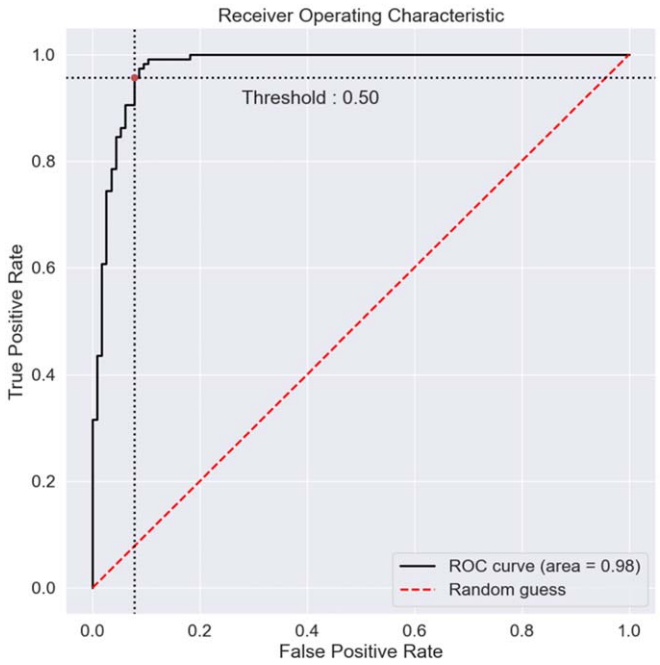


Figure 11. The ROC AUC curve of the Xception model.

with the state of the art tire defect detection methods. When examining this table, it can be seen that our results achieve similar accuracy where other methods did not have any clear information for other important metrics.

The existing literature on shearography remains limited and requires further exploration. As far as the authors know, only Chang *et al* [10] focused on the bubble defect detection of tire shearography images. As shown in table 6 the proposed methods achieve better accuracy for all transfer learning-based models with over 93% when comparing Chang *et al* [10]'s 89% accuracy. In the bubble detection problem, the other essential metric is recall. Proposed models achieve 95%, 92%, and 87% recall values for DenseNet201, Xception, and VGG19, respectively. Chang *et al* [10] study was claimed to achieve very good recall values, but their other metrics are not good enough where they did not give any confusion matrix in their paper. Overall metrics must be high for a quality control system. Because, in the industrial process, not only focusing on the defect is enough,

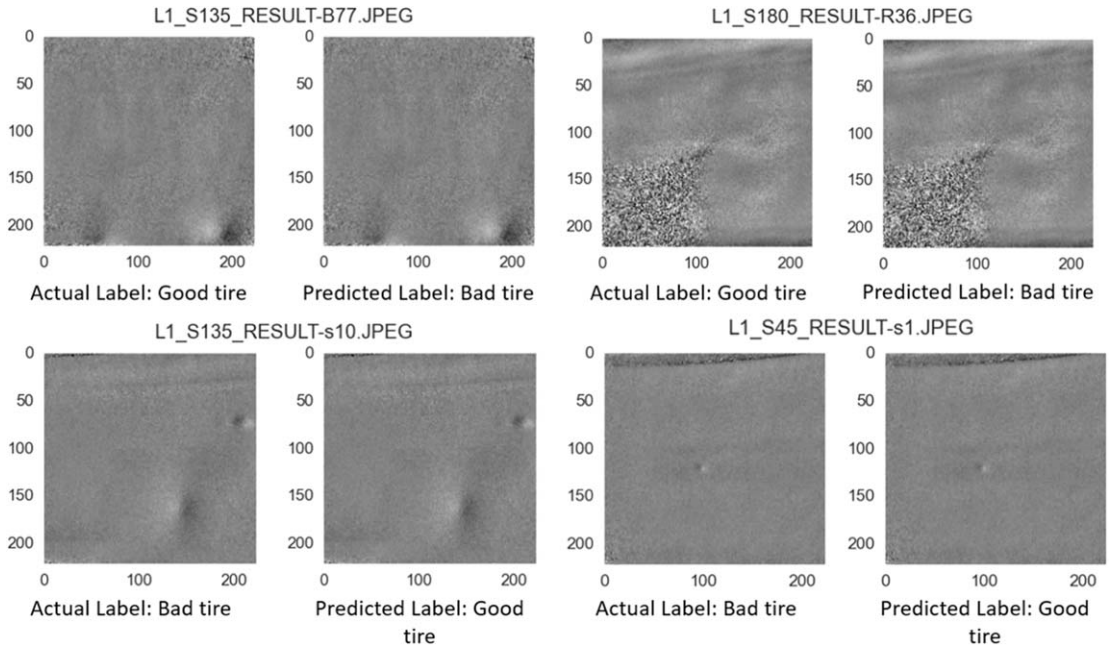


Figure 12. False positives examples (Top), and false negative examples (bottom).

Table 5. Comparison with the state-of-the-art tire defect detection methods.

Method	Model architecture	Dataset size	Imaging method	Accuracy	Precision	Recall	F1_score
Zhu and Ai [3]	Faster RCNN	—	x-ray	96.80%	—	—	—
Zhang *et al* [12]	Support Vector Machine	400	x-ray	95.76%	—	—	—
Wang *et al* [14]	VGG16	914	x-ray	78.91%	—	—	—
Yang *et al* [15]	Image processing techniques	200	x-ray	91.50%	—	—	—
	DenseNet201	1392	Shearography	95.71%	95.72%	95%	95.71%
The Proposed	VGG19	1392	Shearography	93.56%	94.29%	87%	93.53%
	Xception	1392	Shearography	93.13%	93.15%	92%	93.13%

Table 6. Comparison with the shearography based state of the art tire bubble detection methods.

Method	Architecture	Imaging technic	Dataset size	Accuracy	Precision	Recall	F1_score
Chang *et al* [10]	SVM	Shearography	4076	50.13%	36.56%	92.65%	52.43%
	RF	Shearography	4076	44.61%	26.88%	96.32%	42.03%
	Faster Region-based CNN	Shearography	4076	89.16%	83.09%	100%	90.76%
The proposed	DenseNet201	Shearography	1392	95.71%	95.72%	95%	95.71%
	VGG19	Shearography	1392	93.56%	94.29%	87%	93.53%
	Xception	Shearography	1392	93.13%	93.15%	92%	93.13%

but also keeping good tires is important. The model of the Chang *et al* [10] has overfitting and classified a lot of good tires into the defective class.

5. Conclusion

Air bubbles in tires pose significant safety risks, potentially leading to traffic accidents and endangering human lives. To mitigate these risks, tire manufacturers conduct inspections during the production process to detect air bubbles and ensure safe and reliable travel. Automated classification of air bubbles in tires, therefore, plays a critical role in enhancing safety standards and reducing the reliance on manual inspection. In this study, we present a comparative analysis of transfer learning-based deep convolutional neural networks (CNNs) for detecting air bubble defects in tire shearography images. Our findings demonstrate the efficacy of models such as DenseNet201, which achieved the highest performance metrics, surpassing 95% in accuracy, recall, precision,

and F1-score. This level of performance is highly competitive and exceeds the results documented in the current literature, underscoring the potential of the proposed model for real-world industrial implementation. However, despite these promising outcomes, the study has limitations, particularly concerning the size of the dataset and the challenges associated with overfitting. Expanding the dataset or incorporating advanced data augmentation techniques could significantly enhance the model's generalization capabilities and mitigate these issues. The proposed models' hyperparameters are also flexible and can be further optimized through various methods, providing opportunities to fine-tune and maximize performance metrics even further. This highlights a clear pathway for future research, including the extension of the methodology to other industrial defect detection applications. Furthermore, practical implementation challenges, including model deployment strategies, real-time processing requirements, and data quality considerations, can influence the scalability and robustness of the proposed approach. Tackling these challenges will be essential for translating this research into a fully operational and industry-ready solution.

Future work will focus on developing a larger and more diverse dataset to validate and retrain the models, while also exploring and implementing advanced architectures to enhance the system's scalability and robustness, thereby fully meeting the demands of industrial defect detection tasks. By addressing these limitations and exploring additional applications, subsequent studies aim to enhance the real-world applicability and effectiveness of the proposed approach.

Acknowledgments

This study was supported by Kocaeli University Scientific Research Projects Coordination Unit and Pirelli Automobile Tires İzmit Factory with Project Number: FKA-2022-2772. All results were carried out in Kocaeli University Software Engineering Department Software Technologies Research Laboratory (STAR Lab).

Data availability statement

The data cannot be made publicly available upon publication because they contain commercially sensitive information. The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Farid Al-Areqi https://orcid.org/0000-0001-7607-993X
Mehmet Zeki Konyar https://orcid.org/0000-0001-8914-5553
Kaplan Kaplan https://orcid.org/0000-0001-8036-1145
H Metin Ertunç https://orcid.org/0000-0003-1874-3104

References

[1] Lelli F 2019 Interoperability of the time of industry 4.0 and the internet of things *Futur. Internet* **11** 36
[2] Zhao Q, Dan X, Sun F, Wang Y, Wu S and Yang L 2018 Digital shearography for NDT: phase measurement technique and recent developments *Appl. Sci.* **8** 2662
[3] Zhu Q and Ai X 2018 The defect detection algorithm for tire x-ray images based on deep learning *IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)* **2018** 138–42
[4] Hung J Y, Gao W and Hung J C 1993 Variable structure control: a survey *IEEE Trans. Ind. Electron.* **40** 2–22
[5] Steinchen W, Yang L X, Schuth M and Kupfer G 1994 Electronic shearography (ESPSI) for direct measurement of strains *Optical Measurements and Sensors for the Process Industries* **2248** 210–21
[6] Creath K 1988 V phase-measurement interferometry techniques *Progress in Optics* **26** 349–93
[7] Saleh N A, Konyar M Z, Kaplan K, Ongir S and Ertunç H M 2022 Detection of air bubbles from tire shearography images *In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)* (IEEE) 1–4
[8] De Angelis G, Meo M, Almond D P, Pickering S G and Angioni S L 2012 A new technique to detect defect size and depth in composite structures using digital shearography and unconstrained optimization *Ndt & E International* **45** 91–6
[9] Gayer A and Saya A 1988 The use of X-radiography and computer software for detecting defects during the manufacture of steel-belt tyres *NDT Int.* **21** 333–6
[10] Chang C-Y, Srinivasan K, Wang W-C, Ganapathy G P, Vincent D R and Deepa N 2020 Quality assessment of tire shearography images via ensemble hybrid faster region-based ConvNets *Electronics* **9** 45
[11] Zhang Y, Li T and Li Q 2013 Defect detection for tire laser shearography image using curvelet transform based edge detector *Opt. Laser Technol.* **47** 64–71
[12] Zhang Y, Lefebvre D and Li Q 2015 Automatic detection of defects in tire radiographic images *IEEE Trans. Autom. Sci. Eng.* **14** 1378–86
[13] Cui X H, Liu Y, Wang C X and Li H 2016 A novel method for feature extraction and automatic recognition of tire defects using independent component analysis *Int. Conf. on Intelligent Manufacturing and Materials (ICIMM 2016) Statesboro, GA, U.S* (https://doi.org/10.12783/dtmse/icimm2016/6240)

[14] Wang R, Guo Q, Lu S and Zhang C 2019 Tire defect detection using fully convolutional network *IEEE Access* **7** 43502–10
[15] Yang H, Jiang Y, Deng F, Mu Y, Zhong Y and Jiao D 2022 Detection of bubble defects on tire surface based on line laser and machine vision *Processes* **10** 255
[16] Guo T, Dong J, Li H and Gao Y 2017 Simple convolutional neural network on image classification *In 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (IEEE) 721–4
[17] Ahmed M I, Mamun S M and Asif A U Z 2021 DCNN-based vegetable image classification using transfer learning: a comparative study *5th International Conference on Computer, Communication and Signal Processing (ICCCSP)* 235–43
[18] Chollet F 2017 Xception: deep learning with depthwise separable convolutions *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1251–8
[19] Lo W W, Yang X and Wang Y 2019 An xception convolutional neural network for malware classification with transfer learning *10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* 2019, 1–5
[20] Hridayami P, Putra I K G D and Wibawa K S 2019 Fish species recognition using VGG16 deep convolutional neural network *J. Comput. Sci. Eng.* **13** 124–30
[21] Shehab L H, Fahmy O M, Gasser S M and El-Mahallawy M S 2021 An efficient brain tumor image segmentation based on deep residual networks (ResNets) *Journal of King Saud University-Engineering Sciences* **33** 404–12
[22] Rezende E, Ruppert G, Carvalho T, Ramos F and De Geus P 2017 Malicious software classification using transfer learning of resnet-50 deep neural network *16th IEEE International Conference on Machine Learning and Applications (ICMLA)* 2017, 1011–4
[23] Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L C 2018 MobileNetV2: inverted residuals and linear bottlenecks *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 4510–20
[24] Huang G, Liu Z, Pleiss G, Van Der Maaten L and Weinberger K 2019 Convolutional networks with dense connectivity *IEEE Trans. Pattern Anal. Mach. Intell.* **44** 8704–16
[25] Chollet F *et al* 2015 Keras: the Python deep learning API (https://keras.io/)
[26] Abadi M *et al* 2015 TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.