Konya Mühendislik Bilimleri Dergisi, c. 12, s.3, 700-713, 2024
Konya Journal of Engineering Sciences, v. 12, n. 3, 700-713, 2024
ISSN: 2667-8055 (Elektronic)
DOI: 10.36306/konjes.1424329

A DEEP LEARNING ENSEMBLE APPROACH FOR X-RAY IMAGE CLASSIFICATION

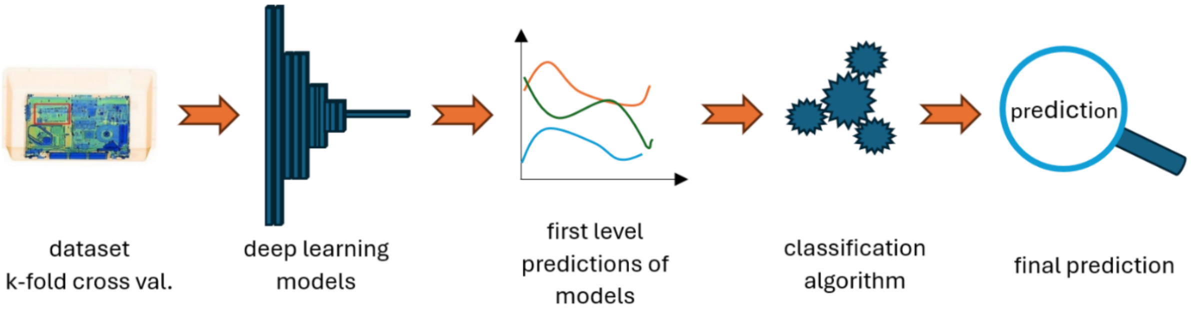
[1]**Engin ESME**, [2,*] **Mustafa Servet KIRAN**

[1]*Konya Technical University, Engineering and Natural Sciences Faculty, Software Engineering Department, Konya, TÜRKİYE*
[2]*Konya Technical University, Engineering and Natural Sciences Faculty, Computer Engineering Department, Konya, TÜRKİYE*
[1]eesme@ktun.edu.tr, [2]mskiran@ktun.edu.tr

Highlights

- Automating X-ray imaging using deep learning methods to overcome time and personnel constraints.
- Training deep learning models on X-ray images for detecting hidden explosive circuits.
- Developing an ensemble approach to enhance classification success.

Graphical Absract



***Corresponding Author:** Mustafa Servet KIRAN, mskiran@ktun.edu.tr

Konya Mühendislik Bilimleri Dergisi, c. 12, s.3, 700-713, 2024
Konya Journal of Engineering Sciences, v. 12, n. 3, 700-713, 2024
ISSN: 2667-8055 (Elektronic)
DOI: 10.36306/konjes.1424329

A DEEP LEARNING ENSEMBLE APPROACH FOR X-RAY IMAGE CLASSIFICATION

[1]Engin ESME, [2,*] Mustafa Servet KIRAN

[1]Konya Technical University, Engineering and Natural Sciences Faculty, Software Engineering Department, Konya, TÜRKİYE
[2]Konya Technical University, Engineering and Natural Sciences Faculty, Computer Engineering Department, Konya, TÜRKİYE
[1]eesme@ktun.edu.tr, [2]mskiran@ktun.edu.tr

(Received: 23.01.2024; Accepted in Revised Form: 19.07.2024)

ABSTRACT: The application of deep learning-based intelligent systems for X-ray imaging in various settings, including transportation, customs inspections, and public security, to identify hidden or prohibited objects are discussed in this study. In busy environments, x-ray inspections face challenges due to time limitations and a lack of qualified personnel. Deep learning algorithms can automate the imaging process, enhancing object detection and improving safety. This study uses a dataset of 5094 x-ray images of laptops with hidden foreign circuits and normal ones, training 11 deep learning algorithms with the 10-fold cross-validation method. The predictions of deep learning models selected based on the 70% threshold value have been combined using a meta-learner. ShuffleNet has the highest individual performance with 83.56%, followed by InceptionV3 at 81.30%, Darknet19 at 78.92%, DenseNet201 at 77.70% and Xception at 71.26%. Combining these models into an ensemble achieved a remarkable classification success rate of 85.97%, exceeding the performance of any individual model. The ensemble learning approach provides a more stable prediction output, reducing standard deviation among folds as well. This research highlights the potential for safer and more effective X-ray inspections through advanced machine learning techniques.

Keywords: Deep Learning, Ensemble Learning, Object Classification, X-Ray

1. INTRODUCTION

Passenger and cargo transportation, customs inspections, public buildings, and public areas commonly use X-ray imaging devices to detect hidden or prohibited objects and identify potential hazards. However, in places with heavy traffic, X-ray inspections can become inefficient and even unsafe due to time limitations, caseload, and lack of qualified personnel. Especially when items are stacked on each other, the images become complex, and it becomes challenging for operators to focus on the scanning screen for long periods. Moreover, situations like detecting explosive materials and related circuits require additional expertise. As a result, the problem of identifying threat objects concealed inside electronic devices arises. In this context, deep learning-based intelligent systems can automate the imaging process and object detection, making inspections safer. Deep learning algorithms are being explored in some research to evaluate X-ray security images since they can automatically extract high-level features compared to traditional image processing methods. Akçay, Kundegorski et al. examine transfer learning by applying Convolutional Neural Networks (CNNs) to the image classification problem used in X-ray luggage screening. Due to less training data, fully training CNN is difficult. Therefore, the last layers of the network are frozen, and only the initial layers are fine-tuned and optimized. The proposed method achieves more successful results than previous studies and is effective in detecting firearms [1]. Benedykciuk, Denkowski et al. address the material detection problem in the X-ray scanners used for security purposes. Scanner images are divided into six main categories based on whether organic or metallic. Feature extraction and classification are performed using deep learning methods. During

*Corresponding Author:** Mustafa Servet KIRAN, mskiran@ktun.edu.tr training, the images are split into parts representing six different material types, and a multi-scale network structure consisting of five sub-networks is used to handle size variations. Additionally, the effects of regularization and activation approaches such as (Exponential Linear Unit) ELU and Rectified Linear Units (RELU) on the architecture are also investigated [2]. Miao, Xie et al. conduct research on the detection of prohibited items in X-ray images. For this purpose, they create a dataset called SIXray, which is 100 times larger and resembles real-world data. The dataset contains approximately 1% of prohibited items. They develop an algorithm to hierarchically and iteratively enhance features to suppress the number of overlapping objects and irrelevant information in X-ray images. To improve the efficiency of the slow-working iterative algorithm, they limit forward and backward passes with a pruning mechanism. Additionally, they introduce a new loss function to address the data imbalance between positive and negative classes and achieve impressive results by testing it with different network frameworks [3]. Chang, Zhang et al. address the object detection problem in X-ray security imaging using a two-stage network. This method aims to reduce false alarms by considering the physical size of prohibited items. They also apply a method called "Hard-negative-example selection" to reduce the low performance caused by the imbalance between positive and negative examples. In the developed solution, they conduct experiments on SIXray and OPIXray datasets using the Feature Pyramid Network method and the Faster R-CNN method, along with the physical size and hard-negative selection mechanism [4]. Shao, Liu et al. separate prohibited objects from background images by highlighting the problem of object overlapping in X-ray scanning. In this method, they obtain features using Cross Stage Partial Darknet53 (CSPDarknet53), spatial pyramid pooling, and yolov4-tiny networks, and then separate foreground and background [5].

As revealed in these studies, the performance of deep learning models varies depending on network architectures and applications, and each network may exhibit different individual performances in different scenarios. In this regard, instead of designing a new network architecture in this research, we propose applying a method that takes advantage of using multiple deep learning networks. Ensemble learning is the general expression of approaches that aim to create a better model with improved generalization capability by bringing together a set of learners. The cumulative decision-making of multiple models on a problem highlights their strengths while compensating for their weaknesses, leading to enhanced performance. Each model is trained with different algorithms, parameters, or datasets to provide different perspectives on the problem. This way, the ensemble model becomes more reliable, producing consistent results while revealing relationships in a broader pattern space [6, 7]. Ensemble learning methods may involve various techniques for combining predictions, such as majority voting, weighting, or using an optimized machine learning model on the predictions. These methods target different objectives, including ensuring learning diversity, statistical stability, minimizing errors, and improving generalization performance[7, 8]. Ensemble approaches are widely used in the processing of X-ray images, particularly in the medical field [9-11] and biochemistry [12, 13] and physics [14, 15]. Nevertheless, the utilization of ensemble techniques on X-ray images is restricted within the realm of security. Kolte et al. use an architecture called Skip-GANomaly to overcome the problem of limited data in X-ray security applications and design an updated version using a UNet++ style generator. Then, they combine these two architectures using an ensemble method. It is reported that the ensemble method learns better features to distinguish the abnormal class from the normal class compared to individual architectures [16]. Kong et al. propose an approach that benefits from a classifier ensemble using multi-modal information from X-ray images of a single-view object. They use deep neural networks to learn a good representation for each method used to train the base classifiers. To achieve high overall classification performance, they estimate the reliabilities of the base classifiers by considering natural properties of an object in an X-ray image, such as color and shape. They perform tests on a dataset with 15 classes to evaluate the method's competitive performance [17]. Zhou et al. develop an adaptive weighted ensemble model for carotid ultrasound image segmentation, bringing together the advantages of different CNN models. During the joint training of ensemble networks, model weights and sample weights are combined to improve segmentation performance. The method evaluates three different UNet++ models (ResNet152, DenseNet169, and VGG19) on carotid ultrasound images and achieves higher accuracy compared to other methods [18]. Ahmad et al. propose an ensemble-based classification network for classifying baggage X-ray images. The method utilizes joint learning of a deep CNN combined with a Principal Component Analysis (PCA)-based Support Vector Machine (SVM) classifier. The suggested method exhibits high performance in classifying baggage X-ray images [19]. The summarised studies are listed in Table 1 with their salient features. These studies demonstrate that the use of ensemble based deep learning networks can enhance the classification accuracy in X-ray security images compared to a single network.

It is seen that deep learning models have high success in subjects related to X-Ray images. In the methods and datasets available in the literature, object detection (gun, scissors, knife, etc.) is generally performed. Since the relevant datasets consist of images of these objects, their functions are limited to the detection of these objects. In very few of these studies, anomalies etc. are detected. In this study, a unique dataset was created by embedding the electronic circuits potentially belonging to explosives in laptops. The dataset contains different combinations of laptops and circuits. In this way, anomaly detection of electronic circuits can also be provided.

This study aims to efficiently differentiate a series of laptops, some containing a foreign circuit and others normal, in X-ray security image analysis using deep learning algorithms to determine whether they contain hazardous substances. For this purpose, an ensemble methodology is adopted to achieve more accurate and reliable detection of prohibited substances. In the proposed approach, individual outputs of each model are combined with an optimized machine learning algorithm which is a meta-learner. The individual performances of a total of 11 different deep learning models are compared and evaluated against the results of the community approach. The goal of using ensemble learning is to achieve more effective results in prohibited substance detection and provide a more efficient solution for security applications.

The project contributes to both security and deep learning literature in the following aspects.

- Detecting an explosive circuit hidden inside a laptop is a challenging problem for both human operators and deep learning models. Current X-ray devices do not have such detection software/hardware.
- There is a gap in the intersection between deep learning models and X-ray image analysis, especially the topic of X-ray image analysis as it relates to security.
- The existence of such a problem is not mentioned in the scientific literature.
- With this project, the problem is treated as a classification problem and a dataset is created to train the classification methods.
- In addition, an ensemble system with higher accuracy is developed.

2. MATERIALS AND METHODS

2.1. Ensemble Learning

Ensemble learning is a machine learning approach that aims to combine multiple individual models or learners to create a model with higher generalization capability. The fundamental idea behind ensemble learning is that relying on the predictions of various individual learners makes the final prediction more stable, reliable, and generalized compared to a single model. The main ensemble learning approaches can be listed as follows [7, 20]:

Table 1. Related research in the literature

Reference	Task	Methods	Notes
Akçay et al. [1]	Object Classification	Transfer learning using convolutional neural networks	Two class (gun/no gun) handgun detection problem, 98.92% detection accuracy.
Benedykciuk et al. [2]	Object Classification Material Detection	Multi-scale convolutional neural network	The method classify the materials into six groups: background, light organic, heavy organic, light metals, heavy metals and impenetrable. 95.5% detection accuracy.
Miao et al. [3]	Object Classification Threat Detection	The class-balanced hierarchical refinement is applied to ResNet, Inception, DenseNet.	Securty Inspection Xray dataset is presented, it consists of 1,059,231 X-ray images, in which 6 classes of 8,929 prohibited items.
Chang et al. [4]	Object Classification Threat Detection	Faster R-CNN	The proposed method consists of convolutional feature maps, the reconstructed feature maps, binary masks. The physical size constraint formulated as a regularization term during the process of training the proposed detection network.
Shao et al. [5]	Object Classification Threat Detection	Foreground and background separation, YOLOv4	On the GDTIPXray, OPIXray, SIXray datasets, higher success was achieved than the other methods compared.
Zhao et al. [6]	Object Classification	Ensemble Learning, BoostForest, RandomForest, Extra-Trees, XGBoost, LightGBM, GBDT-PL	The research compares the performances of 7 different ensemble approaches on 30 different datasets.
Nasser and Akhloufi [9]	Image Classification Disease Detection	Deep Learning Models, Ensemble Learning	The review article summarizes the approaches and data sets used to diagnose chest disease with xray images.
Radak et al. [10]	Image Classification Disease Detection	Machine Learning, Deep Learning Models,	The review article summarizes the approaches and data sets used to diagnose breast cancer with medical images.
Khan et al. [11]	Image Classification Disease Detection	Machine Learning, Deep Learning Models, Ensemble Methods	The review article summarizes the approaches and data sets used to diagnose chest disease with xray images.
Wang et al. [12]	Classification and Segmentation	DenseNet, ResNet, Random Forest, CNN, Deep Neural Networks, xgBoost	Six feature extraction methods are integrated into proposed deep learning method respectively to form six baseline models. The weighted voting strategy is used to integrate the results from six different classifiers.
Putin et al. [13]	Prediction	Deep Neural Networks, Stacking Ensemble Model	The best performing DNN in the ensemble demonstrated 81.5% accuracy, while the entire ensemble achieved 83.5% accuracy.
Xie and Marsili [14]	Image Classification	Ensemble of Deep Belief Networks	Random energy model applied to the deep learning.
Hoffmann et al. [15]	Prediction	Ensemble of Deep Neural Networks	The paper applies the deep learning hybrid approach to measurement data from a real specimen of an asphere.
Kolte et al. [16]	Object classification Threat detection	GAN based ensembles	Proposed ensemble-based architecture achieved a 75.3% AUC on the SIXray dataset.
Kong et al. [17]	Object Recognition	Deep Neural Networks Based Ensemble Learning	The research reports the comparative results using a dataset with 15 classes.
Zhou et al. [18]	Image Segmentation	CNN, Adaptively weighted ensemble algorithm	Training multiple networks in the ensemble algorithm costs greater computational resources than a single network.
Ahmed et al. [19]	Object classification Threat detection	Deep Neural Networks, PCA, SVM	CNN-based classification models are hybridized with classical machine learning models.

1) **Bagging**
   Data sets are randomly created, and learners are trained in parallel. The models' predictions are combined using majority voting.
2) **Boosting**
   It is a method where each learner is trained to compensate for the errors of the previous learner. Learners are trained sequentially.
3) **Stacking**
   A machine learning model learns from the outputs of individual learners. This meta-model combines the outputs of individual models to make the final prediction.

Commonly used ensemble techniques include bagging, boosting, and stacking, as shown in Figure 1 [21].

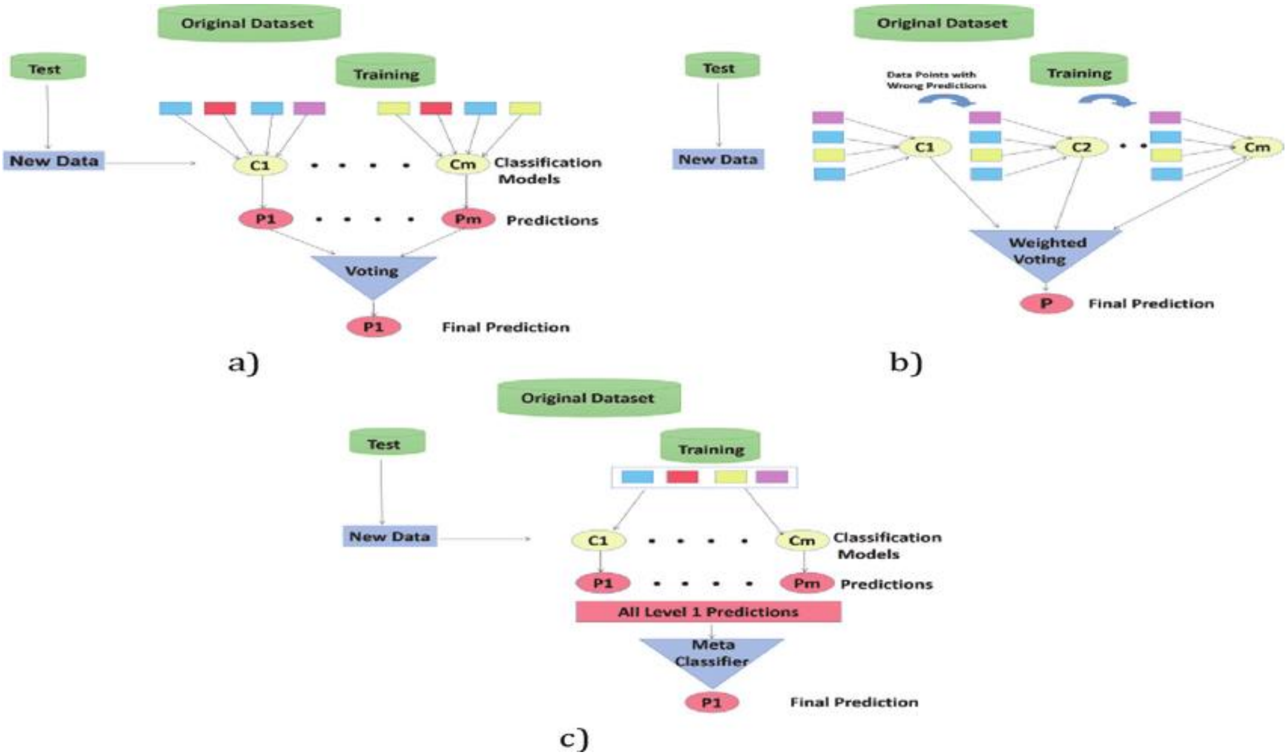


Figure 1. Ensemble techniques a) Bagging, b) Boosting, c) Stacking.

2.2. Deep Learning Models

1) SuffleNet

ShuffleNet is designed by Zhang et al. in 2018, specifically for mobile devices with limited computational capability. The cost is reduced with operations such as point group convolution and channel shuffling. The group convolution method, which is first used with AlexNet, is introduced as a new method in ShuffleNet architecture by using it together with the shuffling process [22].

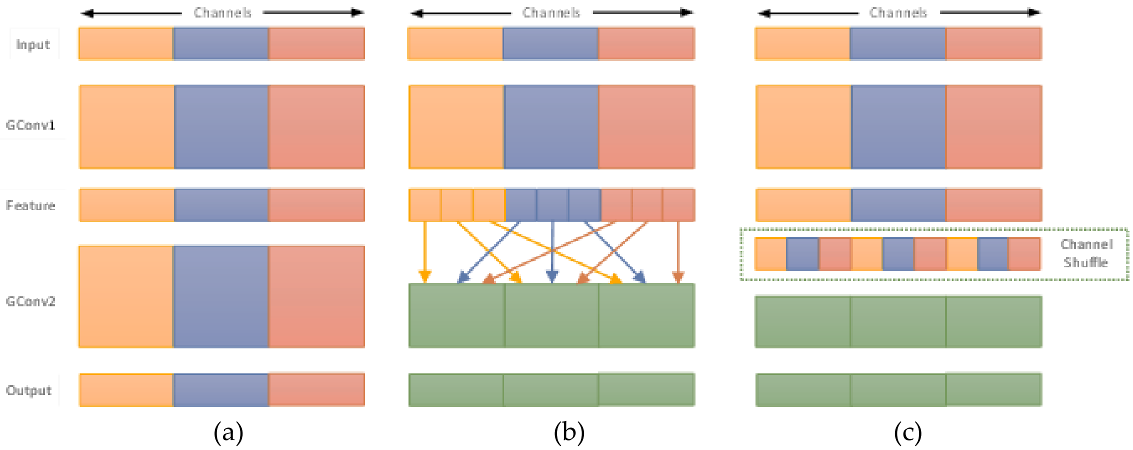
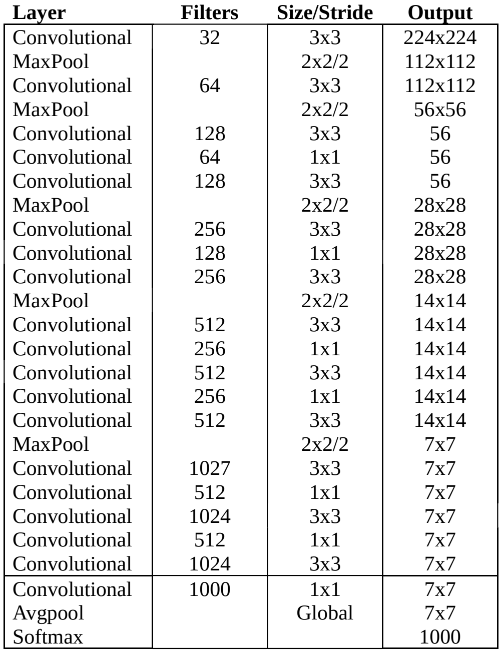


Figure 2. The channel diagrams for convolution a) Group convolution with no cross-talking b) Group convolution with cross talking c) Shuffled channels.

As shown in Figure 2(a), the input value inside group convolution is processed independently on different graphics processing units through convolution and batch normalization. In this case, only the filters in the relevant group affect the output result. However, kernels from a different group than the relevant group may have a significant impact on some outputs. Therefore, in some methods, the groups in the relevant outputs are applied to the next convolution process. As shown in Figure 2(b), each sub-group is mixed in a way to contribute to a different group, which improves performance. Instead of this, in Figure 2(c), the mixing process is performed using matrix transpose and flattening operations. This way, compared to Figure 2(b), a less costly mixing process is achieved. The most significant advantage of these processes in ShuffleNet is the ability to achieve similar classification performance with much less complexity in models.

2) DarkNet

DarkNets are deep learning architectures designed for single-shot tasks, incorporating YOLOv2 and YOLOv3 in their backbone structures. There are two DarkNet models available: DarkNet-19 [23] and DarkNet-53 [24], comprising 19 and 53 convolution layers, respectively. DarkNet-53, in addition to its extensive convolution layers, also employs residual connections to address degradation issues. The configuration of layers for DarkNet-19 are illustrated in Figure 3.



Layer	Filters	Size/Stride	Output
Convolutional	32	3x3	224x224
MaxPool		2x2/2	112x112
Convolutional	64	3x3	112x112
MaxPool		2x2/2	56x56
Convolutional	128	3x3	56
Convolutional	64	1x1	56
Convolutional	128	3x3	56
MaxPool		2x2/2	28x28
Convolutional	256	3x3	28x28
Convolutional	128	1x1	28x28
Convolutional	256	3x3	28x28
MaxPool		2x2/2	14x14
Convolutional	512	3x3	14x14
Convolutional	256	1x1	14x14
Convolutional	512	3x3	14x14
Convolutional	256	1x1	14x14
Convolutional	512	3x3	14x14
MaxPool		2x2/2	7x7
Convolutional	1027	3x3	7x7
Convolutional	512	1x1	7x7
Convolutional	1024	3x3	7x7
Convolutional	512	1x1	7x7
Convolutional	1024	3x3	7x7
Convolutional	1000	1x1	7x7
Avgpool		Global	7x7
Softmax			1000

Figure 3. The layers of DarkNet-19

3) Inception

The first CNN model that increased the network's width using modules called "Inception" is introduced by Szegedy et al. in 2015. The inception architecture aims to approximately mimic the optimal local sparse structure within a convolutional network. It performs 1x1, 3x3, and 5x5 convolutions, as well as 3x3 maximum pooling, in parallel within the convolutional layers. To reduce computational complexity, 1x1 convolution layers are added before the parallel Inception convolutional layers which are shown in Figure 4 [25].

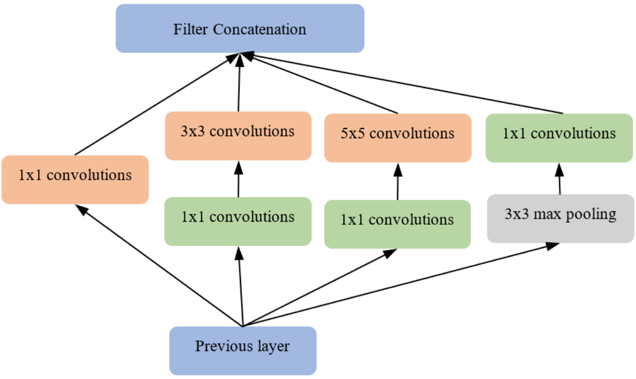


Figure 4. Inception convolutional layers

4) DenseNet

DenseNet, is an architecture introduced by Gao Huang et al. in 2017. It is specifically designed to tackle the vanishing gradient problem and enhance feature reuse within deep neural networks. This is achieved through the utilization of dense connections, where each layer is directly connected to every other layer in a feed-forward manner. Within the Dense Blocks depicted in Figure 5, each layer is linked with corresponding feature map sizes. Every layer not only forwards its feature maps to subsequent layers but also receives supplementary inputs from the preceding layers, ensuring the preservation of an uninterrupted information flow. These attributes contribute to making DenseNet a robust and effective type of CNN, showcasing outstanding performance across diverse computer vision tasks [26].

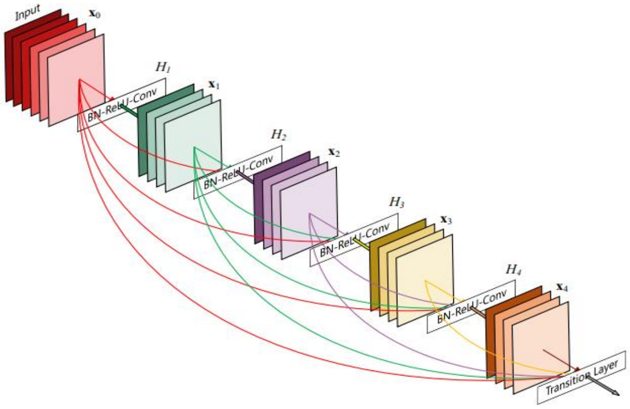


Figure 5. Dense Blocks

5) Xception

Xception is a deep learning model introduced by François Chollet in 2016. The name "Xception" means Extreme Inception, indicating that it is an extension of the Inception architecture. Instead of spatial filters, depthwise separable convolutions are developed to separate spatial and channel-wise filtering, aiming to enhance the performance of CNNs. The convolution process, as shown in Figure 6, consists of deep convolution where each channel evolves independently and pointwise convolution is applied for inter-channel interactions using a 1x1 convolution [27].

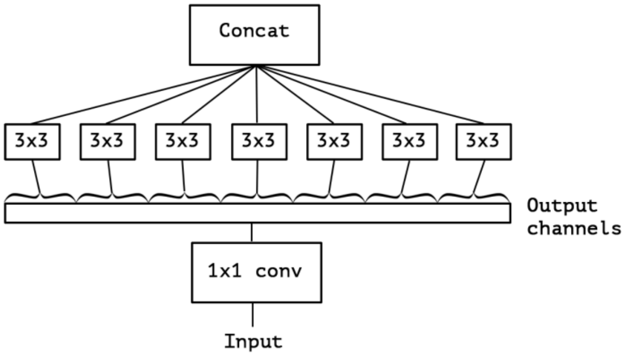


Figure 6. Xception Module

2.3. Dataset

The aim of the research is to use deep learning methods to detect circuits hidden in laptops. In this context, a dataset is needed to train and test deep learning methods. For this reason, the data set was created by us. Arduino uno, nano, bluetooth boards that are easily available in the market were preferred as circuits. Since many different laptops were needed, second-hand laptops were purchased from the market. X-ray images of 60 laptops in different configurations were obtained by using the X-ray devices at the airport with the permission of the Konya Airport Administrative Authority. X-ray images of the laptops taken from different perspectives are given in Figure 7. The areas enclosed in red rectangles in images contain hidden circuits that do not belong to the computer motherboard.

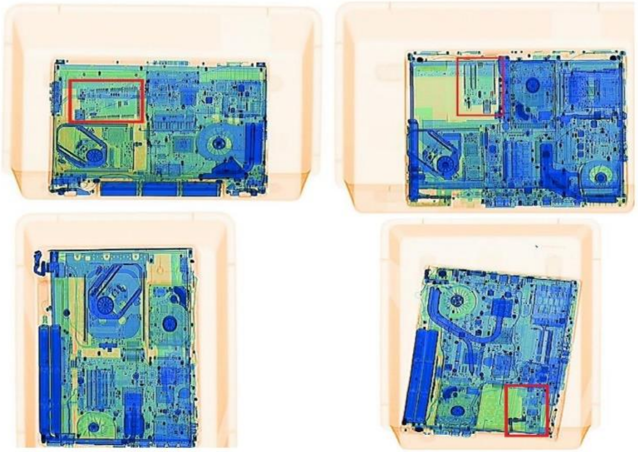


Figure 7. X-ray images of the laptops

In a period of 5 months, a total of 6395 X-ray images were taken. Of these, 2545 have hidden circuitry and 3850 do not. In order to keep the data balanced, the number of images without circuitry was reduced to 2549 and a total of 5094 X-ray images were used in the experiments. Since the problem is considered as a classification problem, it is necessary to have the labels of the images during the training and testing process with deep learning methods. The 5094 X-ray images were labeled as normal or abnormal and stored in different folders. The background and object images were segmented and the clean image shown in Figure 8 was obtained. Since deep learning architectures have different input sizes, all images were resized for each deep learning algorithm to match the input size of the images. Since the number of images was sufficient to train the deep learning models, no data augmentation was performed.

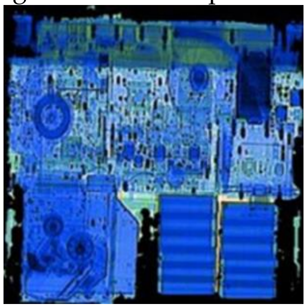


Figure 8. Preprocessed image sample

2.4. Proposed Method

The dataset consists of a total of X-ray images of 5094 laptops, including 2545 images of laptops with foreign circuits hidden inside and 2549 images of normal laptops without any modifications. A total of 11 deep learning algorithms were trained with Adaptive Moment Estimation optimizer. Adam is more efficient in situations where gradient-based optimization algorithms have problems such as local minima and slow training speed. Since Adam is a more efficient and faster optimization algorithm, it is often the default optimization algorithm in deep learning models. In the experimental study, 100 epochs and 3000 iterations were sufficient to determine whether the deep learning models trained or not. The batchsize value of all models was set to 32 depending on the hardware used. The 10-fold cross-validation method was applied, and the training and test sets for each fold were recorded.

To ensure the effectiveness of the ensemble learning approach, algorithms with validation accuracy above 70% were preferred, as very poorly performing learners could negatively impact the ensemble result. The individual decisions of learners were combined using an optimized machine learning method. The choice and optimization of the machine learning algorithm (meta-learner) were performed using the fitcauto function with options for all learners and all OptimizeHyperparameters in Matlab software. fitcauto automatically tries different classification models with various settings. It uses Bayesian optimization to select the best model and cross-validation to evaluate their performance, ultimately determining the best model for predictions. The applied method is illustrated in Figure 9.

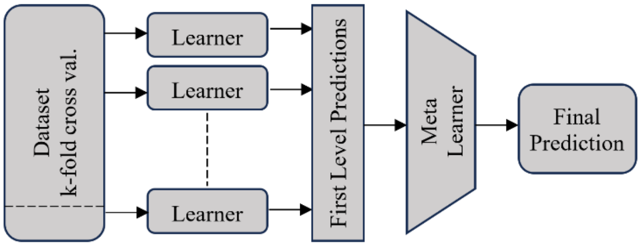


Figure 9. Proposed ensemble method

This ensemble learning model is based on a meta-model that learns from the output of individual learners. Deep learning models with high classification accuracy are used as individual learners. The predictions of the deep learning models provide the first level predictions. Based on these predictions, a machine learning model is optimized on the problem as a meta-learner. The fitauto function includes the following classifiers: Discriminant analysis classifier, Ensemble classification model, Kernel classification model, k-nearest neighbor model, Linear classification model, Naive Bayes classifier, Neural network classifier, Support vector machine classifier, Binary decision classification tree. The Linear Classification model was used as a meta-learner since it provides higher correct classification success than the others.

3. RESULTS

In the study, X-ray images containing normal and abnormal classes were trained on 11 deep learning models listed in Table 2. The table presents the validation accuracy values of each model during the training process for each fold.

Table 2. Individual classification accuracy of deep learning algorithms

					k-fold						
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Avg.**
efficientnetb0	59.49	63.35	59.05	57.38	58.55	60.87	59.45	60.65	58.28	56.47	59.35
resnet18	63.27	65.16	63.93	67.49	64.65	63.42	64.46	64.58	67.44	63.88	64.83
resnet50	59.64	60.51	64.87	60.36	61.02	61.24	61.85	60.29	67.08	63.81	62.07
resnet101	64.29	62.33	61.96	63.05	59.71	62.18	61.77	62.18	63.88	59.96	62.13
darknet19	77.60	77.96	77.02	75.64	79.20	78.55	80.31	75.35	79.36	76.38	**77.74**
darknet53	68.58	67.42	61.82	65.53	62.76	64.58	68.39	69.53	66.86	68.60	66.41
mobilenetv2	61.53	59.71	61.38	62.25	60.51	60.00	61.63	60.22	62.21	61.56	61.10
shufflenet	**82.76**	**83.13**	**83.64**	**82.25**	**80.80**	**85.24**	80.67	83.56	83.28	**83.94**	**82.93**
inceptionv3	82.04	80.73	77.45	81.38	80.15	84.51	**85.10**	**84.15**	**83.94**	82.41	**82.19**
densenet201	74.84	78.40	79.64	76.73	**80.80**	78.76	79.43	79.85	76.67	76.74	**78.19**
xception	71.93	74.04	70.04	69.60	68.80	70.04	69.91	71.64	75.73	70.42	**71.21**

Using 10-fold cross validation, the results on the validation set obtained during the training process are presented with averages. The model with the lowest performance is efficientnetb0. The average accuracy is 59.35%. The highest performing model is shufflenet. The average accuracy is 82.93%. Among the other models, resnet18, resnet50, darknet19, darknet53, inceptionv3, and densenet201 perform quite competitively. On the other hand, models such as resnet101, mobilenetv2, and xception perform slightly worse. They have accuracy rates in the low 60% range.

The problem of very low-performing learners can negatively impact the ensemble decision since they do not learn the problem effectively. Therefore, a threshold value of 70% was set based on the individual results of deep learning models, and models belonging to DarkNet19, Shufflenet, InceptionV3, DenseNet201, and Xception architectures were selected for use in the ensemble approach. Table 3 provides the individual accuracy of the correct classification performances of the top learners and the accuracy of the correct classification performances of the predictions produced by the meta-learner on the test set. Table 4 provides the evaluation metrics as an average of 10 folds.

Table 3. The classification accuracy of top models and meta-learner on test set

						k-fold						
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Avg.**	**σ**
darknet19	80.35	77.60	81.53	78.19	77.60	78.98	**81.93**	71.32	81.14	80.55	78.92	2.95
shufflenet	**83.10**	**83.89**	**86.05**	**81.14**	**81.93**	**88.80**	80.55	**82.51**	**84.09**	**83.50**	**83.56**	2.30
inceptionv3	81.93	77.60	78.19	80.16	80.94	84.68	84.68	79.96	83.50	81.34	81.30	2.34
densenet201	74.66	77.01	77.41	72.89	82.12	78.59	78.98	78.98	75.83	80.55	77.70	2.61
xception	75.25	70.33	68.57	68.57	69.35	70.73	71.91	69.16	75.64	73.08	71.26	2.50
ensemble	**86.64**	**83.89**	**86.64**	**82.12**	**86.25**	87.62	**88.02**	**85.46**	**85.66**	**87.43**	**85.97**	**1.72**

The ensemble model is the highest with an average accuracy of 85.97%. Darknet19, shufflenet and inceptionv3 models also performed quite well, but no deep learning model achieved a higher percentage accuracy than the ensemble model. This indicates that an ensemble model performs better than the others on the test set. Although densenet201 and xception have a lower average accuracy percentage, they still perform reasonably well. Furthermore, the standard deviation (σ) value shows how consistent the performance of each model is. The Ensemble model has the lowest standard deviation, indicating that its performance is more consistent.

Table 4. Evaluation metrics of top models and meta-learner on test set

	Accuracy	Specificity	Precision	Recall	F-Measure	G-mean
darknet19	78.92	80.75	80.27	76.98	78.48	78.76
shufflenet	83.56	85.33	84.93	81.99	83.23	83.53
inceptionv3	81.30	84.44	83.69	77.96	80.62	81.06
densenet201	77.70	78.81	78.43	76.58	77.41	77.63
xception	71.26	71.81	71.75	70.84	71.00	71.08
ensemble	**85.97**	**84.54**	**84.98**	**87.37**	**86.12**	**85.92**

Table 4 presents the accuracy, specificity, precision, precision, recall, F-Measure and G-mean metrics for assessing model performance [28]. Proposed ensemble learning achieved better correct classification performance than the individual highest performing ShuffleNet model in 9 out of 10 folds. When looking at the average accuracy values, the ensemble approach showed 14.72% higher performance compared to the worst model and 2.42% higher performance compared to the best model. Additionally, it provided a more stable prediction output according to the standard deviation (σ). Ensemble method shows the highest accuracy, precision, and F-measure, indicating superior performance compared to individual models. Table 5 shows the confusion matrix of the average of the 10-fold cross-validation results of the ensemble model.

Table 5. Confusion matrix of the ensemble model

		Normal	Abnormal
True Class	Normal	219	39
	Abnormal	32	219
		Predicted Class	

4. CONCLUSION

In environments where security controls are tightened, X-ray devices are commonly used with human operators. X-ray systems become inefficient in high-traffic areas or when expertise is required for the recognition of a threat object. Applications such as classification or anomaly detection on X-ray images using deep learning have been widely used, especially in the medical field, but have not received sufficient attention for security purposes. In this research, deep learning methods were examined on laptop X-ray images, some of which contain threat objects, in terms of both individual correct classification performance and proposing an ensemble approach instead of designing a new model architecture. The ensemble approach highlighted the advantages of existing models, resulting in a more stable output with 1.72 lower standard deviation and 87.43% higher correct classification performance compared to all individual models.

There is a limited number of studies in the literature that focus on the ensemble approach for forbidden object detection in X-ray images. In addition to developing new architectures on this topic, different ensemble methods can also be applied.

Declaration of Ethical Standards

The authors declare that they have carried out this completely original study by adhering to all ethical rules including authorship, citation and data reporting.

Credit Authorship Contribution Statement

Engin EŞME: Methodology, Conceptualization, Resources, Investigation, Writing.
Mustafa Servet KIRAN: Methodology, Conceptualization, Resources, Investigation, Writing, Supervision.

Declaration of Competing Interest

The authors declared that they have no conflict of interest.

Funding / Acknowledgment

This study has been supported by The Scientific and Technological Research Council of Türkiye under Grant 122E024. The authors thank the council for the institutional supports.

Data Availability

Data supporting the findings of this study can be obtained from the corresponding author with reasonable requests to assist in scientific studies.

REFERENCES

[1] S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in 2016 IEEE International Conference on Image Processing (ICIP), 25-28 Sept. 2016 2016, pp. 1057-1061, doi: 10.1109/ICIP.2016.7532519.

[2] E. Benedykciuk, M. Denkowski, and K. Dmitruk, "Material classification in X-ray images based on multi-scale CNN," Signal, Image and Video Processing, vol. 15, no. 6, pp. 1285-1293, 2021/09/01 2021, doi: 10.1007/s11760-021-01859-9.

[3] C. Miao et al., "Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2019, pp. 2119-2128.

[4] A. Chang, Y. Zhang, S. Zhang, L. Zhong, and L. Zhang, "Detecting prohibited objects with physical size constraint from cluttered X-ray baggage images," Knowledge-Based Systems, vol. 237, p. 107916, 2022.

[5] F. Shao, J. Liu, P. Wu, Z. Yang, and Z. Wu, "Exploiting foreground and background separation for prohibited item detection in overlapping X-Ray images," Pattern Recognition, vol. 122, p. 108261, 2022.

[6] C. Zhao et al., "BoostTree and BoostForest for ensemble learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[7] Z.-H. Zhou, Ensemble learning. Springer, 2021.

[8] I. H. Witten, E. Frank, and M. A. Hall, "Data mining: Practical machine learning tools and techniques," ed: Morgan Kaufmann, 2016.

[9] A. Ait Nasser and M. A. Akhloufi, "A review of recent advances in deep learning models for chest disease detection using radiography," Diagnostics, vol. 13, no. 1, p. 159, 2023.

[10] M. Radak, H. Y. Lafta, and H. Fallahi, "Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies," Journal of Cancer Research and Clinical Oncology, pp. 1-19, 2023.

[11] W. Khan, N. Zaki, and L. Ali, "Intelligent pneumonia identification from chest x-rays: A systematic literature review," IEEE Access, vol. 9, pp. 51747-51771, 2021.

[12] H. Wang, H. Li, W. Gao, and J. Xie, "PrUb-EL: A hybrid framework based on deep learning for identifying ubiquitination sites in Arabidopsis thaliana using ensemble learning strategy," Analytical Biochemistry, vol. 658, p. 114935, 2022.

[13] E. Putin et al., "Deep biomarkers of human aging: application of deep neural networks to biomarker development," Aging (Albany NY), vol. 8, no. 5, p. 1021, 2016.

[14] P. Peng, C. Marceau, and D. M. Villeneuve, "Attosecond imaging of molecules using high harmonic spectroscopy," Nature Reviews Physics, vol. 1, no. 2, pp. 144-155, 2019.

[15] R. Xie and M. Marsili, "A random energy approach to deep learning," Journal of Statistical Mechanics: Theory and Experiment, vol. 2022, no. 7, p. 073404, 2022.

[16] S. Kolte, N. Bhowmik, and Dhiraj, "Threat Object-based anomaly detection in X-ray images using GAN-based ensembles," Neural Computing and Applications, pp. 1-16, 2022.

[17] Q. Kong, N. Akira, B. Tong, Y. Watanabe, D. Matsubara, and T. Murakami, "Multimodal Deep Neural Networks Based Ensemble Learning for X-Ray Object Recognition," 2019: Springer, pp. 523-538.

[18] R. Zhou, F. Wang, X. Fang, A. Fenster, and H. Gan, "An adaptively weighted ensemble of multiple CNNs for carotid ultrasound image segmentation," Biomedical Signal Processing and Control, vol. 83, p. 104673, 2023.

[19] A. H. Ahmed, M. Al Radi, and N. Werghi, "An Ensemble Learning Method Based on Deep Neural and Pca-Based Svm Network for Baggage Threat and Smoke Recognition," in Proc. Advances in Science and Engineering Technology International Conferences, Dubai, 2023, pp. 1-6.

[20] A. Kumar and J. Mayank, "Ensemble learning for AI developers," BApress: Berkeley, CA, USA, 2020.

[21] E. O. Kiyak, "Data Mining and Machine Learning for Software Engineering," Data Mining-Methods, Applications and Systems, 2020.

[22] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in Proc. IEEE conference on computer vision and pattern recognition, 2018, pp. 6848-6856.

[23] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in Proc. IEEE conference on computer vision and pattern recognition, 2017, pp. 7263-7271.

[24] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[25] C. Szegedy et al., "Going deeper with convolutions," in Proc. IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks,", in Proc. IEEE conference on computer vision and pattern recognition, 2017, pp. 4700-4708.

[27] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proc. IEEE conference on computer vision and pattern recognition, 2017, pp. 1251-1258.

[28] Ž. Vujović, "Classification model evaluation metrics," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 599-606, 2021.