

Detecting and classifying media images of athletes using convolutional neural networks – case study: Individual sports images

Haider Shaker Mizher¹, Montadar Abas Taher², Bashar Saad Falih^{3,4},
Łukasz Adam Gierz^{5*}, Łukasz Warguła⁵, Bartosz Wieczorek⁵

¹ College of Basic Education, University of Diyala, Diyala, Iraq

² Department of Communications Engineering, College of Engineering, University of Diyala, Diyala, Iraq

³ Department of Computer Techniques Engineering, Al Salam University College, Baghdad, Iraq

⁴ Department of Electrical and Electronics Engineering, Cukurova University, Adana, Turkey

⁵ Institute of Machine Design, Faculty of Mechanical Engineering, Poznan University of Technology, Piotrowo 3, 60-965 Poznań, Poland

* Corresponding author's e-mail: lukasz.gierz@put.poznan.pl

ABSTRACT

Sports image classification using neural networks and machine vision is a rapidly expanding field, with applications in highlight reel creation, performance analysis, and illegal play detection. We present an innovative structure for sports image classification using convolutional neural networks (CNNs) based on deep learning in this article. Boxing, gymnastics, swimming, tennis, and weight lifting are five distinct sports that all fall under the umbrella of individual games. In terms of setting and attire, these various forms of athletic competition are very comparable. Specifically, the suggested deep learning model consists of 20-layers, among them, there are four CNN layers. The results show that the proposed model achieved a significant result in terms of accuracy, although the selected sports have similar characteristics to each other. For instance, boxing classification accuracy was 90.63%, gymnastics accuracy was 86.88%, swimming sport image classification achieved 94.06% accuracy, tennis classification accuracy is 88.13%, and the weight lifting was 89.06%, in the testing phase. The obtained results prove that the developed new sports image classification method is effective enough and has been improved.

Keywords: CNN, convolutional neural network, deep learning, deep neural network, sport image classification.

INTRODUCTION

Deep neural networks (DNNs) have recently gained much attention in various applications, especially in combination with CNNs [1]. They have been used to improve safety, such as urban flow prediction [2, 3], automatic video caption matching [4], speed control [5], in medicine for heart disease identifications and bone fractures detection in X-ray [6, 7], in agricultural robotic vision [8–10], fruit quality [11], handwritten digits recognitions [12], and cancer classification [13–17]. That is, the utilization of CNNs become essential in most of the classification issues [18]. Accordingly, sports images could be also classified. For

instance, classifying different sport images into, for instance, football playing, swimming, weight lifting and so on. There are many standard CNN models that can be employed for the abovementioned applications, such as LeNet-5, which structured to involve around 60.000 learnable parameters. The number 5 in its name is due the five layers in the whole structure of the model, two convolutional layers and three fully connected layers, hence, it is the simplest DNN based CNN model introduced in 1998 [19]. An improved version of LeNet-5 is the AlexNet [20], which has three more convolutional layers, then the total number of layers in the AlexNet is 8-layers. AlexNet was first introduced in 2012 where at that time it was

the largest/deepest neural network that was able to deal with ImageNet dataset [21], where there are 60 million learnable parameters in this AlexNet. An improved version of AlexNet is the GoogleNet which includes only 4-millions of learnable parameters with enhanced performance [22], which is also known as Inception-V1. The GoogleNet, which is from Google, constructed from 22-layers. In 2010, the sci-fi movie by Leonardo DiCaprio was the inspiration of introducing the Inception model of versions 2,3, and 4 [22–25]. Inception was constructed from blocks, the blocks are built from convolutional layers, in other words, it is not only stacking layers one after each other, hence, the name Inception. Visual Geometry Group also invented their own DNN called VGG [26]. They used the Rectified Linear Unit (ReLU) activation, which is borrowed from AlexNet, with more convolutional layers of filters that are reduced in size than AlexNet. The network constructed from 13-layers of convolution and three-layers of fully connected type, hence an overall leaning parameter of 138 million, this version is well-known as VGG-16, then the introduced the VGG-19 version. Then other standard models are introduced such as ResNet [27], DenseNet [28], ResNext [29], Channel Boosted CNN [30], and EfficientNet [31].

Most of the abovementioned standard models can be used either as pre-trained using the ImageNet dataset or can be used to be trained from scratch, a fresh-copy. However, most of the authors are interested in the trained versions where transfer learning could be utilized to tune the weights of the model, learnable parameters, to the intended dataset, such as that for the chest x-ray and other datasets. Sports datasets are different in their topics, for instance, images for football or images for various balls, players motion and reactions, and line detection. In this work, a dataset for 22-types of sports will be adopted for classification. This dataset is available online from Kaggle.com and it is called Sports Image Dataset that is collected from google.com image search. The aim of this work is to use a CNN-based DNN to quickly and accurately classify five types of sports: boxing, gymnastics, swimming, tennis, and weightlifting, and to identify which algorithm provides the highest identification accuracy. The reason behind selecting these five types are their challenging correlated features. The proposed model consists of 19-layers, excluding the input layer, in total of around 549000 learnable parameters. The structure of the article is organized as follows: in section 2, the works that are most

related to the discussed topic, i.e. the state of the art, will be presented, then the data set will be discussed in detail, and then the suggested model will be presented and analysed. In the results and discussion section, the complications of the selected 5 types of sports will be shown. In the conclusions section, the most important observations and planned further research directions will be presented.

RELATED WORK

Various published works display varying results from suggested models, whether they are their own or standard models. Furthermore, the literature contains a variety of datasets specific to different sports. In fact, the high correlation between the image's features makes it difficult to distinguish between different sport-type images. For example, features such as colors, lines, uniforms, dimensions, players, and actions can all slightly differ. Some sports, such as swimming, may have completely different features or at least more than 50% differences. The swimming pool's blue-colored water is the dominant feature in swimming sport pictures. Generally speaking, the conceptual gap in picture interpretation makes sports image categorization a challenging problem to solve. Accordingly, in 2007 Rongrong et al. [32] suggested a technique using the support vector machine (SVM) in hierarchical manner, i.e., a multi-SVM to tackle this problem. However, the pictures that have been utilized in their work is for more than five-types. In the next work, VGG-16 was used to classify 18 types of sports [33]. However, it can be seen that there are five sport types with environments that differ significantly from others. These five types – chess, football, swimming, table tennis, and tennis – showed the highest classification accuracy due to their distinct features, making them easier to differentiate from other sports [33]. The dataset used for this analysis was collected from the web by the authors themselves [33]. In the following study, the authors employed a novel combination of CNN and recurrent neural networks (RNN) to classify sequential frames of sports types [34]. This approach, while innovative, yielded less accurate results than the method [33]. Specifically, [34] aimed to classify five sports – basketball, cricket, football, ice hockey, and tennis – where the features are highly correlated, leading to classification challenges. As a

result, the CNN-based approach in [33] demonstrated improved accuracy compared to the combined CNN-RNN model in [34].

Joshi and his team use Inception-V3 to collect features from six types of sports and a neural network (NN) for classification purposes [35]. The adopted six-sport types are: rugby, tennis, cricket, basketball, volleyball, and badminton. Thus, in this research, they offer a solid framework for categorizing sports photos according to their context, such as the environment and surroundings. The dataset is constructed by extracting frames from films belonging to each sports genre on YouTube. The employed model is significantly deep and may face the overfit problem. The four categories of sport types: american football, rugby, soccer, and field hockey, have almost the same features, but significant different features are still available such as the ball's color and shape, shoulder pads, helmets, and visors. Differences in features also allowed for classification of similar sports images using a convolutional neural network [36]. Podgorelec et al. proposed setting hyperparameters using a differential evolution (DE) approach for the CNN model, which was used in two subsequent works [37, 38]. The suggested CNN model is based on the standard model VGG-19 [26] and achieved accuracy of 81.31%. That is, the suggested methodology is very complicated due to the employment of DE-algorithm for feature extraction. The dataset was created by the authors themselves, using google.com image search.

The analysis of sports sequence images using a CNN-based network is the main goal of the study described by Chen [39]. In light of the fact that multiframe detection algorithms are difficult and single-frame detection strategies have a poor detection rate, this study suggests a novel algorithm that combines the two methods. This will increase the detection rate of tiny targets and decrease the detection time. The convolution-layer is able to see data at various sizes and get additional input characteristics because of the network architecture. Furthermore, training becomes easier, the gradient vanishing phenomenon is successfully suppressed, and the network depth is decreased. However, the suggested approach has been trained on a dataset that is not related to sports types, it is trained on CIFAR-10 dataset, which has ten different categories including: airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks.

Combined DNN methods [40] include residual dense module, third-order hourglass networks, and attention-based graph convolution, are suggested

to introduce a novel DNN model for video sports classification. From various sources on the Internet, researchers of [40] collected 2200 sports photographs. Among them, 799 were of football, 689 were of swimming, and 712 were of table tennis. Thus, there are only three sport-types in this experiment. Swimming and football were the most accurate sports in the classification results, whereas table tennis had a low level of accuracy. This is due to the significantly uncorrelated features between the swimming and football sport-types, i.e., two different sport classes as team sport class and individual sport class. On the other hand, recognizing the actions allows one to anticipate the next move of a single athlete, which allows one to investigate the psychological effects of inauthentic moves in sports. In light of this, a three-dimensional convolutional model is optimized and provided [41]. That is, using CNN and recorded footage of basketball players' offensive and defensive processes, researchers can forecast their next moves by combining this data with applicable theories from sports psychology. Hence, only one sport-type was employed in their work. A CNN architecture was implemented [40] for movement classification from video dataset UCF101 [41]. However, the authors did not mention how was the depth of the proposed model, but they compared their work with standard models, that are mentioned previously. The number of sport-types was 10: basketball, diving, golf, horse riding, kicking-front, running, skate boarding-front, swing-bench, walk-front, and weight lifting, in other words, there are two sport classes, the team class and individual class sports.

Soccer, the world's most popular sport, captivates millions of people. Li and Ullah present a DL image classification algorithm to recognize soccer player activities from videos and images [44]. Their novel method uses CNNs and graph convolutional networks (GCNs) to identify complex spatial-temporal patterns in player attitudes and movements. The approach combines CNN and GCN. Continuous convolutional and pooling layers in the CNN generate discriminative visual characteristics using player position input frames. With bones as edges and skeletal joints as graph nodes, the GCN convolutionally combines both temporal and spatial data from neighbouring components of the body. So, local pose dynamics can be captured. Fusion of complementary CNN and GCN feeds through fully connected layers classifies player activities based on visual representations and pose arrangements. A multi-class cross-entropy loss was used

to train the model on extensively labelled soccer videos. Researchers classified seventeen complex football actions. Ablation studies verify CNN, GCN, and fused model contributions. This advances the use of DNNs for precise and detailed sporting event assessment. In the next work, DL was used to automatically classify sports images using large datasets [45]. The MobileNetV3 model and the modified battle royale optimization strategy were combined to achieve this goal. Then, various images from the Internet were processed using the suggested model. The outcomes were then evaluated alongside other cutting-edge techniques, such as InceptionV3, Bayesian classifier, and CNN. Participating sports in the study included six different kinds of games: rugby, cricket, badminton, basketball, tennis, and volleyball (combined classes of sports, individual and team games). Unfortunately, the work relied on the overly complicated standard model, MobileNetV3.

As indicated previously, with so many different kinds of sports to sort through, not to mention issues with feature recognition and less-than-ideal detection results, sports image classification is no easy feat. To tackle the issue of categorizing one hundred distinct types of sports images, the researchers use four pretrained models: ResNet-50, EfficientNet B7, DenseNet-121, and You Only Look Once version 8 (YOLOv8) [46]. The dataset provides a solid experimental basis for this research with 12,200 images of sports. Both the training and inference accuracy of the EfficientNet B7 model were 37.45% and 62.42%, respectively. Perhaps its underwhelming performance is because it struggles with certain types of sports image classification tasks, which are more demanding on its representational abilities. On the training set, DenseNet-121 achieved an accuracy of 71.791%, while on the validation set, it reached an accuracy of 86.211%. Since it outperforms EfficientNet B7, they can infer that the dense connectivity architecture is ideal for feature extraction from images. Averaging 94.90% accuracy on the training set and 96.60% on the validation set, the YOLOv8n model also performed exceptionally well. That is, the authors concluded that the dense connection was ideal for feature extraction process [46], while previous works in this section did not agree completely with this conclusion.

There has been a lot of success with picture classification, but accurately classifying sports images is still challenging. Current methods produce less-than-ideal categorization results because they can't handle the dynamic components and

complicated features found in sports images. Because of the complicated structure of sports images – which can include a wide range of lighting circumstances, backgrounds, and activity patterns— new and improved classification methods must be developed. As a result, complex motions, different camera angles, changing lighting, and overlapping scenes all contribute to the difficulties of sports image categorization. In this paper, we suggested an 18-layers, excluding the input and classification layers, that stucked in two branches, at the beginning then concatenated, as a first part. Thus, the first part consists of two parallel branches, each branch has one convolutional-layer followed by ReLU layer and batch normalization layer (BNL). The results from these two parts will be concatenated (a concatenation layer) to feed it to the second part. The second part consists of 11-layers, this second part is involving two convolutional layers and one-fully connected layers, of course there are ReLU, dropout, and pooling layers for overfit issue overcoming. The model was evaluated based on accuracy, F1-score, and recall. Results reveals that the suggested approach for the selected challenging sport-types are promising for realization.

DATASET DESCRIPTION AND PREPROCESSING

In this work, the adopted dataset was collected using manual search on the Internet by a contributor, and it has been uploaded to Kaggle.com and freely available through the link [<https://www.kaggle.com/datasets/sheikhzaib/sports-image-image-classification>]. There are 663, 635, 677, 667, and 562 images corresponding, respectively, to boxing, gymnastics, swimming, tennis, and weight lifting, as listed in Table 1. The original dataset contains 22 sport types: badminton, baseball, basketball, boxing, chess, cricket, fencing, football, Formula 1, gymnastics, hockey, ice hockey, kabaddi, MotoGP, shooting, swimming, table tennis, tennis, volleyball, weightlifting, wrestling, and WWE.

In this study, we have selected only five sport types: boxing, gymnastics, swimming, tennis, and weightlifting. As mentioned in the previous section, the rationale behind choosing these five types is their challenging, correlated features, which make them suitable for the objectives of this work.

Figure 1 shows a sample of 25 randomly selected images of these five sport types. The

Table 1. Elected sport games from the original dataset

Sport game type	Number of images
Boxing	663
Gymnastics	635
Swimming	677
Tennis	667
Weight lifting	562
Total	3204

image extensions were portable networks graphics (PNG), joint photographic experts group (JPG or JPEG), graphics interchange format (GIF), ASP.NET generic handler (ASHX), content management system (CMS), and Web picture format (WebP). That is, all of these formats have been converted to the PNG format before make use of the dataset. However, as the authors point of view, the five-selected games in Table 1 and Figure 1 are the most challenging types, since they

have mostly the same features. Therefore, we have selected these games for the current work. The adopted images, that are related to the selected five games, have been reviewed carefully, where there are no duplications, no images that are not related to the selected games, and there is no image that has zero size, in other words, there is no empty picture. All of these operations can be considered as a first step of the preprocessing, say manual preprocessing. As a second step of preprocessing operations, the selected-dataset was augmented, such as image-rotation, flip-left, flip-right, and flip-up/down.

However, for more exploration inside the selected games, Figure 2 shows 15-pictures of the selected sports. It can be seen that the dominated color is the blue. Further, lines, outfit, areas and other features are all similar to each other in all of the 15-pictures in Figure 2. That is, first column is Boxing-sport, second column is the

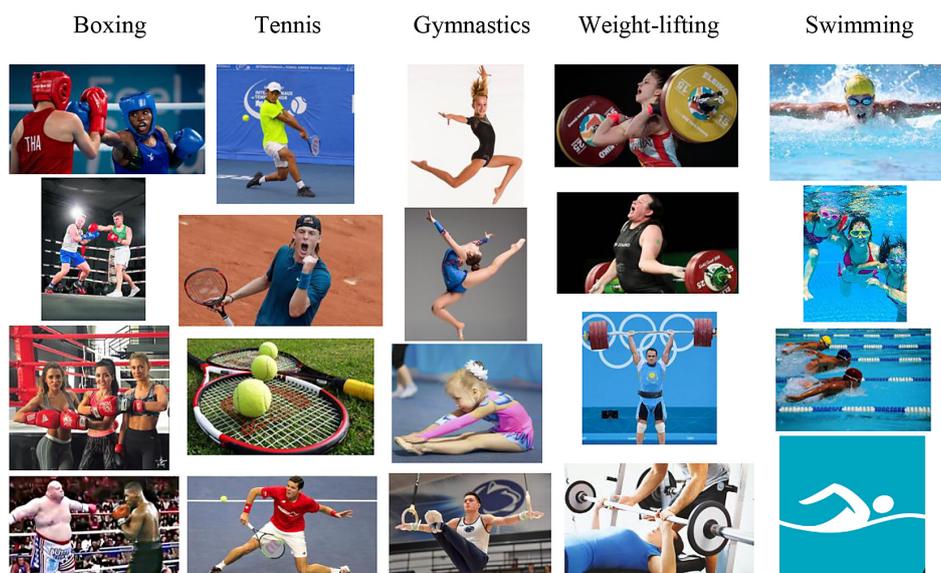


Figure 1. Twenty-samples of randomly selected images of the adopted 5-games dataset

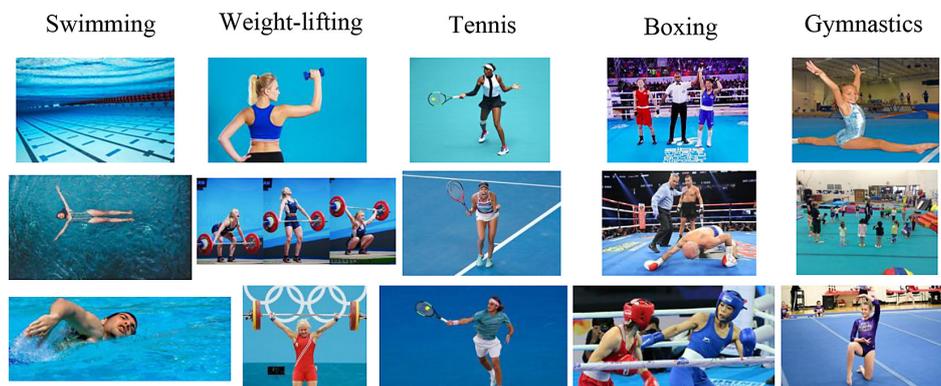


Figure 2. 15 sample images from the adopted 5-games dataset, illustrating the similar features shared among them

Gymnastic-sport, third is the swimming-sport, fourth is tennis-sport, and last is the weight lifting-sport, where in each column there are six-images of the same sport-type. This will make a noticeable confusion and mis-classification. Therefore, it is essential to design a DNN that has the capability to distinguish between these five-sports correctly as much as possible.

SUGGESTED MODEL

The suggested model in this work consists of 20 layers, as shown in Figure 3. The first layer is the image input layer, which expects an input image of size $15 \times 15 \times 3$. To match this requirement, the input images were resized to 15×15 pixels in width and height, with three channels (colored image). Next, the model processes the images through two parallel paths. Each path contains a ConvLyr, a ReLU activation layer, and a batch BNL. The ConvLyr in the first path constructed from 200-filters, with size of 3×3 . The second ConvLyr in the second path was designed to include 200-filters but with size of 5×5 . Each ConvLyr is followed by a ReLU layer and then by a BNL. Note that the ConvLyr padding is calculated during the training process such that the output dimension is equal to the input dimension. In other words, the horizontal or vertical padding will be added to the upper/lower edges equally if stride is set to 1, further, if the amount to be added is odd, then extra line of padding will be added. The same procedure will be followed vertically. However, the added values are zeros in both dimensions. The biasing of the ConvLyr is set to zeros. Each ConvLyr's weights have been initialized differently. Thus, for the first ConvLyr of the first path, the weights have been initialized using 'He' algorithm [47]. The second ConvLyr, which is in the second path has 200-filters with 5×5 as width and height. The initializations of this ConvLyr are similar to that of the first ConvLyr. The second ConvLyr is followed by ReLU and BNL. Each path performs the following mathematical operations:

For each input image or feature map X of size ($H=15 \times W=15$) with 3 input channels, and a filter F of size $K \times K$ (either 3×3 in first path or 5×5 in second path) with 200 filters, the output at position (i, j) in the feature map is given by:

Additionally, a bias b is added to the result of this convolution operation, allowing the network to learn an offset for each filter. The final output

for each position (i, j) in the feature map is given by Equation 1:

$$Y_{i,j} = \sum_{m=1}^K \sum_{n=1}^K \sum_{c=1}^3 X_{i+m-1,j+n-1,c} \cdot F_{m,n,c} + b \quad (1)$$

For ReLU activation, the output $Y_{i,j}$ after applying the convolution is then passed through a ReLU activation function, which is defined as:

$$Y_{i,j}^{ReLU} = \max(0, Y_{i,j}) \quad (2)$$

At the next, BNL, batch normalization can be applied, which normalizes the output of the layer across the batch:

$$Y_{i,j}^{BNL} = \gamma \frac{Y_{i,j}^{ReLU} - \mu}{\sqrt{\sigma^2 + \epsilon}} \quad (3)$$

where: μ and σ^2 are the mean and variance of the mini-batch, respectively. γ is learnable parameters that scale and shift the normalized output. ϵ is a small constant added for numerical stability.

As indicated in Figure 3, the two paths will be concatenated through the concatenation layer. The next part of the model is structured as one path of two ConvLyrs. Thus, the third ConvLyr in the model is represented by the first ConvLyr of the second part of the model, as shown in Figure 3. There are 128-filters of size 3×3 in aforementioned ConvLyr, followed by ReLU, BNL, and dropout layers. The dropout layer will drop connections with probability of 50%, to overcome overfitting problem. A dropout layer randomly sets a fraction of the activations to zero during training to prevent overfitting. If the dropout rate is p , each activation is kept with probability $1 - p$. The output after dropout is given by Equation 4:

$$Y_{i,j}^{Dropout} = Y_{i,j}^{BNL} \cdot M \quad (4)$$

where: M is a mask matrix with values of 0 or 1, sampled independently for each activation, where each value is 1 with probability $1 - p$.

The fourth ConvLyr (last ConvLyr in the suggested model) is designed with 256-filters of size 3×3 followed by ReLU, BNL, and a 25% dropout layer. A global average pooling layer is following which is followed by fully connected layer of five-nodes, according to the number of elected sports, then a SoftMax-layer as a last layer before the classification layer, as shown in Figure 3. However, the same initialization steps of the ConvLyrs of the first and second paths of the

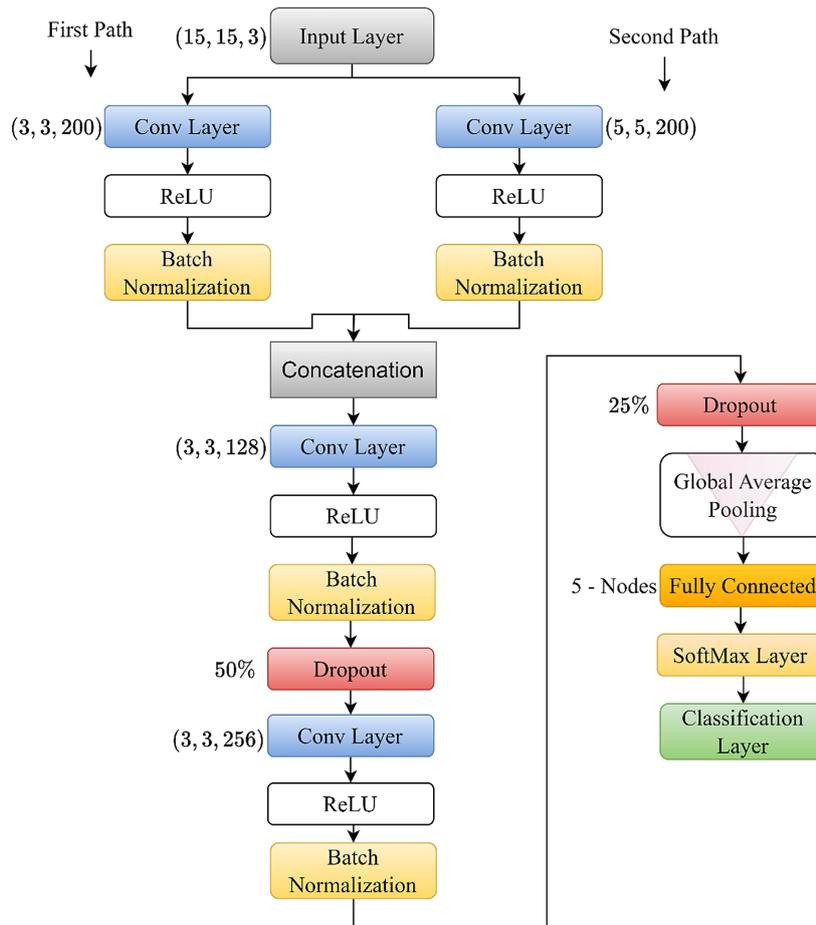


Figure 3. Suggestion of the DNN model that consists of twenty-layer

model was followed for these two last ConvLyr. The whole model was initialized with a specific hyperparameters, as will be seen later in the results section, these hyperparameters were adopted after a comprehensive investigation.

The output of the global average pooling layer is a vector G of length C , where each element G_c represents the average value for channel c .

For each channel c , the global average pooling operation is defined as:

$$G_c = \frac{\sum_{i=1}^H \sum_{j=1}^W Y_{i,j,c}}{H \times W} \quad (5)$$

In this context, $Y_{i,j,c}$ represents the value at spatial position (i, j) within channel c , and $H \times W$ denotes the total number of elements in each channel.

On the other hand, Table 2 lists all the weights of the suggested model in Figure 3. It is shown that there are 549,349 learnable parameters allocated in the 20-layers of Figure 3. Thus, the first layer, which is the image input layer, has no learnable parameters. The layer, which is the ConvLyr (in the first path), has 5600 learnable parameters.

Second ConvLyr, which is the first layer in the second path, has 15200 parameters that are capable to be learned. The difference between the first and second ConvLyr in the number of learnable parameters comes from the number of filters and the size of the filters in each layer. That is, due to 200-filters of size 3×3 within 3-channels, then the total number of learnable parameters will be $3 \times 3 \times 3 \times 200$, adding to them the number of biasing parameters, 200-parameters, then $3 \times 3 \times 3 \times 200 + 200 = 5600$ learnable parameters. In the second ConvLyr, there are 200-filters of size 5×5 for three channels and 200-biasing parameters, then there are $5 \times 5 \times 3 \times 200 + 200 = 15200$ parameters. However, the ReLU, dropout, SoftMax, Classification, Global average pooling, and Concatenation layers have no learnable parameters. Consequently, there are 400-learnable parameters in each BNl layer, as listed in Table 2. There are 128-filters of size 3×3 of 200-channels (that are the number of filters of the previous layer) in the third-ConvLyr then the total number of learnable parameters will be $3 \times 3 \times 200 \times 128$ with

Table 2. Learnable parameters of the suggested model

Layer index	Layer type	Number of learnable
1	Image Input	0
2	ConvLyr (1st path)	5600
3	ConvLyr (2nd path)	15200
4	ReLU (1st path)	0
5	ReLU (2nd path)	0
6	BNL (1st path)	400
7	BNL (2nd path)	400
8	Concatenation	0
9	ConvLyr	230528
10	ReLU	0
11	BNL	256
12	Dropout	0
13	ConvLyr	295168
14	ReLU	0
15	BNL	512
16	Dropout	0
17	Global average pooling	0
18	Fully connected	1285
19	SoftMax	0
20	Classification	0
Total		549.349

128-parameters of biasing, i.e., there are 230528 learnable parameters. The next BNL has 256 parameters followed by the fourth ConvLyr which is constructed from 256 filters of size 3×3 of 128 channels (according to the last layer size), then total number of parameters will be 295168 learnable parameters. Last but not least, there are another BNL of 512 parameters and 1285 parameters for the fully connected layer.

RESULTS AND DISCUSSION

The model that has been discussed in the previous section, which consists of 20-layers, will be initialized with a specific hyperparameters. The hyperparameters are chosen after tens of trials. For instance, the optimizer was adaptive moment estimation, adam, with gradient decay factor of 0.9. Further, the initial learning rate was set to 0.001, when the maximum number of epochs was 50. The learning rate will be decreased by a factor of 0.5 every 5-epochs. Note that at each epoch, the data will be shuffled to get real world results. The batch size was set to 56 samples from the dataset. Table 3 lists the hyperparameter settings

used in the model. To initiate the training process, the dataset of the selected five sports was divided into three subsets: training (70%), testing (20%), and validation (10%).

Figure 4 shows the training progress for 50-epochs. That is, there are 40-iterations per single epoch. The best validation point was at epoch number 48, at which the train, validation, and test accuracies were 81.32%, 74.38%, and 74.103%, respectively. At this epoch, the learning rate was decreased to 1.953125×10^{-6} as indicated in Figure 5. As indicated previously, most of the features of the adopted sports are similar to each other. Therefore, it is clearly reflected to the training progress in Figure 4. It can be seen that the training accuracy curve did not exceed the 81.32% level, even the validation accuracy was not more than 74.38%. Worth mention that another model was suggested (not shown in this work) was also employed, but the results did not improve.

In the sake of getting higher accuracy levels, training options were also changed. For instance, the initial learning rate value was set to more than 0.001, but the accuracy was not improved, on the contrary, it has been decreased down to around 65%. While decreasing the learning rate to less than 0.001, the accuracy did not exceed 55%. That is, the best learning rate was 0.001 in this work. Decreasing the learning rate period also has been changed, for instance, every 3-epochs or 10-epochs, these changes decreased the accuracy and overfit problem appeared. Increasing number of epochs also did not give better results, it made the model suffers from the overfitting problem. Learning rate drop factor varied in the experiment, for example, it is set to 75% to 10%, however, the

Table 3. Hyperparameters settings to train the suggested model

Hyperparameter	Value set
Optimizer	Adam
Optimizer gradient decay factor	0.9
Initial learning rate	0.001
Number of Epochs	50
Learning rate drop factor	0.5
Learning rate drop period	5 Epochs
Shuffling	Every single Epoch
Batch size	56
Train set ratio	70%
Test set ratio	20%
Validation set ratio	10%

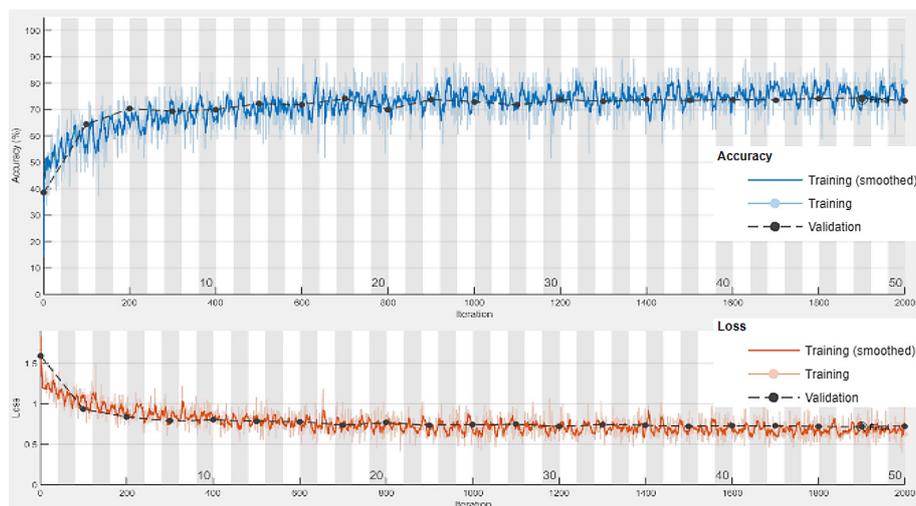


Figure 4. Train progress of the proposed model showing the accuracy and loss curves

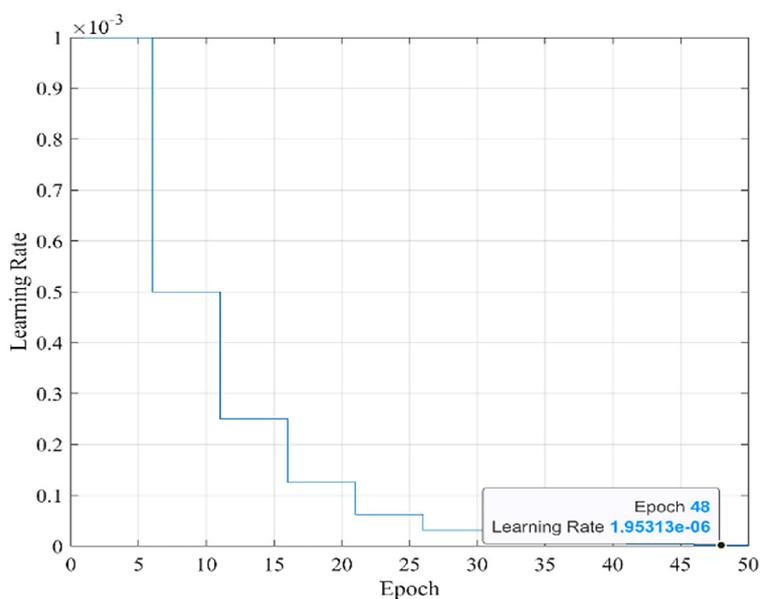


Figure 5. Learning rate dropping progress of the proposed model

best drop factor was 50% every 5-epochs. Moreover, the batch size was varied from 200 records to 20 records, the optimum number of records was found is 56-records, as listed in Table 3.

Consequently, confusion matrix of the training phase is shown in Figure 6. From this confusion matrix, various measurement metrics can be deduced for each class, individually. That is, the accuracy of the first class, boxing sport, was calculated from Figure 6 to be 93.09%, for the second class, gymnastics sport, the accuracy is 89.12%, for the swimming sport the delivered accuracy is 96.88%, Tennis accuracy was calculated as 92.91%, and the weight lifting accuracy is 90.64%. The highest accuracy was registered

in the third class, swimming sport, then the Boxing sport class, as listed in Table 4. Furthermore, in Table 4, there are other measure metrics, including: accuracy, sensitivity/recall, precision, specificity, negative predictive value (NPV), false positive rate (FPR), false detective rate (FDR), and F1-score. The highest values for these measures were found in the swimming sport class, as 93.88%, 93.88%, 98.35%, 97.68%, 1.65%, 6.12%, 92.71%, respectively. Note that the FPR and FDR are the best values among other classes. The degraded results are mostly found in the last class; 67.18%, 93.20%, 6.80%, 32.82%, and 71.54%, corresponding to precision, specificity, FPR, FDR, and F1-score, respectively. This is

Boxing	405	45		25	26
Gymnastics	31	333	8	26	67
Swimming	2	11	445	19	9
Tennis	9	20	13	377	27
Weight Lifting	17	36	8	20	264
	Boxing	Gymnastics	Swimming	Tennis	Weight Lifting

Figure 6. Confusion matrix of the training phase

clearly seen in Figure 6, where there are 26-Boxing records, 67-Gymnastic records, 9-swimming records, and 27-tennis records are all falsely classified as weight lifting-records. The measure metrics, including: accuracy, sensitivity/recall, precision, specificity, NPV, FPR, FDR, and F1-score as evaluated by mathematical Equation 6, Equation 7, Equation 8, Equation 9, Equation 10, Equation 11, Equation 12, and Equation 13 respectively.

$$\text{Accuracy} = \frac{\sum_{i=1}^5 TP_i}{\sum_{i=1}^5 (TP_i + TN_i + FN_i + FP_i)} \quad (6)$$

$$\text{Sensitivity or recall} = \frac{\sum_{i=1}^5 TP_i}{\sum_{i=1}^5 (TP_i + FN_i)} \quad (7)$$

$$\text{Precision} = \frac{\sum_{i=1}^5 TP_i}{\sum_{i=1}^5 (TP_i + FP_i)} \quad (8)$$

$$\text{Specificity} = \frac{\sum_{i=1}^5 TN_i}{\sum_{i=1}^5 (TN_i + FP_i)} \quad (9)$$

$$\text{NPV} = \frac{\sum_{i=1}^5 TN_i}{\sum_{i=1}^5 (TN_i + FN_i)} \quad (10)$$

$$\text{FPR} = \frac{\sum_{i=1}^5 FP_i}{\sum_{i=1}^5 (FP_i + TN_i)} \quad (11)$$

$$\text{FDR} = \frac{\sum_{i=1}^5 FP_i}{\sum_{i=1}^5 (TP_i + FP_i)} \quad (12)$$

$$\text{F1 - Score} = \frac{2 \sum_{i=1}^5 \text{Sensitivity}_i \cdot \text{Precision}_i}{\sum_{i=1}^5 \text{Sensitivity}_i + \text{Precision}_i} \quad (13)$$

where: True positives (*TP*) refer to the instances that are correctly predicted as belonging to a specific class. False positives (*FP*) are instances incorrectly predicted as belonging to that class. False negatives (*FN*) are instances that actually belong to the class but are predicted as belonging to a different class. True negatives (*TN*) are instances that are correctly predicted as not belonging to the class.

Figure 7 shows the confusion matrix for the validation phase. In the validation phase. Most of the sport game pictures are confused with each other, this can be observed clearly in the training progress of Figure 4 (black bold line). However, to

Boxing	107	12		10	15
Gymnastics	11	89	4	8	17
Swimming		2	127	15	4
Tennis	3	9	3	90	15
Weight Lifting	12	15	1	10	62
	Boxing	Gymnastics	Swimming	Tennis	Weight Lifting

Figure 7. Confusion matrix of the validation phase

Table 4. Train phase metrics results

Metric	Boxing	Gymnastics	Swimming	Tennis	Weight lifting
Accuracy	93.09%	89.12%	96.88%	92.91%	90.64%
Sensitivity or recall	80.84%	71.61%	93.88%	84.53%	76.52%
Precision	87.28%	74.83%	93.88%	80.73%	67.18%
Specificity	96.61%	93.70%	98.35%	94.99%	93.20%
NPV	94.60%	92.66%	97.68%	96.11%	95.62%
FPR	3.39%	6.30%	1.65%	5.01%	6.80%
FDR	12.72%	25.17%	6.12%	19.27%	32.82%
F1-score	83.94%	73.19%	92.71%	82.58%	71.54%

understand it deeply, the confusion matrix delivers various measurement metrics, as listed in Table 5. It can be deduced from Table 5 that the swimming sport results are dominated with respect to other sports. The worst results in Table 5 are those of the weight lifting sport. However, the accuracies of the validation phase are: 90.17%, 87.83%, 95.48%, 88.61%, and 86.12% corresponding to boxing, gymnastics, swimming, tennis, and weight lifting, respectively. There is no much difference between tennis and weight lifting sport accuracies in the validation phase, which is 0.49%, while in the training phase the difference is 2.27%, which is a logical result. Note that the FPR of the swimming sport is the lowest one that is equal to 1.62%, which is the goal, while the other sports are higher: 5.23%, 7.42%, 8.25%, and 9.43% for the boxing, gymnastics, tennis, and weight lifting, respectively.

Last but not least, testing phase confusion matrix is shown in Figure 8. As expected, swimming sport results are the superior among other sports.

However, the deduced measurement metrics from the confusion matrix in Figure 8 may

Boxing	59	11		5	7
Gymnastics	2	41	2	7	9
Swimming		3	61	8	1
Tennis	2	6	1	42	4
Weight Lifting	3	2	4	5	35
	Boxing	Gymnastics	Swimming	Tennis	Weight Lifting

Figure 8. Confusion matrix of the validation phase

produce deep exploration in the results, as indicated in Table 6. For instance, starting from the boxing sport game; there are 2-gymnastic images, 2-tennis images, and 3-weight lifting images are all wrongly classified as boxing. The most wrongly classified images are in the tennis column, where there are 5 images classified as boxing, 7 images classified as gymnastics, 8 images as swimming, and 5 images as weight lifting. Accordingly, the FPR of the tennis sport game is higher than others, 9.43%, and its specificity is the lowest, 90.57%, as listed in Table 6. The highest FDR is captured in the weight lifting sport images, 37.50% as worse, this is due to 3-boxing, 2-gymnastics, 4-swimming, and 5-tennis images are classified as weight lifting images, leading to second worse FPR in Table 6 as 7.75%. Nevertheless, the accuracies of the five sports were: 90.63%, 86.88%, 94.06%, 88.13%, and 89.06% for the boxing, gymnastics, swimming, tennis, and weight lifting, respectively. That is, the highest accuracy was 94.06% for the swimming sport and the lowest was 86.88% for the gymnastic sport. Thus, the suggested model was capable to distinguish between the sport games, although the images are too similar to each other, as shown in Figure 1 and Figure 2.

Comparing our results with other similar works, it can be seen that the results from this work are more accurate and generalizable, and that the selected sports belong to one class, which are individual sports. Wang and Sofla achieved high accuracy results, which are around 95% and 99%, but these results concern 6 sports that are significantly uncorrelated with each other [45]. The six sports are volleyball, basketball, badminton, rugby, tennis and cricket, where these sports belong to the classes of team games and individual games. In other words, the differences between the images of these six sports are significantly pronounced

Table 5. Validation phase metrics results

Metric	Boxing	Gymnastics	Swimming	Tennis	Weight lifting
Accuracy	90.17%	87.83%	95.48%	88.61%	86.12%
Sensitivity or recall	74.31%	68.99%	94.07%	75.00%	62.00%
Precision	80.45%	70.08%	94.07%	67.67%	54.87%
Specificity	94.77%	92.58%	98.38%	91.75%	90.57%
NPV	92.72%	92.22%	95.85%	94.09%	92.80%
FPR	5.23%	7.42%	1.62%	8.25%	9.43%
FDR	19.55%	29.92%	5.93%	32.33%	45.13%
F1-score	77.26%	69.53%	89.75%	71.15%	58.22%

Table 6. Test phase metrics results.

Metric	Boxing	Gymnastics	Swimming	Tennis	Weight lifting
Accuracy	90.63%	86.88%	94.06%	88.13%	89.06%
Sensitivity or recall	71.95%	67.21%	89.71%	76.36%	71.43%
Precision	89.39%	65.08%	89.71%	62.69%	62.50%
Specificity	97.06%	91.51%	97.17%	90.57%	92.25%
NPV	90.94%	92.22%	95.24%	94.86%	94.70%
FPR	2.94%	8.49%	2.83%	9.43%	7.75%
FDR	10.61%	34.92%	10.29%	37.31%	37.50%
F1-score	79.73%	66.13%	86.52%	68.85%	66.67%

and therefore the classification accuracy was high. In the next work, the accuracy for the same six sports mentioned above was achieved, which is around 86% and 98% [35]. Moreover, Podgorelec et al. achieved accuracy of around 81.31%, also for the same six sports [36] that are mentioned above. This means that the works that were found in the literature combined two main classes of sports games: team games and individual games, unlike our adopted sports that are related only to individual sports. Hence, the results that are achieved in this work can be considered acceptable with respect to other works.

CONCLUSIONS

The need for accurate sports image categorization has grown in tandem with the media's lavish coverage of sporting events. Conventional methods are inadequate for handling massive volumes of data and distinguishing between extremely similar images due to the human intervention required to select relevant attributes.

The learning process of the proposed model involves several steps to achieve optimal classification results. The optimization algorithm used is Adam, which updates the model's weights based on adaptive estimations of the second-order moments of gradients. The initial learning rate was set to 0.001 and reduced by a factor of 0.5 every five epochs to enhance convergence. During training, the total number of steps was 768, with 56 samples processed per epoch. The data was shuffled at the beginning of each epoch to prevent overfitting and ensure that the model does not rely on specific patterns within the data.

Backpropagation was employed to calculate the gradient of the loss function with respect to the weights, which were iteratively adjusted to

minimize error. To further reduce overfitting, regularization techniques such as dropout layers were applied. These layers randomly disable a portion of the neuron connections during training, enhancing the model's ability to generalize to unseen data. This systematic approach ensures effective learning while maintaining the model's robustness and applicability in various scenarios.

In this work, an architecture was developed that uses a specific method of combining multiple convolutional layers. Then, the results of applying the suggested model to different images downloaded from kaggle.com were compared with the results of other state-of-the-art techniques. The selected images are from a specific category of sports: individual sports, where the focus is on the individual athlete rather than the team. Some examples of these sports are weightlifting, swimming, gymnastics, tennis, and boxing. Due to the high correlation between the selected sports, the model achieved accuracy of 93.09%, 89.12%, 96.88%, 92.91%, and 90.64% for boxing, gymnastics, swimming, tennis, and weightlifting in the training phase. In the validation phase, the accuracies reached 90.17%, 87.83%, 95.48%, 88.61%, and 86.12%, respectively. During the testing phase, the accuracies were: boxing 90.63%, gymnastics 86.88%, swimming 94.06%, tennis 88.13% and weightlifting 89.06%. So, in this work, the problem of almost similar sports environments is solved and the proposed model can be used to classify such images with acceptable accuracy.

Based on the positive findings of this study, future research will seek to improve the generalisability of the model and extend the study to other sports categories other than individual sports. One of the areas of improvement that has been planned is to improve the architecture to work in complex environments where athletes may

be partially occluded or in group settings, such as in team sports. Moreover, the routines for real-time processing will enable the categorization of material in live sports broadcasts promptly and accurately.

The other direction pertains to extending current approaches toward extracting the nuanced characteristics in sports images, including athlete stance and motion that can be improved by applying transfer learning and hybrid models. Adding more images to the dataset that are different from the ones used in training and are of high quality will enhance the generalization ability of the model. The intended result is a robust, accurate, and flexible sports image classification system that can be used and implemented in real-world applications, like automated sports content categorization that may be useful to media companies or bringing improved analytical tools to coaches and sportscasters.

REFERENCES

1. Albawi S., Mohammed T. A., and Al-Zawi S., Understanding of a convolutional neural network, in 2017 International Conference on Engineering and Technology (ICET), 21–23 Aug. 2017, 1–6, <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
2. Salman E. H., Taher M. A., Hammadi Y. I., Mahmood O. A., Muthanna A., and Koucheryavy A., An Anomaly Intrusion Detection for High-Density Internet of Things Wireless Communication Network Based Deep Learning Algorithms, *Sensors*, 2023; 23(1), 206, [Online]. Available: <https://www.mdpi.com/1424-8220/23/1/206>
3. Hussain A. H. A. et al., Urban traffic flow estimation system based on gated recurrent unit deep learning methodology for internet of vehicles, *IEEE Access*, 2023; 11, 58516–58531, <https://doi.org/10.1109/ACCESS.2023.3270395>
4. Yousif A. J. and Al-Jammas M. H. Real-time Arabic video captioning using CNN and transformer networks based on parallel implementation, *Diyala Journal of Engineering Sciences*, 03/07 2024; 17(1), 84–93, <https://doi.org/10.24237/djes.2024.17108>
5. Abbas H. I. and Abd A. N. Adaptive inverse neural network based DC motor speed and position control using FPGA, *Diyala Journal of Engineering Sciences*, 2018; 11(3), 71–78, <https://doi.org/10.24237/djes.2018.11311>
6. Al-Saffar B., Ali Y. H., Muslim A. M., and Ali H. A. ECG Signal Classification Based on Neural Network, in International Conference on Emerging Technologies and Intelligent Systems, Cham: Springer International Publishing, September 202; 23–11.
7. Saad A., Sheikh U. U., and Moslim M. S. Developing convolutional neural network for recognition of bone fractures in x-ray images, *Advances in Science and Technology. Research Journal*, 2024; 18(4), 228–237, <https://doi.org/10.12913/22998624/188656>
8. Falih B. S., Ali Y. H., Alabbas A. R., and Arica S. Optimising yield estimation for grapes: utilising the sliding window technique for visual counting of bunches and berries, *Pakistan Journal of Agricultural Sciences*, 2024; 61(2).
9. Al-Saffar B., Arica S., and Tangolar S. Automatic counting of grapes from vineyard images, *Pakistan Journal of Agricultural Sciences*, May 2022; 59(3).
10. Falih B. S., Gierz Ł., and Al-Sammarraie M.A.J. fruit classification by assessing slice hardness based on rgb imaging. Case study: apple slices, *J. Appl. Math. Comput. Mech.*, 2024; 23(3), 7–18, <https://doi.org/10.17512/jamcm.2024.3.01>
11. Falih, B. S., Gierz, Ł. A., and Al-Zaidi, G. A. Detecting clustered fruits using a hybrid of convolutional neural networks and machine learning classifiers – Case study: Grapes. *Adv. Sci. Technol. Res. J.* 2024; 19(1), 1–9.
12. Al-Saffar B., Al-Abbas A. R., Özel S. A. A comparative study on the recognition of English and Arabic handwritten digits based on the combination of transfer learning and classifier, In International Conference on Emerging Technologies Intelligent Systems, Cham: Springer International Publishing, September 2022; 95–107.
13. Taha Ahmed S. and Malallah Kadhem S. Using machine learning via deep learning algorithms to diagnose the lung disease based on chest imaging: A survey, *International Journal of Interactive Mobile Technologies (iJIM)*, 08/23 2021; 15(16), 95–112, <https://doi.org/10.3991/ijim.v15i16.24191>
14. Ahmed S. T. and Kadhem S. M. Early Alzheimer’s disease detection using different techniques based on microarray data: A review, *International Journal of Online and Biomedical Engineering (iJOE)*, 2022; 18(4), 106–126, <https://doi.org/10.3991/ijoe.v18i04.27133>
15. Nogay, H. S., Akinci, T. C., Yilmaz, M. Comparative Experimental Investigation and Application of Five Classic Pre-Trained Deep Convolutional Neural Networks via Transfer Learning for Diagnosis of Breast Cancer. *Advances in Science and Technology Research Journal*, 2021; 15(3), 1–8. <https://doi.org/10.12913/22998624/137964>
16. Abdullah A. J., Hasan T. M., and Waleed J. An expanded vision of breast cancer diagnosis approaches based on machine learning techniques, in 2019 International Engineering Conference (IEC), 23–25 June 2019; 177–181, <https://doi.org/10.1109/IEC47844.2019.8950530>

17. Saad A., Muhanad Hameed A., and Jumana W. Skin cancer classification dermatologist-level based on deep learning model, *Acta Scientiarum. Technology*, 12/19 2022; 45(1), <https://doi.org/10.4025/actasci-technol.v45i1.61531>
18. Jawad A. T., Maaloul R., and Chaari L. A comprehensive survey on 6G and beyond: Enabling technologies, opportunities of machine learning and challenges, *Computer Networks*, 2023/12/01; 237, 110085, <https://doi.org/10.1016/j.comnet.2023.110085>
19. Lecun Y., Bottou L., Bengio Y., and Haffner P. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 1998; 86(11), 2278–2324, <https://doi.org/10.1109/5.726791>
20. Krizhevsky A., Sutskever I., and Hinton G. E. ImageNet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 2012; 25, <https://doi.org/https://doi.org/10.1145/3065386>
21. Deng J., Dong W., Socher R., Li L. J., Kai L. and Li F.-F. ImageNet: A large-scale hierarchical image database, in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 20–25 June 2009; 248–255, <https://doi.org/10.1109/CVPR.2009.5206848>
22. Szegedy C. et al. Going deeper with convolutions, in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015; 1–9, <https://doi.org/10.1109/CVPR.2015.7298594>
23. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., and Wojna Z. Rethinking the Inception Architecture for Computer Vision, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 27–30 June 2016, 2818–2826, <https://doi.org/10.1109/CVPR.2016.308>
24. Ioffe S. and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift, presented at the Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, Lille, France, 2015.
25. Szegedy C., Ioffe S., Vanhoucke V. and Alemi A. A. Inception-v4, inception-ResNet and the impact of residual connections on learning, presented at the Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017.
26. Simonyan K. and Zisserman A. Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
27. He K., Zhang X., Ren S., and Sun J. Deep Residual Learning for Image Recognition, Los Alamitos, CA, USA: IEEE Computer Society, 2016/06; 770–778, <https://doi.org/10.1109/CVPR.2016.90>
28. Huang G., Liu Z., Maaten L. V. D., and Weinberger K. Q. Densely Connected Convolutional Networks, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017, 2261–2269, <https://doi.org/10.1109/CVPR.2017.243>
29. Xie S., Girshick R., Dollár P., Tu Z., and He K. Aggregated Residual Transformations for Deep Neural Networks, in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017; 5987–5995, <https://doi.org/10.1109/CVPR.2017.634>
30. Khan A., Sohail A., and Ali A. A new channel boosted convolutional neural network using transfer learning, arXiv preprint arXiv:1804.08528, 2018.
31. Tan M. and Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, presented at the Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2019. [Online]. Available: <https://proceedings.mlr.press/v97/tan19a.html>
32. Ji R., Yao H., Xu P., and Zhang Z. Sports image classification based on hierarchical multi-svm bagging network, in *Information Sciences* 2007; 973–979.
33. Farhad M. Y., Hossain S., Tanvir M. R. K., and Chowdhury S. A. Sports-Net18: Various Sports Classification using Transfer Learning, in 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), 19-20 Dec. 2020; 1–4, <https://doi.org/10.1109/STI50764.2020.9350415>
34. Russo M. A., Filonenko A., and Jo K. H. Sports Classification in Sequential Frames Using CNN and RNN, in 2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT), 6–8 Sept. 2018; 1–3, <https://doi.org/10.1109/ICT-ROBOT.2018.8549884>
35. Joshi K., Tripathi V., Bose C., and Bhardwaj C. Robust sports image classification using inceptionV3 and neural networks, *Procedia Computer Science*, 2020/01/01; 167, 2374–2381, <https://doi.org/https://doi.org/10.1016/j.procs.2020.03.290>
36. Podgorelec V., Pečnik Š., and Vrbančič G. Classification of Similar Sports Images Using Convolutional Neural Network with Hyper-Parameter Optimization, *Applied Sciences*, 2020; 10(23), 8494, [Online]. Available: <https://www.mdpi.com/2076-3417/10/23/8494>
37. Storn R. and Price K. Differential evolution-A simple and efficient adaptive scheme for global optimization over continuous spaces, *Technical Report TR-95.012, ICSI*, 1995.
38. Storn R. and Price K. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, *Journal of Global Optimization*, 1997/12/01; 11(4), 341–359, <https://doi.org/10.1023/A:1008202821328>
39. Chen Y. Sports sequence images based on convolutional neural network, *Mathematical Problems in*

- Engineering, 2021; 2021(1), 3326847, <https://doi.org/10.1155/2021/3326847>
40. Gao M., Cai W., and Liu R. [Retracted] AGTH-Net: Attention-Based Graph Convolution-Guided Third-Order Hourglass Network for Sports Video Classification, *Journal of Healthcare Engineering*, 2021; 2021(1), 8517161, <https://doi.org/10.1155/2021/8517161>
41. Liu J. Convolutional Neural Network-Based Human Movement Recognition Algorithm in Sports Analysis, (in English), *Frontiers in Psychology*, Original Research June-25 2021; 12, <https://doi.org/10.3389/fpsyg.2021.663359>
42. Ramesh M. and Mahesh K. Sports video classification framework using enhanced threshold based keyframe selection algorithm and customized CNN on UCF101 and Sports1-M Dataset, *Computational Intelligence and Neuroscience*, 2022; 1, 3218431, <https://doi.org/10.1155/2022/3218431>
43. Idrees H. et al. The thumos challenge on action recognition for videos “in the wild”, *Computer Vision and Image Understanding*, 2017; 155, 1–23.
44. Li X. and Ullah R. An image classification algorithm for football players’ activities using deep neural network, *Soft Comput.*, 2023; 27(24), 19317–19337, <https://doi.org/10.1007/s00500-023-09321-3>
45. Wang B. and Rezaei Sofla A. Solution for sports image classification using modified MobileNetV3 optimized by modified battle royal optimization algorithm, *Heliyon*, 2023; 9(11), e21603, <https://doi.org/10.1016/j.heliyon.2023.e21603>
46. L. Xiangchen, “Comparison of Four Convolutional Neural Network-Based Algorithms for Sports Image Classification,” in *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, 2024/02/14 2024: Atlantis Press, 178–186, https://doi.org/10.2991/978-94-6463-370-2_20
47. He K., Zhang X., Ren S., and Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, in *2015 IEEE International Conference on Computer Vision (ICCV)*, 7–13 Dec. 2015; 1026–1034, <https://doi.org/10.1109/ICCV.2015.123>