International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Classification of Turkish Folk Dances using Deep Learning

Hamidullah Nazari[1], Sümeyye Kaynak*[2]

Submitted: 22/07/2022 **Accepted**: 25/09/2022

Abstract: The folk dances, which reflect the cultural values of the society consist of regular and continuous activities performed singly or in groups, accompanied by music. Folk dances show differences according to climate, geographical position and socio-economic status of the society. Classification of Turkish folk dances is an interesting subject due to the different hand, arm and foot postures of the dancers. The complexity of the background, camera angle, special clothing can cause the algorithms to give incorrect results. It gives a playground to experiment with different deep learning techniques. In this study, the classification of Turkish folk dances is carried out from video images with deep learning methods. Zeybek, Çiftetelli, Horon and Halay dances which are among Turkish folk dances were selected. The dataset consists of 24000 images in total, prepared as 6000 images belonging to each class (Halay, Zeybek, Çiftetelli, Horon). 75% of the images make up the training set and the validation set is 25% of the images. CNN model and pre-trained VGG16 and ResNet50 architectures with transfer learning technique are used in the classification of Turkish folk dances in video images. The models have been tested with YouTube data. The accuracy rates of the proposed CNN model, VGG16 and ResNet50 architectures are 94%, 97% and 98%, respectively.

Keywords: *Deep learning, ResNet50, Turkish folk dances, VGG16, CNN*

1. Introduction

Culture is the way people and societies perceive, make sense of and interpret themselves and their environment. As a being who changes the world and lives in this world, people are in the position of creating and processing this culture. People and societies protect their cultures in the world they have built and at every stage of their life. Folk dances are one of the most important cultural assets of a country/society.

Folk dances are a phenomenon that reflects the life of the people of a certain country or region and contains emotions such as sadness, happiness and anxiety [1]. Folk dances can be found in every type of society, from the most isolated and not acquainted with any type of modernization to the most developed community [2].

In folk dances, dancers perform regular and continuous rhythmic movements, individually or in groups, in harmony with the music. Zeybek, Horon, Halay, Çiftetelli, Bar, Spoon dances are the most common folk dances in Turkey. Zeybek, which means bravery, valor, agility and assertiveness, is a folk dance from the Aegean region. The postures and hand gestures of the dancers in the Zeybek dance contain these characters. Horon is a folk dance specific to the Eastern Black Sea region. Horon means grass that has been mown or baled. Therefore, there is a parallelism between the way the dance is performed and the dictionary meaning of the word. Horon dancers stay close to each other much more than in other folk dances. The trembling, shaking, shuddering figures in the Horon express the sea, the dying of the fish coming out of the sea. Halay is a folk dance of Eastern and Southeastern Anatolia. Halay folk dances are mostly played in East, Southeast and Central Anatolia. Dancers hold their hands and form a row and then a circle, according to the rhythm of the music, feet combinations are very important. There are many halay describing historical events, rebellion and love. Halay is a game from prehistoric times. It conveys the energy, cycle and solidarity of life to the audience. There are many different Halay dances that reveal historical events, rebellion and love. As can be understood from the descriptions of folk dances, the poses of the dancers have a semantic meaning.

It is a difficult process to automatically identify the postures of the dancers with a system and to determine which folk dance they belong to. The dancer wears special clothes that affect classification performance. The complexity of the background, camera angle and special clothes cause the pose prediction algorithms to give mistaken results [3]. Deep learning has achieved great success in providing solutions to classification problems in the field of image processing. In this study, deep learning algorithms are used to identify the postures of the dancers and to determine which folk dance they belong to. Firstly, a dataset including Zeybek, Çiftetelli, Horon and Halay dances, which are among the Turkish folk dances, was created from YouTube videos. Then a convolutional neural network (CNN) was proposed to classify them. As a result, it has been shown that a trained CNN model can classify the folk dances from dataset of YouTube videos with high accuracy.

Transfer learning refers to the reuse of a model trained for one task for another related task. The pre-trained models are created using a large dataset. In transfer learning, it is aimed to improve learning in a new task by transferring information from a learned task [4]. Transfer learning allows us to save time or get better performance. VGG16 and ResNet50 are CNN-based models with different architectures that were pre-trained using ImageNet [5]. ImageNet is an image database containing more than 14 million labeled

[1] *Sakarya University, Department of Computer Engineering, Sakarya, TÜRKİYE*
ORCID ID : 0000-0002-0212-511X
[2] *Sakarya University, Department of Computer Engineering, Sakarya, TÜRKİYE*
ORCID ID : 0000-0002-7500-4001
** Corresponding Author Email: sumeyye@sakarya.edu.tr*

images and 1000 total classes [6]. In this study, transfer learning with VGG16, ResNet50 was used.

The proposed CNN model is trained with the dataset obtained from YouTube. Optionally, the pre-trained model may need to be adapted to the target data for the respective task. VGG16 and ResNet50 models were trained on our own dataset utilizing fine tuning.

The pre-trained (VGG16 and ResNET50) models and the proposed CNN model were compared on YouTube videos. The accuracy rates of the trained models are 94%, 97% and 98% for the CNN, VGG16 and ResNet50 models, respectively.

This paper is organized as follows. Section 2 describes the prior work pertaining to the recognition of human poses and hand, arm gestures. In section 3, architectural details of the proposed CNN, VGG16 and ResNet50 models for the classification of Turkish folk dances are given. Section 4 presents a comparison of the proposed CNN-based approaches (CNN, VGG16 and ResNet50). Finally, in section 5, the article is concluded by giving a summary of the article and proposing some further extensions to the research.

2. Literature Review

Many studies have been conducted on the classification of human activities and postures [7]. Human activities can be conceptually divided into 4 different classes: gestures, actions, interactions and group activities [8]. Approaches in the 1990s were generally oriented towards describing gestures and simple actions [9]. Approaches in the 1990s were insufficient to describe and explain interactions and group activities. Then, recognition methodologies designed for complex human activities such as the interaction of the two or more people and/or objects and group activities were developed [10].

The folk dances contain complex human activities. In this review, in addition to estimating and defining the dancer’s posture and hand-arm gestures, studies on the classification of dance forms are included. Indian classical dance (ICD) videos (Bharatnatyam, Kathak and Odissi) are classified using support vector machine (SVM) with inter-section kernel [11]. An accuracy of 86.67% is achieved on ICD dataset created from YouTube videos. In [12], the SVM classifier is used to classify 2 folk dances from the Western Macedonia region. A CNN model is proposed to classify dance posture images of Indian folk dance [3]. YouTube videos and dataset produced using the Kinect sensor are used in the training phase. With the proposed CNN model, an accuracy performance of 97.22% is achieved. Transfer learning technique is utilized by using two large, labeled datasets, namely, MNIST and CIFAR-10. With the pre-trained CNN model, an accuracy rate of 99.6% is achieved. The authors in [13] present a novel framework to classify ICD forms from videos. The representations are then extracted through Deep Convolution Neural Network (DCNN) and Optical Flow. A multi-class linear SVM is used for training these representations. With the framework, an accuracy performance of 75.83% is achieved. In this paper [14], the Adaboost multiclass classifier is used to classify ICD forms from video and the Adaboost classifier gives better classification accuracy compared to AGM and SVM. The research article in [15] presents research on the recognition of classical dance mudras in India. The images of hand mudras of various classical dances are collected from the internet. Histogram of oriented (HOG) features of hand mudras are given as input to the classifier. SVM is used as the classifier. In another study [16] using the CNN model in the classification of ICD, a recognition rate of 93.33% is achieved. There are also studies using the RNN classifier in the classification of Indian dances [17]. The authors in [18] focus on an ICD form known as Bharatanatyam. The dance sequence is recognized using Hidden Markov Model (HMM). In this paper [19], researchers propose a deep CNN model using ResNet50 to classify ICD forms into eight categories. The proposed model gives an accuracy score of 0.911. Hand postures and signs provide an important clue in the classification of dances. The research article in [20] focuses on producing a model that can recognize hand gestures and gestures. A computer program is developed using python language which is used to train the model based on the CNN algorithm.

When the literature studies are examined, it is seen that the studies are mostly on the classification of ICD. There have also been studies on the classification of different folk dances. But no study has been found on the classification of Turkey's folk dances.

3. Convolutional Neural Network: The proposed Deep Learning Framework

A convolutional neural network (CNN or ConvNet) is a type of neural network with a grid-like topology that takes its name from the mathematical operation of convolution [5]. CNN has become the dominant and powerful type of network architecture in field of computer vision. A classic CNN architecture is given in

Layer

Fig. *1*.

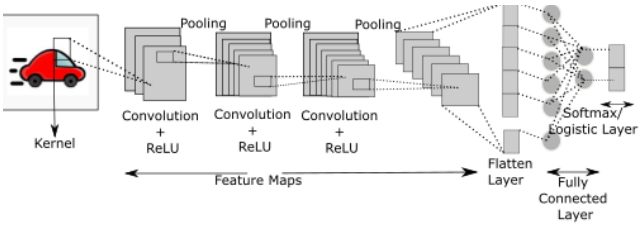


Fig. 1. Classic CNN architecture

The different layers involved in the architecture of CNN are as follows:

1) Input layer: The image data is taken from the CNN’s input layer.
2) Convolutional layer: In the convolution layer, the image is filtered with a mask.
3) ReLU: After the convolution operation, an activation function is applied. In deep learning systems, this activation function is usually ReLU. One of the reasons for using "ReLU" instead of "hyperbolic" "tangent" and "sigmoid" functions used in artificial neural networks is to observe that the system works faster [5]. In this layer, all negative pixels are converted to zero. The final output is a rectified feature map.
4) Pooling layer: A down-sampling operation is performed, which reduces the size of the feature map. A 2x2 filter is usually selected and this filter hovers over the input data. According to the selected method, values such as maximum and average are transmitted to the next layer.
5) Flatten layer: The pooled feature map produced in the pooling layer is converted into a one-dimensional vector, which will be the input layer for the artificial neural network.

6) Fully connected layer: Objects in images are identified and classified.

7) Softmax/Logistic layer: The softmax or logistic layer is located just before the output layer. The softmax or logistic function is used as the final activation function of the neural network. The logistic function is used in binary classification problems, and the softmax function is used in the expression of multiple classification problems.

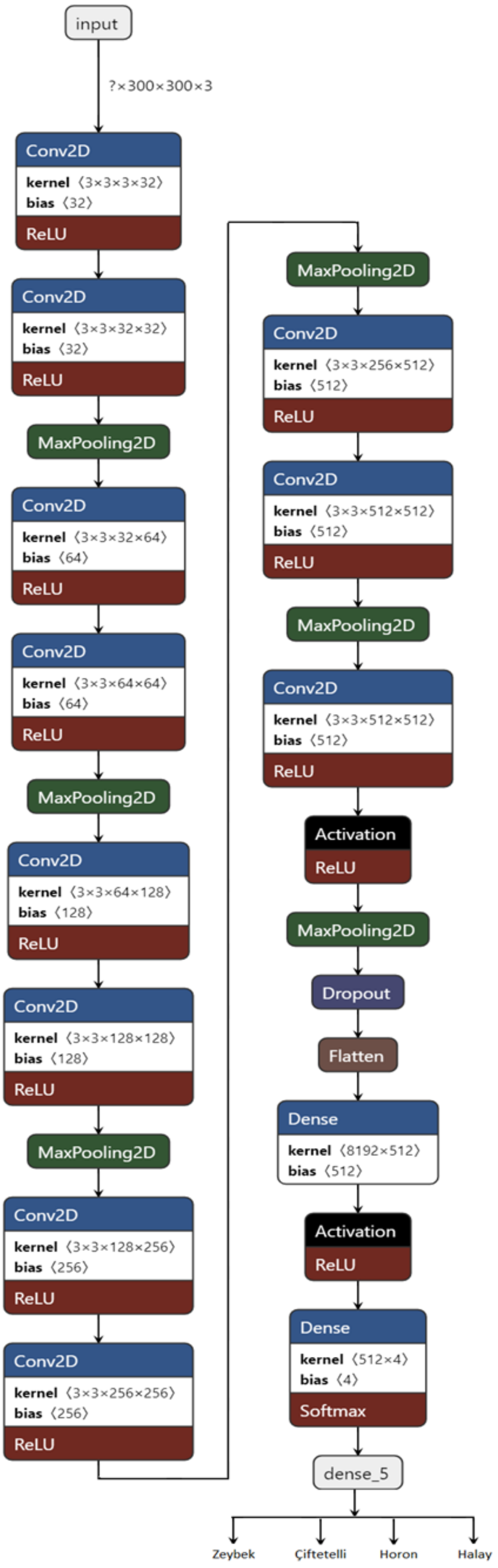


Fig. 2 Architecture of the proposed CNN model

Continuous improvements are made on the basic CNN architecture. Different CNN architectures with different number of layers have been developed.

3.1. The Proposed Architecture

The architecture of the CNN model proposed for this study is given in Fig. 2. There are many convolution layers. A convolution filter called kernel runs over all pixels of the image. The kernel size here refers to the width x height of the filter mask. A kernel size of 3x3 has been applied to all convolution layers. The input image size is set to 300×300×3 (width × height × channel). More than 3 convolution kernels are commonly used in the convolution layer in CNN models. Different convolution kernels each act as different filters. With different filters, feature maps are created. In the proposed model, the number of kernels in each convolution layer is 32, 32, 64, 64, 128, 128, 256, 256, 512, 512, and 512, respectively. The pool size in all max pooling layer is 2×2.

The flatten layer is followed by the dense layer. Dense layer is a layer that is deeply connected to its preceding layer. Each of the neurons of the dense layer is directly connected to each neuron of its preceding layer. The number of neurons belonging to each dense layer in this model (see Fig. 2) was determined as 512.

In the proposed model, the softmax function is used in the expression of multiple classifications. The softmax function converts the value to a normalized probability distribution which can be displayed to user. The number of outputs in the last layer is set to 4. The output values range from 0 to 1. The output values (estimated values) indicate that which dance the input image belongs to.

The performance of the proposed CNN model will increase with the use of a larger and diverse dataset. Many researchers prefer to use the pretrained models on a large dataset such as ImageNet, as large datasets are rarely available. VGG16 network models and ResNet50 model are pre-trained models. The pretrained models reduce errors and time required for training and increase the accuracy value.

In this study, VGG16 and ResNet50 models are used. The weights associated with these models are used directly in the app. A max_pooling, a flatten layer, two dropout functions, a dense layer consisting of 512 nodes, and a softmax layer have been added to the last layers of the ResNet50 and VGG16 models (see Fig.). The weights for these layers are determined using the training dataset created from YouTube videos. By using these developed models, it is determined to which class the dance images obtained from the YouTube video belong.

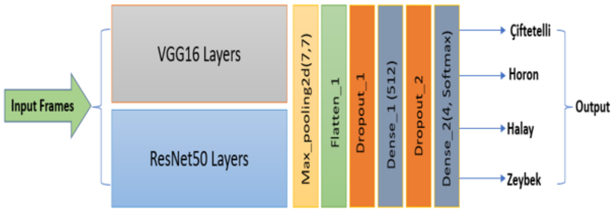


Fig. 3 ResNet50 and VGG16 models used in the study

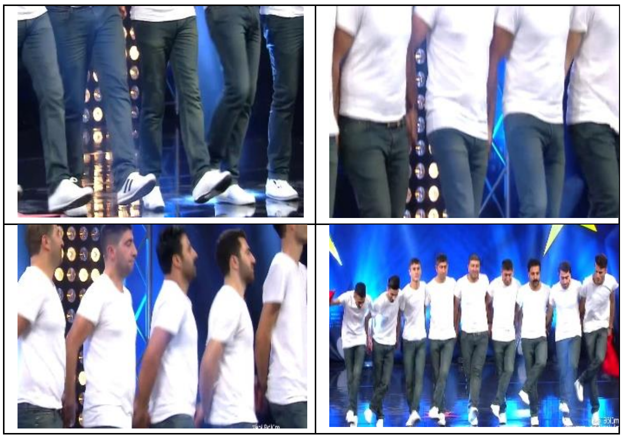
3.2. Dataset

In this paper, four types of Turkish folk dances, namely Çiftetelli, Zeybek, Horon, Halay are used. A training and validation dataset was prepared by collecting dance images from 720 pixel-sized

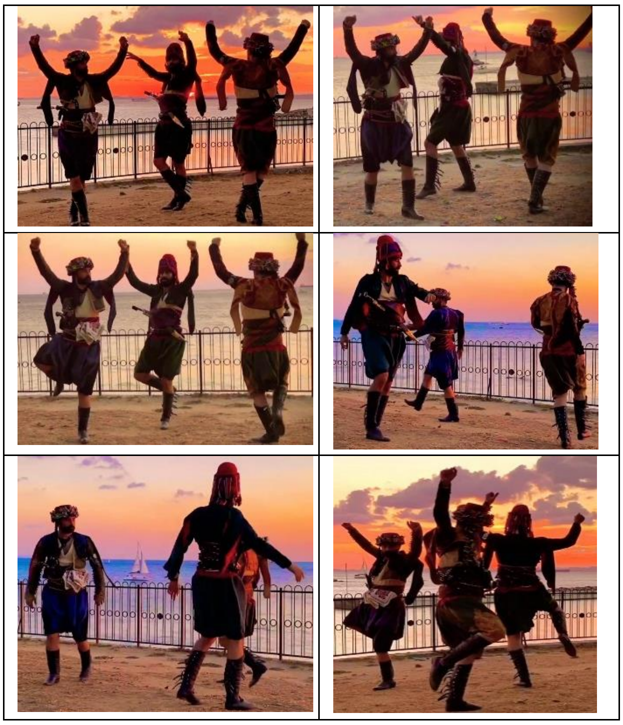
YouTube videos. The number of images belonging to each dance type is 6000. 75% of the dataset is reserved for the training and 25% is reserved for the validation. There are 4500 images belonging to each of the dance types in the training dataset. There are 1500 images belonging to each of the dance types in the validation dataset. There are 18000 images in the training folder and 6000 images in the validation folder. Sample images of the dataset are given in Figure 4.

(a)

(b)



(c)



(d)

Fig. 4 Sample images of the dataset. (a) Çiftetelli folk dance. (b) Horon folk dance. (c) Halay folk dance. (d) Zeybek folk dance.

4. Experimental results

In this study, deep learning models are developed using Keras library. The proposed CNN model is trained with the training dataset. Pre-trained VGG16 and ResNet50 architectures using the ImageNet database were fine-tuned with the training dataset.

The accuracy and loss function graphs of the proposed CNN model are shown in Fig. . As seen in Fig. , the training accuracy and the validation accuracy are almost same. In other words, there is no overfitting. The metrics in the training set provide information about how the model is progressing in terms of training. The metrics in the validation set provide insights into the quality of the model - how good it is at making new predictions based on data it hasn't seen before. The training loss and validation loss curves show a downward trend. Validation accuracy rate has reached 94% and the loss value has decreased from 1.4 to less than 0.34.

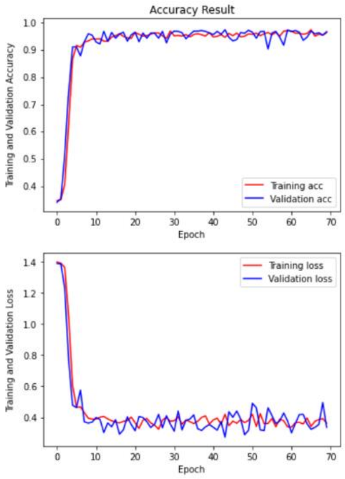


Fig. 5. Training and validation accuracy graph (above) and training and validation loss graph (below)

The proposed CNN model was tested with 120 dance images from YouTube videos for each dance genre. The confusion matrix of the CNN model is given Fig. 6.

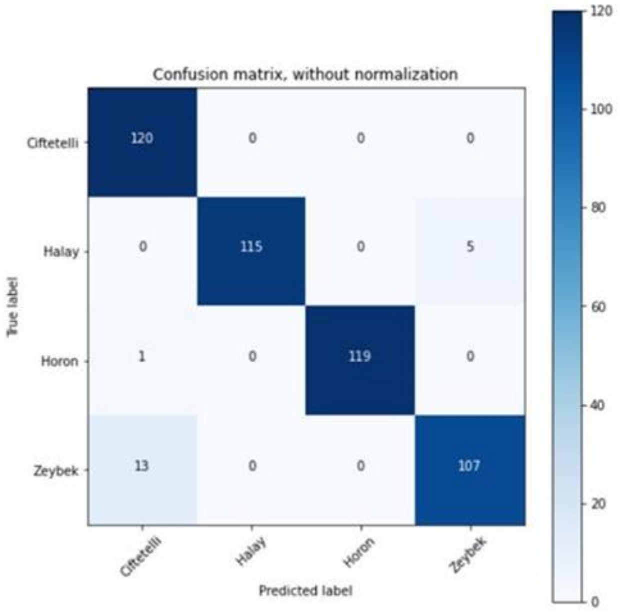


Fig. 6. The confusion matrix of the CNN model

As seen in the confusion matrix, the CNN model correctly predicted all 120 images of the Çiftetelli dance. The model estimated 115 of the 120 Halay dance images as correct and 5 as Zeybek dances.

The elements of the confusion matrix are utilized to find parameters named accuracy, sensitivity (recall), precision and f1-score. The performance of the models has been tested with these metric functions.

Precision quantifies the number of correct positive predictions made. The accuracy metric shows the performance of the model in all classes. The sensitivity gives a measure of how many positive samples the model correctly predicted. The value of the sensitivity affects the accuracy value. The higher the sensitivity, the model is the better in correctly identifying the positive cases. F1-score value is the harmonic average of precision and sensitivity values. The f1 score value gives more objective results in data sets that are not evenly distributed. It is a measure of how well the classifier is performing and is often used to compare classifiers.

Table 1. Performance metrics of models for each class

		Çiftetelli	Horon	Zeybek	Halay
CNN	Precision	0.90	1	0.96	1
	recall	1	0.99	0.89	0.96
	F1-score	0.94	0.99	0.92	0.98
VGG16	Precision	0.98	0.97	0.96	0.99
	recall	0.99	0.95	0.93	0.98
	F1-score	0.98	0.96	0.94	0.98
ResNet50	Precision	0.99	0.99	0.97	0.98
	recall	0.97	0.96	0.98	0.99
	F1-score	0.98	0.97	0.97	0.98

The performance evaluation of the CNN, VGG16 and ResNet50 models for each class is given in Table 1. In the CNN model, the most correct predictions were made in the Horon folk dance. In the CNN model, the most incorrect predictions were made in the Zeybek folk dance. The performance behavior of the VGG16 model for Çiftetelli and Halay folk dances is very similar. In the VGG16 model, the most incorrect predictions were made in the Zeybek folk dance. In the ResNet50 model, Çiftetelli folk dance has the highest f1-score.

As can be seen in Fig. 4, the postures of dancers in the Zeybek and Çiftetelli folk dances are similar to each other. The best performing model in Zeybek folk dance is ResNet50.

Table 2. Comparison of models

CNN	Accuracy	0.95
	Validation accuracy	0,94
VGG16	Accuracy	0,96
	Validation accuracy	0,97
ResNet50	Accuracy	0,98
	Validation accuracy	0,98

The accuracy metric shows the correct predictive performance of the model on real-world data or test data. Validation accuracy indicates the correct prediction performance of the model on the validation dataset. The accuracy metric provides clues about how ready the model is for the real world. The accuracy and validation accuracy values of the CNN, VGG16 and ResNet50 models are given in Table 2. ResNet50 model showed better accuracy performance than other models.

It is not enough to examine the accuracy and validation accuracy values alone. In addition to the accuracy metric, the values of performance metrics such as precision, recall, f1-score should also be evaluated. F1-score is a widely used reliable metric for comparing models. Model selection should be made by evaluating performance metrics in a way specific to the problem. For example, precision is important when the cost of false positive is high.

The performance metrics of the models developed in this application were evaluated. Precision and recall metrics were evaluated under the F1-score metric. The ResNet50 model has better performance in model evaluation with the F1-score metric.

5. Conclusion

Folk dances are one of the most important cultural values that transfer the feelings and life of a society to generations. In folk dance, dancers perform dance-specific hand, arm and foot movements and they wear folk dance costumes. It is difficult process to automatically identify the dancer's movements with a system and to determine which folk dance they belong to. The complexity of the background, camera angle, and special clothing cause the prediction algorithms to give erroneous results. Deep learning has achieved great success in classification problems in image processing. In this study, deep learning methods were used in the classification of Turkish folk dances. There are many studies in the literature on the classification of folk dances. However, it was not encountered to a study in the literature on the classification of Turkish folk dance. Zeybek, çiftetelli, horon and halay dances which are among Turkish folk dances were selected. 6000 images of each type were collected from YouTube videos. The developed CNN model was trained with 18000 images. Pre-trained VGG16 and ResNet50 architectures using the ImageNet database were fine-tuned with the training dataset. The proposed CNN model, VGG16 and ResNet50 models were compared with various performance metrics. Accuracy and f1-score metrics, which are frequently preferred to compare the performances of the models were used to selection of the best model. The ResNet50 model gave the best performance with an accuracy rate of %98. The results of the model are promising. We believe that this study will guide the studies in this field. It also contributes by sharing the data set. In our further work, we aim to expand the dataset and focus on the classification of different Turkish folk dances by using different deep learning techniques.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1]. T. Eroğlu, Doğu ve Güneydoğu anadolu'da halk oyunları ve halayların incelenmesi, vol. 1, Ankara: Kılıçaslan matbaacılık, 1995.

[2]. S. Aydın, "Halk Oyunları," Kültür Turizm Bakanlığı, Ankara, 2009-Kasım.

[3]. Vanitha, D. D. . (2022). Comparative Analysis of Power switches MOFET and IGBT Used in Power Applications. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 01–09. https://doi.org/10.17762/ijrmee.v9i5.368

[4]. Aparna Mohanty, Pratik Vaishnav, Prerana Jana, Anubhab Majumdar, Alfaz Ahmed, Trishita Goswami and Rajiv R. Sahay, "Nrityabodha: Towards understanding Indian classical dance using a deep learning approach," Signal Processing: Image Communication, p. 529–548, 2016.

[5]. K. Fırıldak and M. F. Talu, "Evrişimsel Sinir Ağlarında Kullanılan Transfer Öğrenme Yaklaşımlarının İncelenmesi," Anatolian Journal of Computer Science, vol. 4, no. 2, pp. 88-95, 2019.

[6]. Ş. Kılıç, "Derin öğrenme yöntemleri kullanılarak giyilebilir sensörlerden kişi tanıma," Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2021.

[7]. Bulla, P. . "Traffic Sign Detection and Recognition Based on Convolutional Neural Network". International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 43-53, doi:10.17762/ijritcc.v10i4.5533.

[8]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," Int J Comput Vis, pp. 211-252, 2015.

[9]. Weilong Yang, Yang Wang and Greg Mori, "Recognizing Human Actions from Still Images with Latent Poses," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Canada, 2010.

[10]. J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," ACM Comput. Surv., vol. 43, no. 3, 2011.

[11]. J. K. aggarwal and Q. Cai, "Human motion analysis: A review," Comput. Vision Image Understand, pp. 428-440, 1999.

[12]. Bangpeng Yao , Aditya Khosla and Li Fei-Fei, "Classifying Actions and Measuring Action Similarity by Modeling the Mutual Context of Objects and Human Poses," in Proc. 28th Int. Conf. Mach. Learn. ICML , 2011.

[13]. Soumitra Samanta, Pulak Purkait and Bhabatosh Chanda, "Indian Classical Dance classification by learning dance pose bases," in IEEE Workshop on Applications of Computer Vision (WACV), USA, 2012.

[14]. Ioannis Kapsouras, Stylianos Karanikolos, Nikolaos Nikolaidis and Anastasios Tefas, "Folk dance recognition using a bag of words approach and ISA/STIP features," in BCI '13: Proceedings of the 6th Balkan Conference in Informatics, Greece, 2013.

[15]. Ankita Bisht, Riya Bora, Goutam Saini, Pushkar Shukla and Balasubrmanian Raman, "Indian Dance Form Recognition from Videos," in International IEEE Conference on Signal-Image Technologies and Internet-Based System, Jaipur, India, 2017.

[16]. Agarwal, D. A. . (2022). Advancing Privacy and Security of Internet of Things to Find Integrated Solutions. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(2), 05–08. https://doi.org/10.17762/ijfrcsce.v8i2.2067

[17]. K. V. V. Kumar, P. V. V. Kishore and D. Anil Kumar, "Indian Classical Dance Classification with Adaboost Multiclass Classifier on Multifeature Fusion," Mathematical Problems in Engineering, 2017.

[18]. K. V. V. Kumar and P. V. V. Kishore, "Indian Classical Dance Mudra Classification Using HOG Features and SVM Classifier," Smart Computing and Informatics, vol. 77, pp. 659-668, 2018.

[19]. P. V. V. Kishore, K. V. V. Kumar, E. Kiran Kumar, A. S. C. S. Sastry, M. Teja Kiran, D. Anil Kumar and M. V. D. Prasad, "Indian Classical Dance Action Identification and Classification with Convolutional Neural Networks," Advances in Multimedia, 2018.

[20]. Swati Dewan, Shubham Agarwal and Navjyoti Singh, "A deep learning pipeline for Indian dance style classification," in Tenth International Conference on Machine Vision (ICMV 2017), Vienna, Austria, 2018.

[21]. Agarwal, A. . (2022). Symmetric, e-Projective Topoi of Non-Solvable, Trivially Fourier Random Variables and Selberg's Conjecture. International Journal on Recent Trends in Life Science and Mathematics, 9(1), 01–10. https://doi.org/10.17762/ijlsm.v9i1.136

[22]. T. a. D. P. a. M. A. Mallick, "Posture and sequence recognition for Bharatanatyam dance performances using machine learning approach," Preprint, 2019.

[23]. Kabisha, M. S., Rahim, K. A., Khaliluzzaman, M., & Khan, S. I. (2022). Face and Hand Gesture Recognition Based Person Identification System using Convolutional Neural Network. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 105–115. https://doi.org/10.18201/ijisae.2022.273

[24]. Nikita Jain, Vibhuti Bansal , Deepali Virmani , Vedika Gupta, Lorenzo Salas-Morera and Laura Garcia-Hernandez, "An Enhanced Deep Convolutional Neural Network for Classifying Indian Classical Dance Forms," Applied Sciences, vol. 11, no. 14, 2021.

[25]. N. A. Libre. (2021). A Discussion Platform for Enhancing Students Interaction in the Online Education. Journal of Online Engineering Education, 12(2), 07–12. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/49

[26]. D. Bhavana, K. Kishore Kumar, Medasani Bipin Chandra, P.V. Sai Krishna Bhargav, D. Joy Sanjana and G. Mohan Gopi, "Hand Sign Recognition using CNN," International Journal of Performance Analysis in Sport, vol. 17, no. 3, pp. 314-321, 2021.