Performance Evaluation of the Convolutional Neural Networks for Object Identification Using RGB and Binary Images

Jasim Mohammed Dahr and Alaa Sahl Gaafar
Directorate of Education in Basrah, Basrah, Iraq.
E-mail: Jmd20586@gmail.com, alaasy.2040@gmail.com

Keywords: machine learning, CNN, RGB colour image, binary image, accuracy, speed

Received: July 6, 2024

Convolutional Neural Network (CNN) is a topmost deep learning technique because of its capability to learn features autonomously with domain-specific images similar to the classical machine learning approaches. One common tactic for training CNN architecture is to transfer learned knowledge of a pre-trained network used to perform one task into a fresh task. Accurate object detection remains an active area of research. The use of traditional classification methods is widespread, but limited by time-ineffectiveness, and subjectiveness. The advent of CNN attempts to estimate and extract features inside images for the improved precision of image classification. This paper evaluates the two CNN models (CNN-1 and CNN-2) for object identification with two input image formats. The CNN structure for the binary image was composed of (4, 4, 16) outputs were flattened into vector shape of 1024 before going through Dense layers. While, CNN structure for the RGB image was composed of (4, 4, 64) outputs flattened into vector shape of 256 before going through Dense layers. The CNNs were experimented using standard image datasets: binary images (MINST) and RGB images (CIFAR-10) with 50,000 and 10,000 images samples respectively. The training and validation of the proposed CNN-based image identification systems were carried out on Google Colaboratory simulator. The outcomes showed that, the image identification system model with RGB images outperformed binary images-based system in terms of average duration of 14.20ms to 29.00ms. Conversely, the image identification system modelled with CNN and binary images was superior over the RGB images for loss function of 0.00320 (0.3611%) and 0.8830 (99.64%) because of simpler network structure and features map. The same trend was observed for the accuracy metric, CNN model with binary images achieved an accuracy of 0.9910 (58.24%) more than RGB images at 0.7106 (41.76%). These results were comparable to existing studies using the average duration and accuracy, which are statistically significant using Wilcoxon Test of p-value of 0.007 < significant value (0.05). Therefore, CNNs could be applied to high-precision objects identification tasks like surveillance, medical diagnosis, agriculture, security, and road transportation systems.

Povzetek: Študija ocenjuje zmogljivost konvolucijskih nevronskih mrež (CNN) za prepoznavanje objektov na RGB in binarnih slikah. Model z binarnimi slikami je izkazal višjo natančnost (99,10 %) in nižje izgube, medtem ko je model z RGB slikami omogočal hitrejšo obdelavo. Predlagani modeli so primerni za aplikacije, kot so medicinska diagnostika, nadzor in transportni sistemi.

1 Intorduction

Machine learning (ML) approaches are applied to causal inference improvement with supervised, unsupervised, and models to assist in accurate effects estimation [1]. Deep Learning (DL) is a field in ML that employs representation learning through expressing the input data in multiple levels of simpler representations. A CNN is a special type of DL having shown unprecedented success in image-related problems including: image classification, image semantic segmentation, object detection in images, and others [2]. The input data in a CNN is processed in a grid-like topology [3].

With the availability of several annotated images, DL approaches have demonstrated their superiority over the classical ML approaches. The CNN architecture is a popular DL approaches with superior achievements in the medical imaging domain. The primary success of CNN is due to its ability to learn features automatically from domain-specific images, unlike the classical machine learning methods. The popular strategy for training CNN architecture is to transfer learned knowledge from a pre-trained network that fulfilled one task into a new task [4], [5].

Traditional classification methods (like naked-eye observation and laboratory tests) have many limitations, such as being time consuming and subjective. Presently, DL methods, especially those based on CNN, have gained widespread application in plant disease classification [6], [7].

The success of CNNs in computer vision is being extended for accurate object detection-based on DL like agricultural robots whose outcomes exceeds traditional detection methods-based on manually designed features like histogram of oriented gradients, and scale invariant feature transform [8]. Precise identification of farmland obstacles is an important environmental-perception task for agricultural vehicles. The orchard environment is a complex and unstructured environment making obstacles detection complex, less accurate and effective [8].

The performance of the image classification algorithms crucially relies on the features used to feed them. Each layer has many neurons that respond to different combinations of inputs from the previous layers [9]. CNN is a well-known technique for classifying objects in distinctive specialties as it has an extraordinary capability of discovering complex patterns [10]. Thus, detecting objects accurately is necessary and urgent. Traditional classification methods have several limitations such as being time consuming and subjective [6]. However, recently, a CNN founded on deep learning hold prospects in its capability to estimate and extract features to enhance the precision of object identification tasks [11], [12]. In this paper, CNN models are used to identify patterns in different image-types and their subsequent identification. The specific contributions are:

i. To preprocess binary and RGB colour images from standard databases and repositories.
ii. To build two CNNs for operating on two-image input datasets.
iii. To evaluate the performance of the built CNNs models for object identification tasks.

The remaining sections of the paper include: section 2 is the literature review, section 3 is the research methodology, section four is the results, and conclusion is presented in section 5.

2 Literature review

The CNN begun in 1962 when Hubel and Wiesel analyzed the structure of visual cortex in the cat brain and found that the biological visual information is transferred through multi-layer receptive domain [13]. They tried to construct similar algorithms to make the machine recognizes images. The construction and application of CNN developed rapidly after 1980. Basically, the authors in [14] developed VGG16 (13 convolutional layers and three fully connection layers) and VGG19 (16 convolutional layers and three fully connection layers). MobileNet is a lightweight CNN model which is introduced by Google in the Conference on Computer Vision and Pattern Recognition. This model utilized depth-wise separable convolutions to compress model parameters and improve computing speed [15].

Knowledge is one of the techniques for improving performance based on participation. DL is a subfield of Artificial Intelligence (AI) and uses algorithms that are inspired by the structure and functionality of the brain neurons. DL techniques are primarily used for improving the enforcement of several ML applications [10]. Deep architecture with trainable parameters attempts to build a higher level of abstraction on data, with linear and nonlinear transformation function [16]. DL are mainly utilized in the computer vision field like object detection, classification, localization, abnormally detection from medical images [17]. DL techniques are also popularly used in the Bangla and Manipuri character recognition field [18], [19], [20].

CNN takes aim in optimizing the speed and accuracy of classification tasks [21].

Image classification is the popular application of CNN. The object detection process matches the objects from predefined categories [22]. It is can be used to determine the existence and region of tumors on organs or tissues of medical images. If the target is present, it could be indicated on spatial location. The object in images is marked by a frame (such as a boundary box) with the confidence on the top of boundary box [23]. Object detection can perform many tasks such as lesion location, lesion tracking and image discrimination. The application of object detection in medical images is extremely wide. Semantic segmentation is another algorithm in which computer segments images based on the pixels presented in the images. The semantic refers to the content of the image, and the segmentation means that different objects in the image are segmented based on pixels. In semantic segmentation analysis, each pixel in the image is labeled [24].

The CNN architecture for fine-grained visual categorization was proposed by [25]. The approach reflected human expert success in classifying bird species. First, the architecture calculates an estimate of the pose of the object which is further used to calculate features of local images. In addition, these characteristics are used for classification purposes. The features are determined by applying deep convolutionary nets to patches of the image that the pose locates and normalizes. The outcomes showed that the bird species identification was highly successful, with a substantial increase in the right levels of classification over the previous methods (75% vs. 55-65%).

2.1 Object identification

The principal study was to classify the bird species from the user's picture as input. Transfer learning is the technology used to fine-tune a pretrained model (like AlexNet) in order to obtain superior classification outcomes as in the case of Support Vector Machine [26]. The advanced algorithms give good accuracy in terms of their numerical precision. CNN includes an input image, and assigns the weights and the distinctions to the various aspects of the images, and then distinguish one image from another. The pre-processing required in CNN compared with other classification algorithms is much lower. In primitive methods, filters were usually hand-engineered; on the other hand, CNN has the ability to learn these filters on its own when subjected to enough number of trainings [27].

CNNs' architecture is quite similar to that of the pattern of neuron connectivity in the human brain, in which individual neurons respond only to stimuli in the receptive field. These receptive areas collectively overlap the entire visual area. The initial parameters to be known are the elements that are significant part in the operation of Convolutional Neural Networks. The main components of the CNN model are made of the following: input image, CNN, output label (image class) [28] as shown in Figure 1.

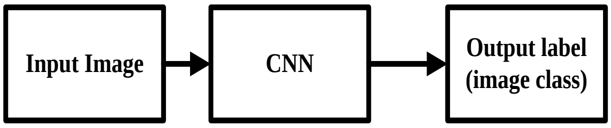


Figure 1: A typical CNN interaction layout of the elements

2.2 Related works

DL methods (like CNN) are widespread in classification tasks such as plant disease classification using agricultural images [6]. Traditional classification methods (such as naked-eye observation and laboratory tests) have several shortfalls including time consumption and subjectivity. A different layout was proposed by [10] for obtaining a unique CNN architecture from scratch, which has manifold advantages over classical ML approaches. This offers a unique ability to consolidate feature extraction and classification altogether using non-linearity in the DL model. The CNN was composed of four layers including: convolutional layer, nonlinear activation layer, pooling layer, and fully connected layer. The evaluation of the model considered the Bangla datasets (that is, cMATERdb and ISI Bangla) and the Manipuri Character dataset (or Mayek27). But, the binary images were considered for the two regional handwritten character identification task [10].

The progression of object detection and semantic segmentation in medical imaging study was introduced by [22]. Both object detection and semantic segmentation algorithms are based on CNN. These are widely applied in various fields of medical imaging study, particularly in the digestive system, respiratory system, endocrine system, cardiovascular system, brain, eye, and breast. These algorithms analyzed multiple images including radiation images (CT, MRI, and PET), pathological images, ultrasound images, and endoscopic images. But, there is no mention of the best performing image format. A study is to examine DL effectiveness for crack detection on images from masonry walls was conducted by [29]. A dataset with photos from masonry structures is produced containing complex backgrounds and various crack types and sizes. Different deep learning networks based on CNNs (such as FCN, FPN, and U-net) are considered and by leveraging the effect of transfer learning crack detection on masonry surfaces is performed on patch level with 95.3% accuracy and on pixel level with 79.6% F1 score. But, other types of surfaces images formats could be experimented.

CNNs provide precise applications and strategy for detecting cellular breakdown in lungs using channels and division methods is proposed [30]. The authors computed CT-Image obtained from cellular fragmentation in patients with computer hemorrhage are dissociated by a digital image-making strategy. The results were similar to the standard features obtained from the ongoing investigation. Therefore, settlement counting techniques can detect the cellular breakdown in the lungs with the middle channels and assembly of medical equipment. This way, clinicians were able to detect the cellular breakdown within the lungs. However, the CNN model holds great potential for a cellular breakdown in early lungs detection with high accuracy.

A study examined DL techniques effectiveness for crack detection on images from masonry walls was conducted by [29]. A dataset with photos from masonry structures is produced containing complex backgrounds and various crack types and sizes. Different CNNs DL networks (such as FCN, FPN, and U-net) were considered by leveraging the effect of transfer learning crack detection on masonry surfaces. It performed on patch level at 95.3% accuracy, and on pixel level with F1 score of 79.6%. In future works, the best performing architectures can be implemented on other surfaces to identify other objects such as doors, ornaments, etc.

The collapse of buildings caused by earthquakes can lead to a large loss of life and property. Rapid assessment of building damage with remote sensing image data can support emergency rescues. To this end, the authors [31] attempted to produce a deep learning network model with strong generalization by adjusting four CNNs for extracting damaged building information and compare their performances. A sample dataset of damaged buildings was constructed by using multiple disaster images retrieved from the xBD dataset. Using satellite and aerial remote sensing data obtained after the 2008 Wenchuan earthquake, authors examined the geographic and data transferability of the deep network model pre-trained on the xBD dataset. The result shows that, the networks pre-trained with samples generated from multiple disaster remote sensing images can extract accurately the collapsed building information from satellite remote sensing data. Among the adjusted CNN models tested, the adjusted DenseNet121 was the most robust with accuracy of 88.9% from 64.3%.

Recently, deep learning methods have been applied in many real scenarios with the development of CNNs. In the study by [32], a camera-based basketball scoring detection (BSD) method with CNN based object detection and frame difference-based motion detection was considered. In the proposed BSD method, the videos of the basketball court are taken as inputs. Thereafter, the real-time object detection, that, you only look once (YOLO) model, is implemented to locate the position of the basket-ball hoop. Then, the motion detection based on frame difference was utilized to detect object motion within the area of the hoop to achieve the basketball scoring condition. The outcomes for Faster RCNN, SSD and YOLO models were 89.26%, 91.30%, and 92.59% respectively. Future work could consider larger datasets, and more efficient detection models for the BSD method.

The facial modality as a fundamental biometric technology has become increasingly important in the field of research. In [11], the authors developed a gender prediction and age estimation system based on CNNs for a face image or a real-time video. The authors created three CNN network models with different architecture (that is, the number of filters, the number of convolution layers, etc.) and validated on IMDB and WIKI datasets. The outcomes showed that, the CNNs greatly improved with accuracy of the recognition (age: 83.97% and 86.20%, gender: 93.56% and 94.49% for WIKI and IMDB datasets). However, other high-performance models can be used future works including SVM.

The specific contributions of authors in terms of objectives, strengths and weaknesses are presented in Table 1. The findings contained in Table 1 revealed that no study undertook a comparative performance analysis of CNNs with both RGB images and Binary Images datasets, which serve as motivation for this paper.

Table 1: Summary of related works

Author(s)	Objectives	Strengths	Weaknesses
Abbas et al. (2021)	Decompose, Transfer and Compose (DeTraC) for chest X-ray image classification.	Detection of COVID-19 cases and acute respiratory syndrome at accuracy of 93.10% and sensitivity of 100.00%.	-Irregular dataset. -Low performance. -Unstandardized dataset.
Lu et al. (2021)	CNN-based plant disease classification.	Faster and better accuracy.	-Insufficient datasets. -Image background instability. -Unstandardized dataset.
Li et al. (2021)	Farmland obstacle detection based on YOLOv3 with MobileNetv2 and Gaussian model.	Minimized running time, improved features extraction and detection rate: F1 of 91.76%.	-Low image size of 416x416. -Unstandardized dataset.
Hazra et al. (2020)	CNN architecture formulation for Bangla datasets detection.	Optimized features extraction and classification. Use of nonlinearity in DL.	Binary images.
R. Yang & Yu (2021)	CNN-based medical images detection.	Segmentation and object detection.	-Images required labeling by experienced doctors. -Limited image types.
Yoon et al. (2021)	DCNN detection of apparent and occult scaphoid fractures within radiographic images.	Detection 20 from 22 cases at 90.90% after EfficientNetB3 optimization of image features.	-Low accuracy.
Zhou, Lu, & Pei (2021)	CNN-based detection of cellular breakdown in lungs.	CT images were used to determine hemorrhage of patients at high accuracy.	-CT images were utilized.
Dais et al. (2021)	DL detection of cracks in masonry walls.	It performed on patch level at 95.3% accuracy, and on pixel level with F1 score of 79.6%.	No fine-grain objects detection.
Yang & Zhang (2021)	Building damage images detection after quakes using CNNs.	DenseNet121 CNN model was best at 88.9%.	-Low accuracy. -Dataset's complexity.
Fu et al. (2021)	CNNs based real-time BSD.	The accuracy of Faster RCNN, SSD and YOLO models were 89.26%, 91.30%, and 92.59%.	-Limited datasets.
Benkaddour et al. (2021)	CNNs based gender and age prediction system using facial modality datasets (IMDB and WIKI).	Age of 83.97% and 86.20%, for WIKI, and gender of 93.56% and 94.49% IMDB.	-Low accuracy. -Limited number of networks.

3 Research methodology

3.1 Description of the proposed objection identification models

The paper proposed image-based object identification models using CNN. These are used to detect class of objects based on the features available inside the RGB and binary images.
The CNN models are composed of the input image block, CNN block and the output label block as depicted in Figure 2.

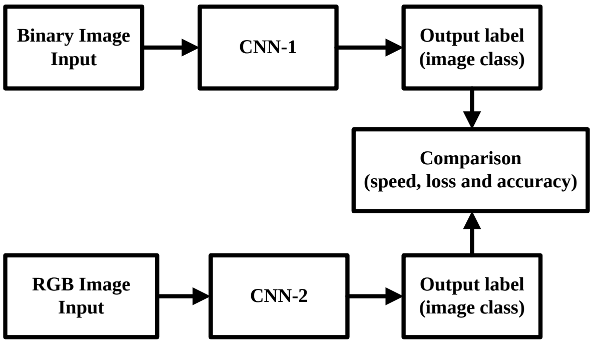

Binary Image Input
CNN-1
Output label (image class)
Comparison (speed, loss and accuracy)
RGB Image Input
CNN-2
Output label (image class)

Figure 2: The CNNs based object detection model for different input data types

From Figure 2, the descriptions of various events taking place at each of the phases of proposed object identification system modelling are provided as follows:

Input image blocks: These are concerned with the collection of images about different objects and their categories dataset from the standard image databases and repositories. Thereafter, the datasets are pre-processed, stemming, tokenization (or removal of noise). The RGB and binary images were selected for this study. This carried out to simplify the adoption of the dataset as input for further processing in the next block.

CNN blocks: These entail training and testing of CNN-1 and CNN-2 models in order to extract and select features. The models are be applied to high dimensional unstructured data after setting the hyperparameters check the relationships within datasets. Again, the attributes are extracted and compiled according through associations and dissociations. The diagrammatic illustration of the details of the CNN models as shown in Figures 3 and 4.

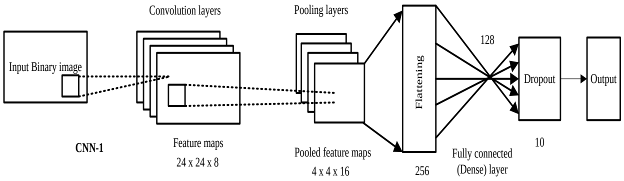


Figure 3: The structure of CNN-1 model with binary image input

From Figure 3, the structure of CNN-1 model for Binary image input is composed of hyperparameters at the input layer, convolution layers, pooling layers, flattening fully connected (dense) layers, and output. In terms of (height, weight, channels) of CNN-1, the convolution layer shape is (24, 24, 8), maximum pooling layers shape is (12, 12, 8), convolution layer 2 shape is (8, 8, 16), maximum pooling layers 2 shape is (4, 4, 16), the shape is flattened to (256) before passing through the two fully connected (dense) layers of shapes (128), and droupout (128), dense layer 2 (10). The number of output channels were controlled by 8 or 16. The total number of trainable parameters is 37,610.

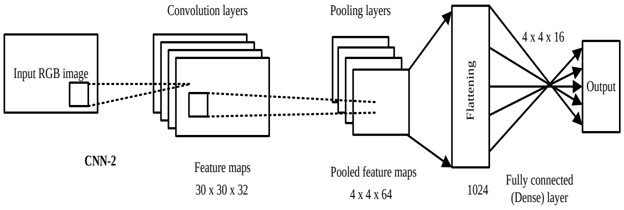


Figure 4: The structure of CNN-2 models with RGB image input

In Figure 4, the structure of CNN-2 model for RGB image input is composed of the hyperparameters at the input layer, convolution layers, pooling layers, flattening, fully connected (dense) layers, and output. In terms of (height, weight, channels) of the CNN-2, the convolution layer shape is (30, 30, 32), maximum pooling layers shape is (15, 15, 32), convolution layer 1 shape is (13, 13, 64), maximum pooling layers 2 shape is (6, 6, 64), the convolution layer 2 shape is (4, 4, 64), then the shape is flattened to (1024) before passing through the two fully connected (dense) layers of shapes is (4, 4, 64). The number of output channels were controlled by 32 or 64. The total number of trainable parameters is 56,320.
Output label blocks: These entail the evaluation and testing outcomes of the CNN-1 and CNN-2 models used for objects identification tasks.

Then, the effectiveness of these objection identification models in terms of accuracy, runtime, training loss, training accuracy, validation loss and validation loss are determined and analyzed. Also, the comparisons of these models are done with speed, loss and accuracy.

3.2 Experimental parameters

The paper adopts RGB and binary images features of objects for the CNN modeling for determining the object types in order to overcome the shortcomings of the traditional methods. The Google Collaboratory virtual machine environment was used for the CNN-based object identification modeling tasks. The procedure of choosing hyperparameters is the major aspect of the most of deep learning approaches, which can be achieved with manual or automatic way. The goals were to minimize cost and memory of execution. The learning algorithm make use of the hyperparameter settings for purpose of training datasets the context-specific dataset on the CNNs. The initial hyperparameters for the experimentations are provided in Table 2.

Table 2: The initial hyperparameters settings

Hyperparameter	Value
Number of convolution layers	2
Number of max pooling layers	2
Number of dense layers	4
Dropout rate	0.2
Optimizer	Adam
Activation function	Relu
Loss function	Binary-crossentropy
Number of epochs	10
Batch size	64
Simulator	Google Colaboratory

From Table 1, the different CNN model structures and associated hyperparameters' impacts on the outputs [33], which were done manually. The deeper network offers better the outcomes for image input classification tasks. Also, the type of the input data significantly impacts on the performance of the networks [34].

3.3 Data collection

The various CNNs parameters generated for determining objects classes through image features exaction, selection and argumentation. The validations of the proposed models were performed with standard datasets collected from standard image databases and repositories such as MINST and CIFAR-10 for binary images and RGB color images respectively. In totality, 60,000 samples of both image types were divided into 50,000 and 10,000 for training and validation purposes.

3.4 System flowchart

The CNN based objects identification models uses the preprocessed datasets previously collected concerning various attributes of image datasets as illustrated in Figure 5.

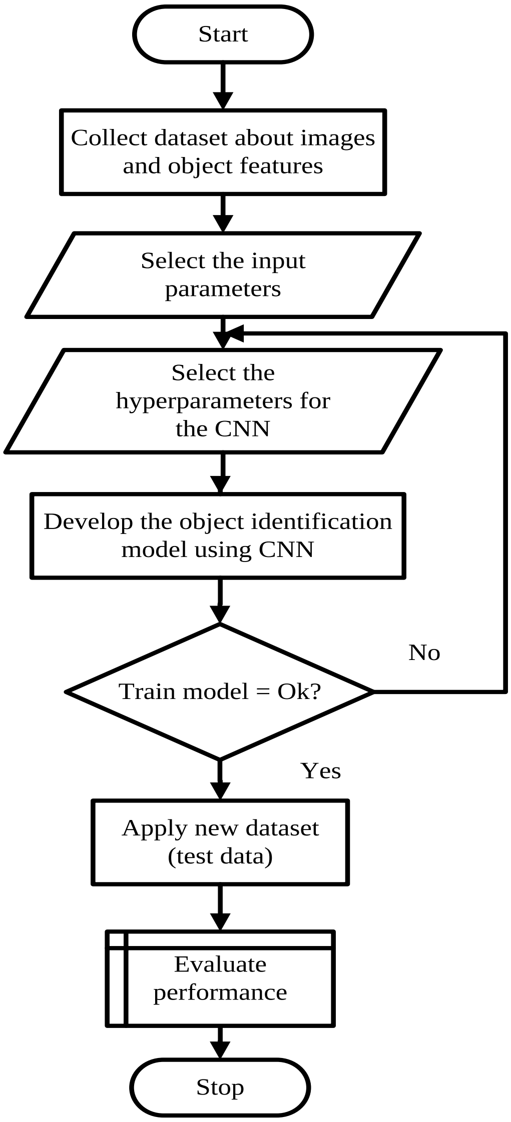

Start
Collect dataset about images and object features
Select the input parameters
Select the hyperparameters for the CNN
Develop the object identification model using CNN
Train model = Ok?
No
Yes
Apply new dataset (test data)
Evaluate performance
Stop

Figure 5: The system flowchart

3.5 Performance evaluation

The performances of the proposed approaches are computed using metrics represented in Equations 1 [35], [36]:

$$Accuracy = \frac{TN + TP}{FN + FP + TN + TP} \quad (1)$$

From Equation 1, the accuracy measures the rate of correctly predicted images to all the samples. The TP, FP, TN, and FN depict the following:

True Positive (TP): If the predicted image is actually object class, the prediction is TP.

False Positive (FP): If the predicted image is actually object class, the prediction is FP.

True Negative (TN): If the predicted image is actually object class, the prediction is TN.

False Negative (FN): If the predicted image is actually the object class, the prediction is FN.
Where, TP, FP, TN, and FN depict the following:
True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN).

Other metrics are provided in the study by [10]. These include: test loss, test accuracy, validation loss, validation accuracy, image size, and runtime.

This paper formulated the null hypothesis (Hn) and alternate hypothesis (Ha) to determine the whether: (a) the CNNs outcomes are not normally distribution as possessing heavy tails or outliers; (b) distribution of the differences between the CNNs are associated and symmetrical in shape.
Hn: The median difference between the pair outcomes of the CNNs is zero.
Ha: The median difference between the pair outcomes of the CNNs is not zero.
The Wilcoxon Signed Ranks Test [37] was adopted for the testing and analysis the stated hypotheses and implications on the outcomes of CNNs.

4 Results

4.1 The RGB image-based object identification

The objects identification using images from the CIFAR10 dataset (Python version) comprising of 60,000 RGB images for 10 classes in which each class consists of 6,000 images. The training dataset is 50,000 images (83.33%) and testing dataset is 10,000 images (16.67%) with classes mutually exclusive and non-overlapping. Class names as defined by Alex Krizhevsky in 2009 include: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The objects classes and label represented in the RGB images acquired are presented in Figure 6.

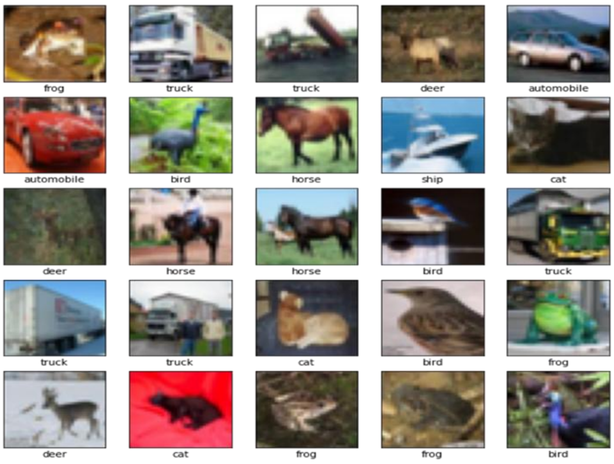
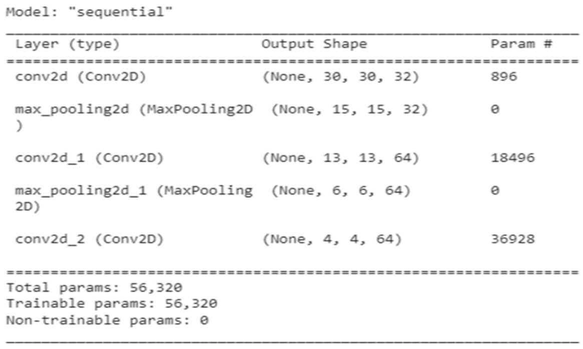


Figure 6: RGB images classes and labels for the objects

The CNN-2 model architecture for the identifying objects based on their RGB image representations in terms of layer type, output shape and parameters are shown in Figure 7.



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 30, 30, 32)        896

 max_pooling2d (MaxPooling2D  (None, 15, 15, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 13, 13, 64)        18496

 max_pooling2d_1 (MaxPooling  (None, 6, 6, 64)         0
 2D)

 conv2d_2 (Conv2D)           (None, 4, 4, 64)          36928

=================================================================
Total params: 56,320
Trainable params: 56,320
Non-trainable params: 0
_________________________________________________________________
```

Figure 7: The architecture of the CNN-2 model

In Figure 7, the output of every Conv2D and MaxPooling2D layer is a 3D tensor of shape (height, width, channels). The width and height dimensions tend to shrink with deeper network. The number of output channels for each Conv2D layer is controlled by the first argument (such as 32 or 64). Typically, as the width and height shrink, it is possible to computationally add more output channels in each Conv2D layer. The network summary shows that, (4, 4, 64) outputs were flattened into vectors of shape (1024) before going through two Dense layers.

The outcomes of performing training and validation of CNN-2 model with the selected RGB images of objects datasets are presented in Table 3.

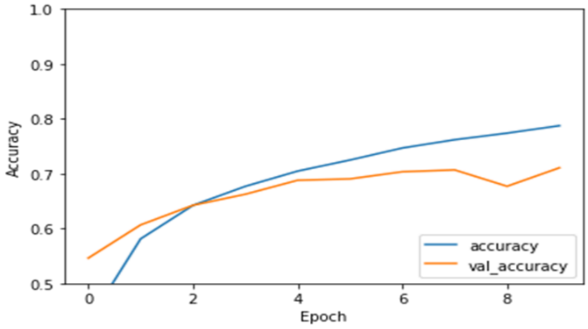


Figure 8: Object identification with CNN-2 model during training and validation

Table 3: CNN-2 model performance for RGB image of objects

Epoch	Duration (secs.)	Training loss	Training accuracy	Validation loss	Validation accuracy	Overall loss	Overall Accuracy
1	25	1.5592	0.4297	1.2685	0.5460	0.8830	0.7106
2	13	1.1814	0.5811	1.1101	0.6065		
3	13	1.0207	0.6421	1.0193	0.6423		
4	13	0.9185	0.6768	0.9548	0.6624		
5	13	0.8396	0.7046	0.8999	0.6879		
6	13	0.7812	0.7248	0.8999	0.6904		
7	13	0.7276	0.7467	0.8618	0.7034		
8	13	0.6823	0.7618	0.8607	0.7067		
9	13	0.6418	0.7738	0.9396	0.6769		
10	13	0.6058	0.7874	0.8830	0.7106		

From Table 3, the CNN-2 model used for identifying RGB images of objects achieved loss of 88.30% and 71.06% accuracy after 1 sec in epoch 1. The performance of CNN-2 model used for detecting RGB objects during training and validation are presented in Figure 8. From Figure 8, the accuracy of the CNN-2 model showed similar trend during training and validation procedures before epoch 2. Both curves increased steadily after epoch 2, but the validation reduced to peak at 0.7-point accuracy at epoch 10. While the training procedure increased steadily through epoch 10 but better than validation curve at 0.8-point accuracy.

4.2 The binary image-based object identification

The objects identification using binary images from the TensorFlow MINST dataset made up of 60,000 28 px by 28 px handwritten digits (0-9). The training dataset is 50,000 images (83.33%) and testing dataset is 10,000 images (16.67%) with each image mutually exclusive and non-overlapping. Though, each image in the dataset is a 28x28 matrix of integers (from 0 to 255), that is, each integer depicts a color of a pixel. One instance of objects is shown in Figure 9.

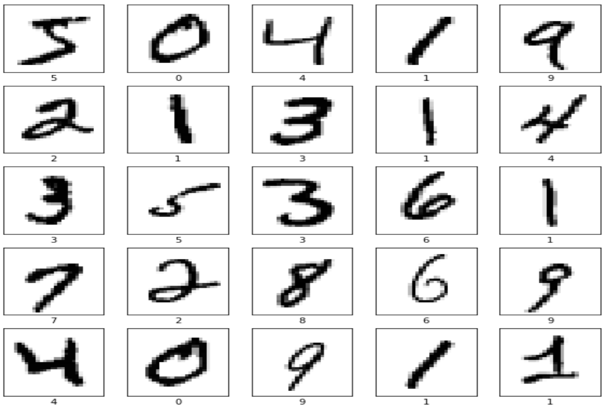
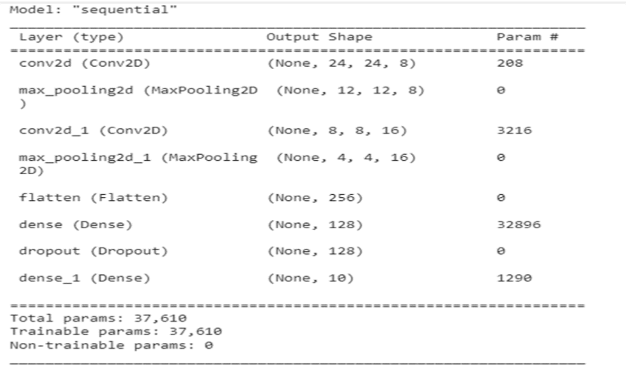


Figure 9: Image representation for matrix vectors of different object

The CNN-2 model architecture for the objects' identification tasks using digital handwritten images in terms of layer type, output shape and parameters are shown in Figure 10.



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 24, 24, 8)         208

 max_pooling2d (MaxPooling2D  (None, 12, 12, 8)        0
 )

 conv2d_1 (Conv2D)           (None, 8, 8, 16)          3216

 max_pooling2d_1 (MaxPooling  (None, 4, 4, 16)         0
 2D)

 flatten (Flatten)           (None, 256)               0

 dense (Dense)               (None, 128)               32896

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 10)                1290

=================================================================
Total params: 37,610
Trainable params: 37,610
Non-trainable params: 0
_________________________________________________________________
```

Figure 10: The architecture of the CNN-2 model

From Figure 10, the output of every Conv2D and MaxPooling2D layer is a 3D tensor of shape (height, width, channels). The width and height dimensions tend to shrink with deeper network. The number of output channels for each Conv2D layer is controlled by the first argument (such as 8 or 16). Typically, as the width and height shrink, it is possible to computationally add more output channels in each Conv2D layer. The network summary shows that, (4, 4, 16) outputs were flattened into vectors of shape (256) before going through two dense layers.

4.2.1 CNN-1 Model training and testing outcomes

The outcomes of the CNN-1 training and validation procedures for the sampled binary images in the objects' datasets are shown in Table 4.

Table 4: CNN-1 model performance for binary images of objects

Epoch	Duration (secs.)	Training loss	Training accuracy	Validation loss	Validation accuracy	Overall loss	Overall Accuracy
1	29	0.1998	0.9384	0.0690	0.9789	0.0320	0.9910
2	29	0.0696	0.9790	0.0488	0.9841		
3	29	0.0502	0.9845	0.0402	0.9865		
4	29	0.0398	0.9891	0.0323	0.9887		
5	29	0.0345	0.9891	0.0359	0.9886		
6	29	0.0267	0.9915	0.0319	0.9897		
7	29	0.0240	0.9922	0.0346	0.9893		
8	29	0.0206	0.9934	0.0307	0.9916		
9	29	0.0189	0.9939	0.0322	0.9899		
10	29	0.0162	0.9950	0.0320	0.9909		

From Table 4, the CNN-1 model approach for identifying binary images of objects achieved loss of 3.20% and 99.10% accuracy after 1 sec in epoch 1. The performance of CNN-1 model used for detecting binary objects during training and validation are presented in Figures 11 and 12.

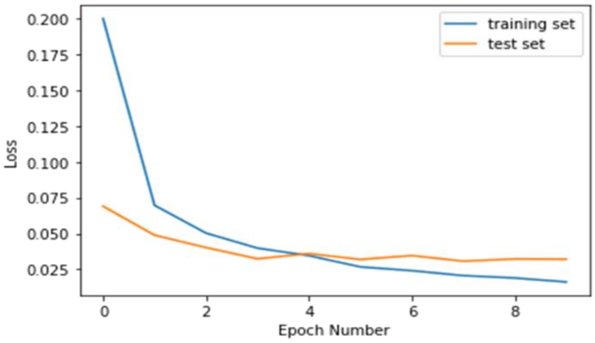


Figure 11: Loss of CNN-1 model during training and testing procedures

From Figure 11, the loss function was changing during the training, which is expect to get smaller and smaller on every next epoch.

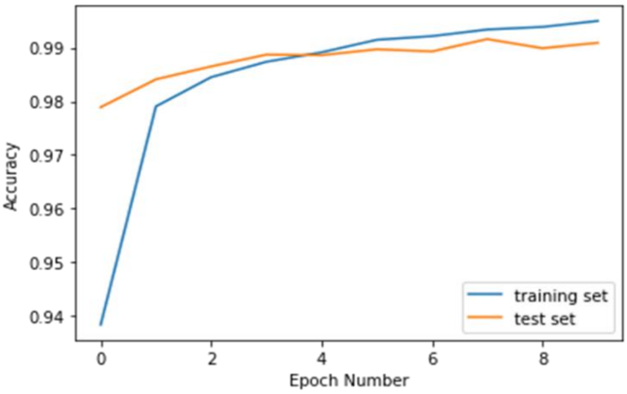


Figure 12: Accuracy of CNN-1 model during training and testing procedures

Similarly, in Figure 12, the accuracy function was changing during the training, which is expect to get larger after epoch 3 better than the test procedure throughout the epoch 10. However, both accuracy curves improved considerably throughout training and testing phases.

4.3 Comparisons of CNN models performances

The summary of performances of the proposed CNNs with different image datasets in terms of accuracy and loss functions are presented in Table 5.

Table 5: The performances of CNN-1 and CNN-2 objection detection compared

Model	Loss function	Accuracy function
CNN-2 with RGB images	0.8830	0.7106
CNN-1with Binary images	0.0032	0.9910

From Table 5, the average loss function computed for CNN-1 with binary images was better than CNN-2 with RGB images by 0.0032 to 0.8830. Similarly, CNN-1 model offered high accuracy of 0.9910 against 0.7106 for the CNN-2 with RGB images. These disparities are attributable to its simpler network structure and features map in the CNN-1 with Binary image data [38]. The graphical representation is illustrated by Figure 13.

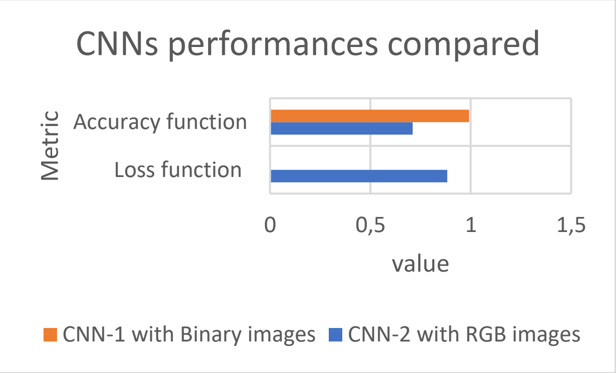


Figure 13: The performances of the CNNs models compared

Again, the summary of CNN models' performances under different image data against comparable works are presented in Table 6.

Table 6: The CNN models-based objects identification performances compared.

Metric	CNN-1 (Binary images)	CNN-2 (RGB images)	(Altwaijry & Al-Turaiki, 2021) –Binary images	(Li et al., 2021) RGB images
Average Duration (ms)	29.00	14.20	22.00	13.00
Loss (%)	3.20	88.30	Unspecified	Unspecified
Accuracy (%)	99.10	71.06	97.00	91.76
No. of Epochs	10	10	Unspecified	Unspecified
Size	60,000	60,000	47,434	4,000
Database	MINST	CIFAR-10	AHCD	YOLOvs

From Table 6, the outcomes revealed that, the average duration for CNN model with RGB images was better than CNN model with binary images (that is, 14.20 ms to 29.00 ms), which is agreement with previous works in RGB images [8], and binary images [3] respectively. Similarly, the loss functions for CNN model with binary images at 0.32% was better than CNN model with RGB images at 88.30%. The reverse was the case of accuracy, the outcomes revealed that, CNN-1 with RGB images was outperformed by the CNN-2 with Binary images by 71.06% to 99.10%. The same trend was observed in the benchmark studies in RGB images [8], and binary images [3].

The performance of the proposed CNNs outperformed existing comparable studies in terms of accuracy, average duration and number of images considered. Though, the performances of the traditional object images are better than handwritten object images in terms of average duration. Conversely, the performance of handwritten object images exceeds traditional object images in terms of loss and accuracy functions.

The reasons of the superior performance of CNN-2 with Binary images include:

1. There are lesser number of channels, height, and width dimensions in the binary image data than RGB image data.
2. There are deeper layers in the CNN-2 against CNN-1.
3. There is faster convergence for CNN-2 due to shorter flatten layer dimension connecting to the fully connected layers.
4. The features and features maps are lesser in the convolution and pooling layers for Binary image data than in RGB image.
5. The CNN-2 output layer is diminished against the CNN-1's.
6. The hyperparameters settings for CNN-1 are much simpler than CNN-2.

Hence, this paper offers useful guides on CNN models capability for solving diverse objects identification problems like surveillance, agriculture, and transportation systems with high level of reliability and efficiency.

4.4 Tests of hypothesis

Wilcoxon Signed Rank Test is used to determine the distribution normality of the CNN-1 and CNN-2 models' outcomes. To this end, this paper analyzed the paired observations of the CNNs without the assumption of normal distribution assumptions. The SPSS 23 Software was used to run the analysis whose outcomes are shown in Tables 6(a) and (b).

Table 6(a): Statistical analysis of hypotheses.

1. **Ranks**

		N	Mean Rank	Sum of Ranks
CNN-2 with Binary images	Negative Ranks (NR)	20[a]	30.50	610.00
CNN-1 with RGB images	Positive Ranks (PR)	20[b]	10.50	210.00
	Ties	0[c]		
	Total	40		

a. CNN-2 with Binary images < CNN-1 with RGB images
b. CNN-2 with Binary images > CNN-1 with RGB images
c. CNN-2 with Binary images = CNN-1 with RGB images

Table 6(b): Statistical analysis of hypotheses.

2. **Test Statistics[a]**

	CNN-2 with Binary images - CNN-1 with RGB images
Z	-2.688[b]
Asymp. Sig. (2-tailed)	.007

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

From Table 6(b), the test summary table showed that, the median differences between CNN-1 with Binary images and CNN-2 with RGB images not equals 0, which implies dissimilarity in the CNNs' performances. The p-value (0.007) computed from the Wilcoxon Signed Ranks Test is less than 0.05, which revealed that, the CNNs models' outcomes were statistically significant. Therefore, the null hypothesis was rejected because of presence of median differences between the CNNs.

5 Conclusion

This study has discovered that, the deep learning concept of CNNs showed greater promise in determining and mining features inside of the RGB and binary images, which produce high precision object identification applications such as agriculture, self-driving cars, surveillance, face recognition, etc., with high-level of performances. CNN models have rapidly increased over time in the field of object detection techniques and many computers vision-based projects. There are two kinds of images used in determining or recognizing objects and their characteristics, that is, binary images and RGB images. The results obtained revealed that, the average duration for CNN-2 model with RGB images was better than CNN-1 model with binary images (that is, 14.20 ms to 29.00 ms), which is agreement with previous works in RGB images [8], and binary images [3] respectively.

Also, the loss functions for CNN-1 model with binary images at 0.32% was better than CNN-2 model with RGB images at 88.30%. The reverse was the case of accuracy, the outcomes revealed that, CNN-2 modelled with RGB images was outperformed by the CNN-1 modelled with Binary images by 71.06% to 99.10% due to its simpler network structure and features map [33]. The same trend was observed in the comparable studies in RGB images [8], and binary images [3] respectively. The paper could not conduct the widescale investigations on the impact of datasets diversity and sizes on the performances of the CNNs, which is an interesting area in the future works.

References

[1] B. Mueller, T. Kinoshita, A. Peebles, M. A. Graber, and S. Lee, "Artificial intelligence and machine learning in emergency medicine: a narrative review," *Acute medicine & surgery*, vol. 9, no. 1, p. e740, 2022, doi: 10.1002/ams2.740.

[2] M. Kumar *et al.*, "Healthcare Internet of Things (H-IoT): Current trends, future prospects, applications, challenges, and security issues," *Electronics (Basel)*, vol. 12, no. 9, p. 2050, 2023, doi: 10.3390/electronics12092050.

[3] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Comput Appl*, vol. 33, no. 7, pp. 2249–2261, 2021, doi: 10.1007/s00521-020-05070-8.

[4] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *Applied Intelligence*, vol. 51, pp. 854–864, 2021, doi: 10.1007/s10489-020-01829-7.

[5] A. Mohiyuddin, A. R. Javed, C. Chakraborty, M. Rizwan, M. Shabbir, and J. Nebhen, "Secure cloud storage for medical IoT data using adaptive neuro-fuzzy inference system," *International Journal of Fuzzy Systems*, vol. 24, no. 2, pp. 1203–1215, 2022, doi: 10.1007/s40815-021-01104-y.

[6] J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network (CNN) applied to plant leaf disease classification," *Agriculture*, vol. 11, no. 8, p. 707, 2021, doi: 10.3390/agriculture11080707.

[7] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, "Comparative analysis of breast cancer detection using machine learning and biosensors," *Intelligent Medicine*, vol. 2, no. 2, pp. 69–81, 2022, doi: 10.1016/j.imed.2021.08.004.

[8] Y. Li, M. Li, J. Qi, D. Zhou, Z. Zou, and K. Liu, "Detection of typical obstacles in orchards based on deep convolutional neural network," *Comput Electron Agric*, vol. 181, p. 105932, 2021, doi: 10.1016/j.compag.2020.105932.

[9] A. Díaz-Álvarez, M. Clavijo, F. Jiménez, and F. Serradilla, "Inferring the driver's lane change intention through lidar-based environment analysis using convolutional neural networks," *Sensors*, vol. 21, no. 2, p. 475, 2021, doi: 10.3390/s21020475.

[10] A. Hazra, P. Choudhary, S. Inunganbi, and M. Adhikari, "Bangla-Meitei Mayek scripts handwritten character recognition using convolutional neural network," *Applied Intelligence*, vol. 51, no. 4, pp. 2291–2311, 2021, doi: 10.1007/s10489-020-01901-2.

[11] M. K. Benkaddour, S. Lahlali, and M. Trabelsi, "Human age and gender classification using convolutional neural network," in *2020 2nd international workshop on human-centric smart environments for health and well-being (IHSH)*, IEEE, 2021, pp. 215–220, doi: 10.1109/ihsh51661.2021.9378708.

[12] S. S. Alqahtany, A. B. Alkhodre, A. Al Abdulwahid, and M. Alohaly, "A Dynamic Multi-Layer Steganography Approach Based on Arabic Letters' Diacritics and Image Layers," *Applied Sciences*, vol. 13, no. 12, p. 7294, 2023, doi: 10.3390/app13127294.

[13] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J Physiol*, vol. 160, no. 1, p. 106, 1962, doi: 10.1113/jphysiol. 1962.sp006837.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] D. Yu, Q. Xu, H. Guo, C. Zhao, Y. Lin, and D. Li, "An efficient and lightweight convolutional neural network for remote sensing image scene classification," *Sensors*, vol. 20, no. 7, p. 1999, 2020, doi: 10.3390/s20071999.

[16] R. Pramanik and S. Bag, "Segmentation-based recognition system for handwritten Bangla and Devanagari words using conventional classification and transfer learning," *IET Image Process*, vol. 14, no. 5, pp. 959–972, 2020, doi: 10.1049/iet-ipr.2019.0208.

[17] A. Chaudhary, A. Hazra, and P. Chaudhary, "Diagnosis of chest diseases in x-ray images using deep convolutional neural network," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, pp. 1–6, doi: 10.1109/icccnt45670.2019.8944762.

[18] M. A. H. Akhand, M. Ahmed, M. M. H. Rahman, and M. M. Islam, "Convolutional neural network training incorporating rotation-based generated patterns and handwritten numeral recognition of major Indian scripts," *IETE J Res*, vol. 64, no. 2, pp. 176–194, 2018, doi: 10.1080/03772063.2017.1351322.

[19] S. Malakar, S. Paul, S. Kundu, S. Bhowmik, R. Sarkar, and M. Nasipuri, "Handwritten word recognition using lottery ticket hypothesis based pruned CNN model: a new benchmark on CMATERdb2. 1.2," *Neural Comput Appl*, vol. 32,

pp. 15209–15220, 2020, doi: 10.1007/s00521-020-04872-0.

[20] R. Ghosh, C. Vamshi, and P. Kumar, "RNN based online handwritten word recognition in Devanagari and Bengali scripts using horizontal zoning," *Pattern Recognit*, vol. 92, pp. 203–218, 2019, doi: 10.1016/j.patcog.2019.03.030.

[21] M. Tan, R. Pang, and Q. V Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp.10781–10790, doi: 10.1109/cvpr42600.2020.01079.

[22] R. Yang and Y. Yu, "Artificial convolutional neural network in object detection and semantic segmentation for medical imaging analysis," *Front Oncol*, vol. 11, p. 638182, 2021, doi: 10.3389/fonc.2021.638182.

[23] F. G. Venhuizen *et al.*, "Deep learning approach for the detection and quantification of intraretinal cystoid fluid in multivendor optical coherence tomography," *Biomed Opt Express*, vol. 9, no. 4, pp. 1545–1569, 2018, doi: 10.1364/boe.9.001545.

[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587, doi: 10.1109/cvpr.2014.81.

[25] J. Zhou, K. Feng, and L. Luo, "Research on fine-grained pattern recognition based on attention pattern-generated model," in *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 2019, p. 022038, doi: 10.1088/1755-1315/300/2/022038.

[26] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele, "Evaluation of output embeddings for fine-grained image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2927–2936, doi: 10.1109/cvpr.2015.7298911.

[27] J. Sai, A. Varma, and W. K. Song, "Human and bird detection and classification based on Doppler radar spectrograms and vision images using convolutional neural networks,". *International Journal of Advanced Robotic Systems, 18(3)*, 2021, doi: 10.1177/17298814211010569.

[28] S. Paisitkriangkrai, J. Sherrah, P. Janney, and V.-D. Hengel, "Effective semantic pixel labelling with convolutional networks and conditional random fields," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 36–43, doi: 10.1109/cvprw.2015.7301381.

[29] D. Dais, I. E. Bal, E. Smyrou, and V. Sarhosis, "Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning," *Autom Constr*, vol. 125, p. 103606, 2021, doi: 10.1016/j.autcon.2021.103606.

[30] Y. Zhou, Y. Lu, and Z. Pei, "Accurate diagnosis of early lung cancer based on the convolutional neural network model of the embedded medical system (vol 81, 103754, 2021) (Retraction of Vol 81, art no 103754, 2021)," 2024, *ELSEVIER RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS*, doi: 10.1016/j.micpro.2020.103754.

[31] W. Yang, X. Zhang, and P. Luo, "Transferability of convolutional neural network models for identifying damaged buildings due to earthquake," *Remote Sens (Basel)*, vol. 13, no. 3, p. 504, 2021, doi: 10.3390/rs13030504.

[32] X.-B. Fu, S.-L. Yue, and D.-Y. Pan, "Camera-based basketball scoring detection using convolutional neural network," *International Journal of Automation and Computing*, vol. 18, no. 2, pp. 266–276, 2021, doi: 10.1007/s11633-020-1259-7.

[33] E. H. Houssein, M. M. Emam, and A. A. Ali, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators' algorithm," *Neural Comput Appl*, vol. 34, no. 20, pp. 18015–18033, 2022, doi: 10.1007/s00521-022-07445-5.

[34] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, and H. M. Gheni, "An effective heart disease detection and severity level classification model using machine learning and hyperparameter optimization methods," *ieee access*, vol. 10, pp. 79974–79985, 2022, doi: 10.1109/access.2022.3191669.

[35] R. J. S. Raj, S. J. Shobana, I. V. Pustokhina, D. A. Pustokhin, D. Gupta, and K. Shankar, "Optimal feature selection-based medical image classification using deep learning model in internet of medical things," *IEEE Access*, vol. 8, pp. 58006–58017, 2020, doi: 10.1109/access.2020.2981337.

[36] R. Prasetya and A. Ridwan, "Data mining application on weather prediction using classification tree, naïve bayes and K-nearest neighbor algorithm with model testing of supervised learning probabilistic brier score, confusion matrix and ROC," *J. Appl. Commun. Inf. Technol*, vol. 4, no. 2, pp. 25–33, 2019, doi: 10.32497/jaict. v4i2.1690.

[37] M. A. Alzarrad, G. P. Moynihan, M. T. Hatamleh, and S. Song, "Fuzzy Multicriteria Decision-Making Model for Time-Cost-Risk Trade-Off Optimization in Construction Projects," *Advances in Civil Engineering*, vol. 2019, no. 1, p. 7852301, 2019, doi: 10.1155/2019/7852301.

[38] E. H. Houssein, M. M. Emam, and A. A. Ali, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators' algorithm," *Neural Comput Appl*, vol. 34, no. 20, pp. 18015–18033, 2022, doi: 10.1007/s00521-022-07445-5.