Knowledge Engineering and Data Science (KEDS)
Vol 7, No 1, April 2024, pp. 27–39

pISSN 2597-4602
eISSN 2597-4637

Convolutional Neural Network in Motion Detection for Physiotherapy Exercise Movement

Dika Fikri Laistulloh [a,1,*], Anik Nur Handayani [a,2], Rosa Andrie Asmara [b,3], Phillip Taw [c,4]

[a] Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang
Jl. Semarang no. 5, Malang 65145, Indonesia
[b] Informatics Engineering Study Program, Information Technology Department, State Polytechnic of Malang
Jl. Soekarno Hatta No.9, Malang 65141, Indonesia
[c] Regional English Language Office (RELO) U.S. Department of State
2201 C Street, NW, Washington, D.C. 20520, United States
[1] dika.fikri.2105348@students.um.ac.id *; [2] aniknur.ft@um.ac.id; [3] rosa.andrie@polinema.ac.id;
[4] phillip.taw.cpeli@gmail.com
* corresponding author

ARTICLE INFO

Article history:
Received 07 November 2023
Revised 04 March 2024
Accepted 15 March 2024
Published online 02 May 2024

Keywords:
Physiotherapy Exercise Movement
Cerebral palsy
Deep learning
CNN

ABSTRACT

Physiotherapy focuses on movement and optimal utilization of the patient's potential. Exercise Therapy is a physiotherapy procedure that specifically focuses exercises on active and passive movements. Cerebral Palsy (CP) patients are one of the sufferers of motor disorders of the upper extremities. Cerebral Palsy (CP) patients suffer from disorders in motor functions of the upper extremities. Physiotherapy Exercise Movement has 4 categories of movement exercises for the therapy of people with upper extremity body disorders: Elbow flexor strengthening in sitting using free weights, lifting an object up, reaching diagonally in sitting, and reaching from a low surface to a high surface. By taking 4 categories of motion movements in exercise therapy, data were taken using normal child subjects as standard movements, which then became a reference for CP child therapy. The limitations of therapy in physical care prompted researchers to investigate the use of image processing as input to Human Computer Interaction (HCI) in the process of motion detection-based therapy. In research using Deep learning as a classifier, namely using the CNN Model (Inception V3, Resnet152, and VGG16 architectural models). The results obtained by the CNN (Inception V3) model have the best performance with an accuracy percentage of 98%.

I. Introduction

Physical therapy is a form of treatment that includes exercises using special equipment to help the process of recovering, maintaining, and improving physical abilities [1]. Physical therapy plays a role in improving the quality of life of patients by improving their abilities so that the ability to independence increases [2]. Exercise therapy is one of the physiotherapy measures that specifically focuses exercise on active or passive movement or physical activities [3]. This training is carried out in a planned [4], systematic and repeatable manner with several iterations. Children with upper extremity motor movement disorders experience gross motor function disorders that limit movement, especially in upper extremity body parts [5]. This exercise therapy can help restore motor function by repeatedly performing movements that provide appropriate and correct information to a brain that has a motor function disorder [6].

Cerebral Palsy (CP) patients are one of the sufferers of motor disorders of the upper extremities. Children with Cerebral Palsy are vulnerable sufferers because patients are at age conditions during growth and development [4]. Patients have permanent neurological disorders that cause impaired gross motor function, fine motor, speech, and other disorders [7]. This disorder can affect the coordination function of his movements and one of them is in the motion of the upper body extremities [8]. Physical therapy efforts will be more effective if it can be done to CP sufferers who are still in the category of children [2]. This is in line with the growth process at an early age which

https://doi.org/10.17977/um018v7i12024p27-39

greatly supports therapeutic efforts by training muscles to reduce the effects of CP symptoms [9]. Ajeng Probowati explained that physical therapy can be very effective in managing symptoms and improving the quality of life of patients with cerebral palsy [1]. Performing movement exercise therapy in cerebral palsy children will be able to improve the functional ability of motor movements that are impaired and prevent further disorders [10].

Physiotherapy Exercise Movement has 4 categories of movement exercises for the therapy of people with upper extremity body disorders. The movement category is elbow flexor strengthening in sitting using free weights, lifting an object up, reaching diagonally in sitting and reaching from a low surface to a high surface [11]. Researchers have a video data source of Physiotherapy Exercise Movement motion detection results. By taking 4 categories of motion movements in exercise therapy videos, data collection uses normal child subjects as standard movements to then become a reference for CP child therapy. During the therapy process, there will definitely be movements that are not in accordance with the therapeutic movement guidelines. Among them are differences in angle, position, and type of movement that are not suitable and even random movements occur, so classifying therapeutic categories is needed.

The limitations of therapists in carrying out physical care encourage researchers to develop therapeutic media that support the therapeutic process by utilizing Human Computer Interaction (HCI) [12]. An interesting therapy process can increase the motivation of patients and encourage enthusiasm to undergo the therapy process. The use of image processing as input to HCI is important to be used in this study in the process of motion detection-based therapy as an effort for an interesting therapeutic process [13]. Video motion therapy is obtained using a kinect sensor that produces video with joint point detection results, coordinates and depth sensor results specifically for right arm movement [14]. This video takes movement data of normal children with an age range of 11-12 years. This movement is taken so that it is used as a movement standard that is suitable for CP child therapy. Furthermore, in the previous study [11], a classification process was carried out using Machine Learning (ML) to be applied to robot actuators as a therapeutic medium with the K-Nearest Neighbor (KNN) Algorithm for a database of numerical values on the video information.

Deep learning (DL) is an Artificial Intelligence (AI) method, one of which is successfully used in Human Action Recognition (HAR) [15]. DL's mathematical architecture resembles a very complex and deep neural network architecture when compared to ML [16]. Deep learning has a Convolutional Neural Network (CNN) architecture that produces the output of important information from an image [17]. A study in Egypt applied the DL method in classifying Human Action Recognition (HAR) datasets and showed significant accuracy results. From this study, the highest accuracy results were obtained at 94.87% [18]. Deep learning in the subject of HAR also showed <90% accuracy results in human abnormal habit detection studies on real-time video. The study also describes spatial, temporal feature extraction, various datasets, augmentation processes [19]. The application of Deep learning models on HAR using wearable sensor data resulted in suggestively better accuracy of 90.44% for CNN than other neural network models such as Inception Time and DeepConvLSTM [20]. The novelty proposed in this study is the implementation of deep learning methods to produce the best accuracy level values in the physiotherapy motion detection video classification process using CNN.

II. Methods

This research goes through several stages which include: data collection, data preprocessing, classification model, and evaluation. The stages of research can be seen in Figure 1.

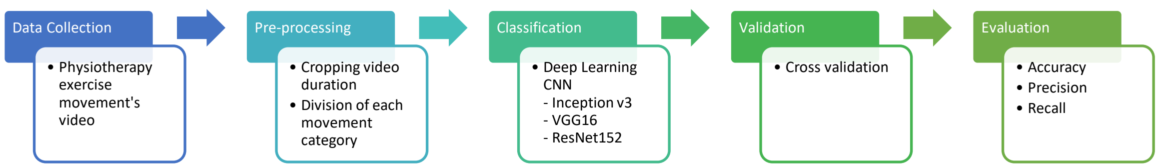


Fig. 1. Research flow diagram

The coding process that contains the core flow model is used such as the feature extraction process, defining the video into sequences, as well as the deep learning and evaluation model process such as in Pseudocode 1.

PSEUDOCODE 1. Hyperparameter tuning using grid search

```
# 1. Data Collection
# Collect training data and test data
train_data, test_data, train_labels, test_labels = load_data()

# 2. Pre-processing Data
# Data normalization and other pre-processing
train_data, test_data = preprocess_data(train_data, test_data)

# 3. Classification using Deep Learning
# Build a CNN model with the desired architecture
model = build_cnn_Inceptionv3_model()
model = build_cnn_resnet152_model()
model = build_cnn_VGG16_model()

# 4. Validation
# Separate validation data from training data
train_data, val_data, train_labels, val_labels =
split_train_validation(train_data, train_labels)

# Model training
train_model(model, train_data, train_labels)

# 5. Evaluation
# Test the model on test data
test_accuracy = evaluate_model(model, test_data, test_labels)

# 6. Best Model
# Choosing the best model (e.g., based on test accuracy)
If test_accuracy > best_accuracy:
    best_model = model
    best_accuracy = test_accuracy

# Save best Model
save_model(best_model)
```

A. Data Collection

Based on physiotherapy exercise movement, there are four categories of movements that will be included in the deep learning process. Four movement therapy recommendations for upper extremity therapy are, elbow flexor strengthening in sitting position using free weights, lifting an object up, reaching diagonally in sitting and reaching from a low surface to a high surface. This is shown in detail in Table 1, with illustrative images of the Physiotherapy Exercise Movement recommendations.

This parameter corresponds to the dataset obtained according to the 4 recommended upper limb therapy movements as in the study of the application of virtual reality for upper limb. This movement is also used for cerebral palsy therapy media in gesture detection research using machine learning.

The dataset to be studied is a video of the results of data collection of the movements of 6 children with normal conditions who are still in elementary school aged 11-12 years with 4 males and 2 females. This video dataset is a video that has been done in previous research [11], so it focuses on the advanced process of the video and does not specifically discuss the work sequence, movement rules and procedures when shooting motion videos. From 6 video sets of 4 Physiotherapy Exercise Movement movements performed 2-3 repetitions in one video dataset. Video datasets have a total duration of between 1-3 minutes. The video has a resolution of 1920 x 1080 pixels with the fps standard producing frames per second of 24 frames which more details can be observed in Table 2.

Table 1. Classified motion and sample per frame

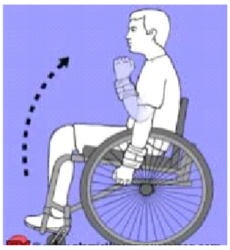
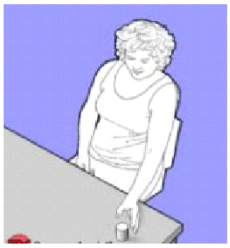
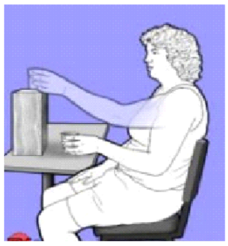
Number	Definition	Labels	Images
1	Elbow flexor strengthening in sitting using free weights	Movement 1	
2	Lifting up an object	Movement 2	
3	Reaching diagonally in sitting	Movement 3	
4	Reaching from a low surface to a high surface	Movement 4	

Table 2. Dataset classification

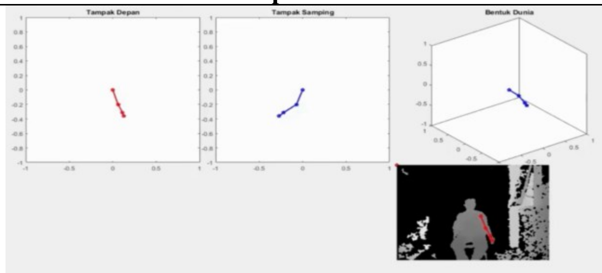
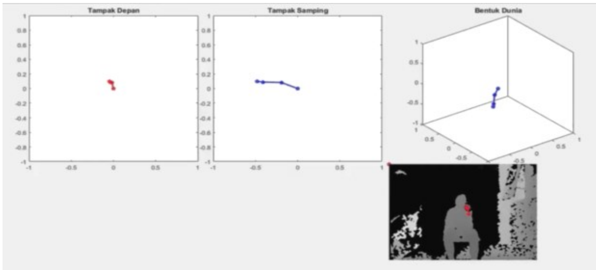
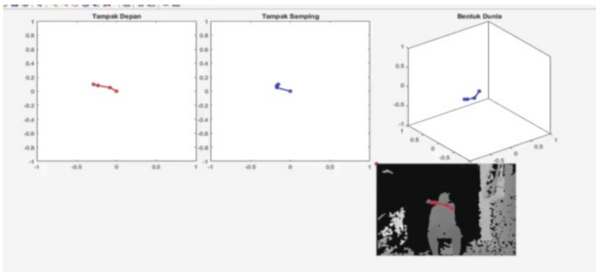
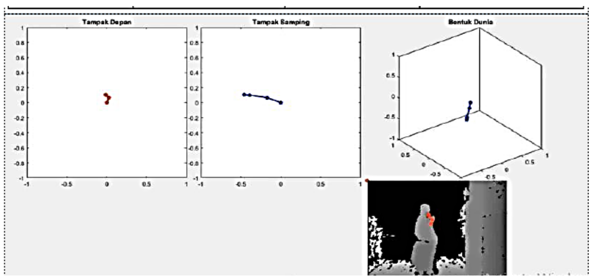
Video	Name	Video Duration (Minutes)	Video Duration (Second)	Frames Video
1	Andika	1.41 m	84.6 s	2,016 frame
2	Bagas	4.31 m	258.6 s	6,192 frame
3	Farel	3.18 m	190.8 s	4,560 frame
4	Naura	6.27 m	376.2 s	9,024 frame
5	Ryand	3.25 m	195 s	4,680 frame
6	Sinta	3.28 m	196.8 s	4,704 frame

B. *Pre-processing*

The visual video dataset is the output of the matlab GUI v3 tool which displays video with visual joint coordinate information [3] on the front, side and 3d coordinate displays, as well as visual information from the depth sensor. As per the sample frame in the image, the video specification has mp4 video format [4] and a frame size of 1920 x 1080 pixels. Table 3 presents the sample frames video observation each movement.

The dataset preprocessing process is carried out by equalizing the size of the frame dimensions [5] and the duration of time on each movement to be equalized. The main video dataset is about 60-120 seconds long containing several repetition movements in a single video. Then cropping [6] per repetition of the movement into 4-5 seconds of core movement video. Furthermore, for every one movement per child cropping is done manually with video editing software illustrated in Figure 2.

Table 3. Sample frames video observation each movement

Labels	Sample Frame
Ssample frame Movemet 1	
Sample frame Movement 2	
Sample frame Movement 3	
Sample frame Movement 4	

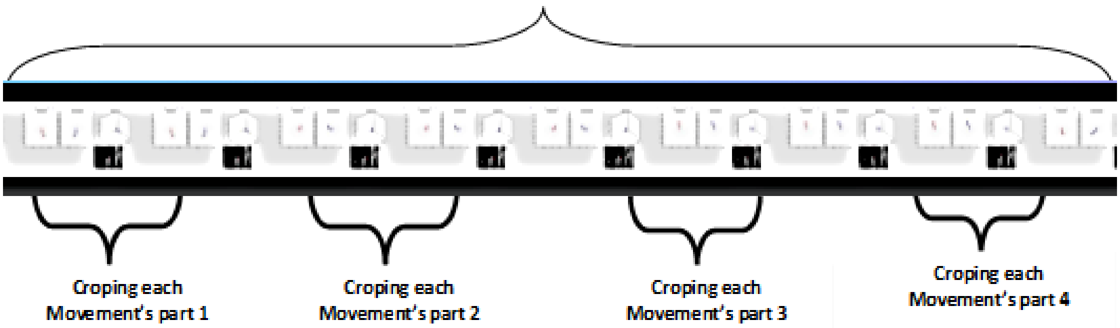
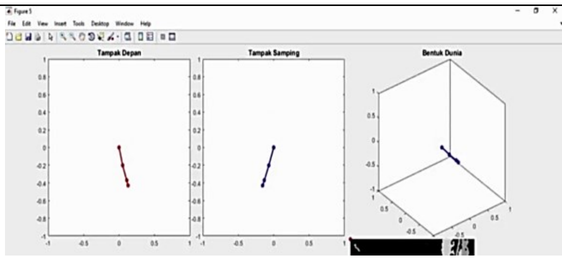
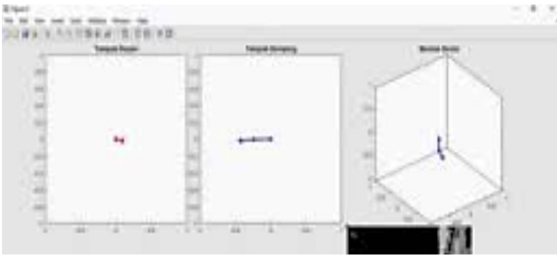
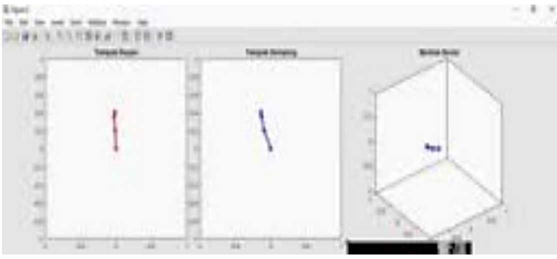
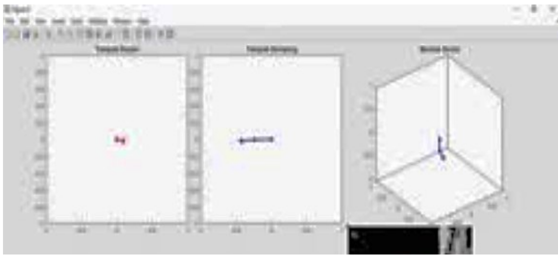
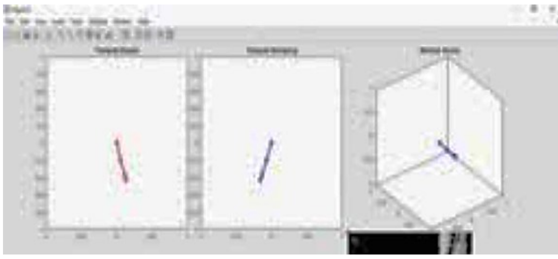


Fig. 2. Croping duration each movement

Manual cropping is done because each movement category in the therapy video has a different start point for each child, and each movement. The cropping video can then be easily categorized per one therapeutic movement for further classification process.

In detail, the video dataset has a duration of 1-2 minutes as shown above, which will be cut per frame with a range of 4-5 seconds to take core movements from the 4 groups of movements studied in each initial video dataset. In a single data set, a set of videos consists of more than one repeated core movement. The sample video frames per motion (Movement 1) present in Table 4.

Table 4. Sample video frames per motion (Movement 1)

Labels	**F(second)**	**Frames ilustration**
Movement 1 (elbow flexor strengthening in sitting using free weights)	F1	
	F2	
	F3	
	F4	
	F5	

After the cropping process per duration and according to the video category, the final video data result produced 93 videos with the division of the number of movements, namely movement 1 = 24 videos, movement 2 = 22 videos, movement 3 = 22 videos and movement 4 = 25 videos. For every 5 second videos that can be observed in Table 5.

Table 5. Result after cropping

Video	Name	After Cropping											
		Movement 1			Movement 2			Movement 3			Movement 4		
		Video	Second	Frame	Video	Second	Frame	Video	Second	Frame	Video	Second	Frame
1	Andika	4	20	480	5	25	600	5	25	600	5	25	600
2	Bagas	4	20	480	3	15	360	3	15	360	4	20	480
3	Farel	4	20	480	3	15	360	4	20	480	5	25	600
4	Naura	4	20	480	3	15	360	4	20	480	4	20	480
5	Ryan	4	20	480	4	20	480	4	20	480	4	20	480
6	Sinta	4	20	480	4	20	480	2	10	240	3	15	360
Total		24	120	2880	22	110	2640	22	110	2640	25	125	3000

C. Classification

Deep Learning (DL) is used to perform tasks such as object recognition, speech recognition, prediction, and data classification [21]. DL allows the system to learn independently from a given dataset or training data [22]. The increase in performance is in line with the increasing amount of data processed. Deep learning in the subject of HAR also showed accuracy results of <90% in human abnormal habit detection studies on real time video [23]. The study also describes spatial, temporal feature extraction, various datasets, augmentation processes. CNN is included in the DL field contained in the ML subfield [24]. This field applies the basic concepts of neural network algorithms with more layers [25]. CNN is a feedforward network because information flows in only one direction from input to output. CNN is very popular and performs well in areas such as computer vision and image processing [26]. The base layer architecture of CNN shows in Figure 3.

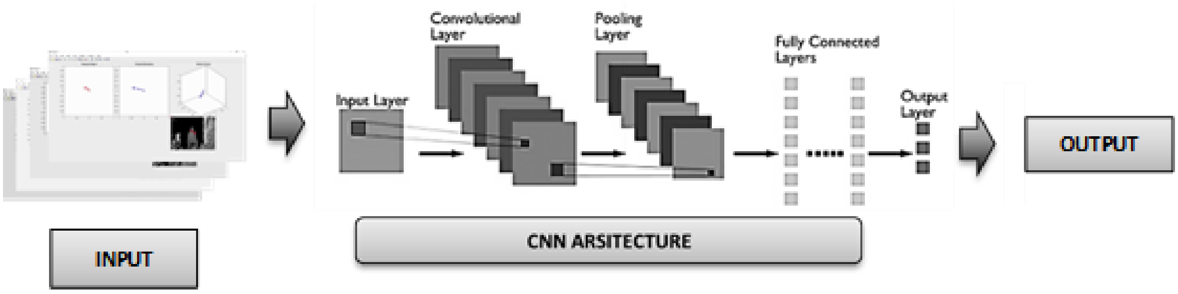


Fig. 3. Base layer CNN

The model training process uses online-based software with Google Collab (GC) which is a cloud-based platform used for computer science project experiments and other science experiments [13]. GC uses an efficient, flexible and effective python language in the latest programming tools [27]. The applied classification is a sequential binary classification of the 1st, 2nd, 3rd and 4th categories of movement.

The process of determining the best model for CNN uses Inception V3, Resnet 152, and VGG16 [28][29][30] architectures. Of the 3 combinations of models and architectures, it is used for comparative analysis of which model's performance value is the best. The selection of 3 architectural models is based on the complexity and quantity of the number of layers of each model. Inception V3 has about 48 layers, generally consisting of a convolution layer, batch normalization, average pooling layer, and a Dense layer (fully connected) [25]. ResNet-152 has a total of 152 layers, which generally consist of many residual blocks containing a convolution layer, batch normalization, activation layer, as well as a global average pooling layer and a fully connected Dense layer at the end [31]. VGG16 is an architecture that has a total of 16 layers, including a convolution layer, a Max Pooling layer, and a Dense (fully connected) layer [32]. The arrangement of layer structure can be seen in Table 6.

Table 6. Result after cropping

Model Inception-V3 (48 Layers)	Model Resnet152 (152 Layers)	Model VGG16 (16 Layers)
	Input	
(Conv2D)-(111, 111, 32)	(ZeroPadding2D)-(230, 230, 3)	(Conv2D)-(224, 224, 64)
(Normalization)-(111, 111, 32)	(Conv2D)-(112, 112, 64)	(Conv2D)-(224, 224, 64)
(Activation)-(111, 111, 32)	(Normalization)-(112, 112, 64)	(MaxPooling2D)-(112, 112, 64)
(Conv2D)-(109, 109, 32)	(Activation)-(112, 112, 64)	(Conv2D)-(112, 112, 128)
(Normalization)-(109, 109, 32)	(ZeroPadding2D)-(114, 114, 64)	(Conv2D)-(112, 112, 128)
(Activation)-(109, 109, 32)	(MaxPooling2D)-(56, 56, 64)	(MaxPooling2D)-(56, 56, 128)
(Conv2D)(109, 109, 64)	(Conv2D)-(56, 56, 256)	(Conv2D)-(56, 56, 256)
(Normalization)-(109, 109, 64)	(Normalization)-(56, 56, 256)	(Conv2D)-(56, 56, 256)
(Activation)-(109, 109, 64)	(Conv2D)-(56, 56, 256)	(MaxPooling2D)-(28, 28, 256)
.	.	.
.	.	.
.	.	.
(Activation)-(5, 5, 192)	(Dropout)-(512)	(Dropout)-(512)
(Concatenate)-(5, 5, 2048)	(Dense)-(26)	(Dense)-(26)
	Output	

D. Validation

The deep learning model used in this research will be tested and validated with model parameters adjusted to match the training data. Cross-validation is an important option to avoid problems when testing-validation by testing K times on data. The classification scheme to be carried out in this study is with the parameters in Table 7.

Table 7. Setting parameter

Parameter	Value
Img_size	224 pixel
Batch_size	5, 10, 15
Max_seq_length	120 frame
Num_features	2048
Epoch	10, 50, 100

The main scheme applied is to use epoch 100, epoch 200 and epoch 300. From the total dataset of 93 videos, an 80:20 ratio was applied for training and testing so that 74 videos were obtained training data, and 19 video testing data. The training process was carried out sequentially from Inception V3, Resnet152, and VGG16 Architectures. With a small dataset quantity, experimental data is obtained by iterating as many as k-3 times the process of taking each model, and each scheme applied.

E. Evaluation

Confusion matrix was used in the process of evaluating deep learning training models for classification tasks performed. By describing how well the model predicts true or false outcomes, a loss matrix helps calculate important evaluation matrices such as accuracy [18] as in (1), precision as in (2), and recall as in (3).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (1)$$

$$Precission = \frac{TP}{TP+FP} \times 100\ \% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

Accuracy is used to analyze the results of prediction values actually. Recall is used to analyze the success performance of the system in recovering information. Precision is used to analyze the value of the level of accuracy between the actual information and the answers given by the system. This allows us to understand the performance of the model by selecting metrics that correspond to the purpose of a particular classification task in more detail. The complete arrangement can be seen in Figure 4.

Class		Actual Class	
		Positive	Negative
Predicated Class	Positive	TP	FP
	Negative	FN	TN

Fig. 4. Classified movement

The contents of the table above are explained as follows, confusion Matrix has several terms including True Positive (TP) which is the amount of positive data that is correctly classified by the system, True Negative (TN) which is the amount of negative data that is correctly classified by the system, False Positive (FP) which is the amount of positive data but misclassified by the system, False Negative (FN) which is the amount of negative data but classified incorrectly by the system.

III. Result and Discussion

In this chapter, the results of the comparative model to determine the best model are discussed using the CNN Deep Learning model with 3 CNN parameters, namely Inception V3, ResNet 152, and VGG16 as presents in Table 8 to Table 10. The following data shows the best accuracy test results on inception V3, Resnet 152, and VGG16 architecture models, with batch size parameters of 5, 10, 15 with epoch numbers of 100, 50, and 10.

Table 8. Setting parameter

No	Batch size/Epoch	Accuracy	Precission	Recall
1	5/100	98.00%	98.33%	98.00%
2	10/50	82.33%	79.67%	82.33%
3	15/10	57.00%	50.67%	57.00%

Table 8 shows an average accuracy result of 98% with epoch 100 and batch size 5. This result is in line with the decreasing loss and increasing accuracy value. At epoch 50 shows an accuracy of 82% and maximum results in batch size 10 and epoch 50. For precision, results, and recall show values of 82%, 79%, 82%. Furthermore, at epoch 10 and batch size 15, the maximum results obtained only reached 57% on accuracy, precision and recall. Graphs the best model training results of Inception V3 at batc size 5 and epoch 100 can be seen in Figure 5.

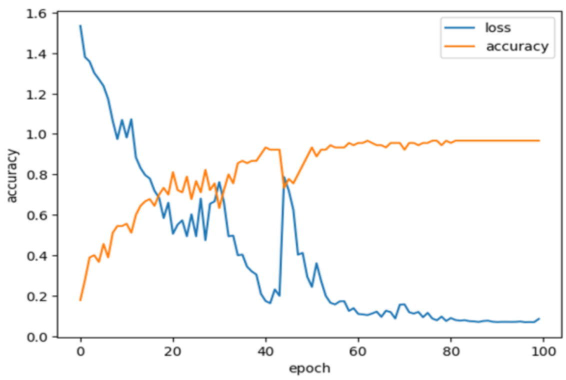


Fig. 5. Inception-v3 best result graph (Epoch 100)

From Figure 5, the results of precision, and recall are also in line with the accuracy results which reach an average value of up to 98% as well. The Loss chart shows a gradual decline along with increasing accuracy in the epoch range 1-20. In the epoch range of 20-40 there is a process of cutting the loss and accuracy chart so that the accuracy of epochs above 50 increases. The cutting process shown on the graph undergoes several cuts because the graph is the average result of 3 tests with different value results when the curve cuts the curve. In substance, towards epoch 100, the accuracy value is getting better and the loss is decreasing. The anomaly in the epoch range of 20-40 charts above arises because early stopping is not applied during the training and testing process, so that the chart fluctuates until the entire epoch tested. When using early stopping, when the accuracy results are good, it will stop so that chart fluctuation anomalies will not appear. Table 9 show the Resnet152 best result and the graph of the best result in Resent152 is present in Figure 6.

Table 9. Resnet152 best result

No	Batch size/Epoch	Accuracy	Precission	Recall
1	15/100	27.67%	11.67%	27.67%
2	10/50	29.67%	15.00%	29.67%
3	10/10	25.67%	07.00%	25.67%

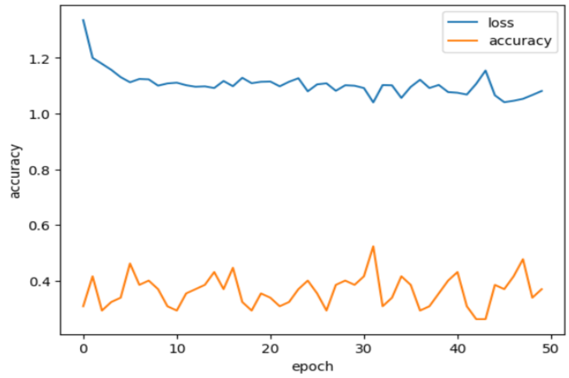


Fig. 6. Resnet152 best result graph (Epoch 50)

From Table 9, we can see that at epoch 100, accuracy values show no development and tend to be low. The best results were obtained in batch size 15 with an accuracy value of 27% followed by precision and recall reaching 11%, and 27%. Furthermore, with epoch 50 shows that the accuracy results only reached 29% and the maximum results in batch size 10. These results are best from epoch model scenarios 100 and 10, as shown in figure 6. It can be observed that the Loss Chart and accuracy values are still stagnant and have not shown that a cut will occur. For precision results, and recall score show values of 15%,29%. These results show similarities with resnet architecture experiments at epoch 100, broadly speaking, have not shown signs of changing values towards the curve cutting process. At the end, shows accuracy results reaching 25% with epoch 10 and batch size 10. These results show that the Resnet 152 algorithm is less powerful and maximal. Table 10 displays the best VGG16 result, while Figure 7 depicts the best VGG16 result graph.

Table 10. Resnet152 best result10 VGG16 Best result

No	Batch size/Epoch	Accuracy	Precission	Recall
1	10/100	29.33%	10.67%	29.33%
2	10/50	31.67%	13.33%	31.67%
3	15/10	25.67%	13.67%	25.67%

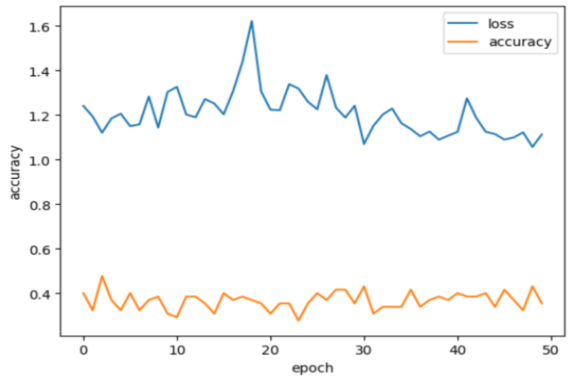


Fig. 7. VGG16 best result graph (Epoch 50)

In Table 10, epoch 100 shows accuracy results only reaching 29% and maximum results in batch size 10. For precision results, recall and f1-score show values of 10%, 29%, 15%. Further result shows accuracy results reaching 31% with epoch 50 and batch size 10. The graph according to Figure 7, shows the Loss line curve as well as accuracy values that have not undergone confluence or cutout. In the epoch range of 10-50, in general, the chart has stagnated and has not shown signs of cutting until the 50th epoch. The last result shows that at epoch 10, the accuracy value shows no development and tends to be low. The best results in epoch 10 at batch size 15 with an accuracy value of 25% followed by precision, and recall reached 13%, and 25%. The application of the test scenario shows the performance of the DL VGG16 model as the number of epochs increases, the accuracy value shows little development but the tendency is still low. The accuracy results in Batch size 10 and Epoch 50 are the highest accuracy values for the VGG16 Model, of the 3 scenarios used. The comparison of all deep learning CNN model presents in Figure 8.

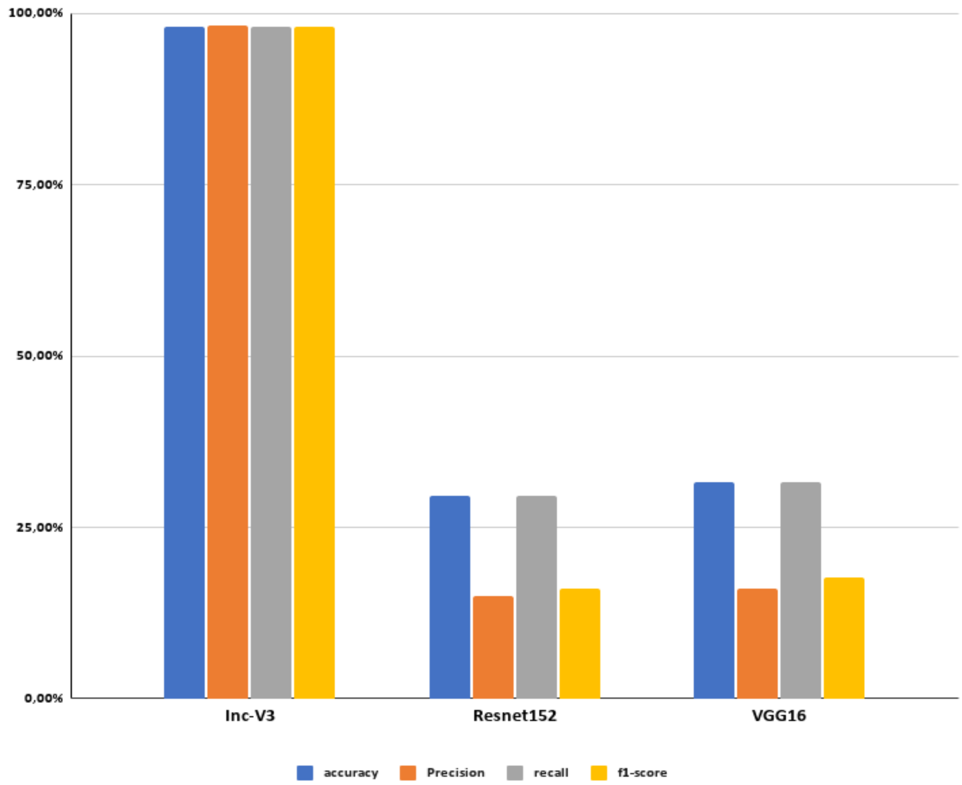


Fig. 8. VGG16 best result graph (Epoch 50)

From Figure 8, we can see that these results are important findings in the application of deep learning models in the classification of physiotherapy movement videos. In previous studies using machine learning, the average accuracy obtained was 93%. CNN Deep Learning models with 3 architectures (Inception V3, VGG16 and resnet152) get the best results on Inception V3 architecture, with 98% accuracy. This makes the Inception V3 model the best of the three models applied to this study, and 5% higher than previous studies with machine learning.

Things that affect the accuracy value and other values include the quantity of the number of datasets. With minimal datasets and k-fold test scenarios that were carried out showing good results, it is also necessary to test with more datasets and more varied K-fold scenarios. It is important to test deep learning models in this research in the future. With the production of deep learning classification models in the context of physiotherapy videos in the future, it is very necessary for the application of models to physiotherapy exercise actuators, one of which is a robotic arm. The findings will be expected to drive robotic arm actuator innovations in physiotherapy exercises for patients with upper extremity disorders or even cerebral palsy patients in the future.

IV. Conclusion

The conclusions obtained include the results of the Deep Learning Model for classification tasks on the physiotherapy video dataset for the therapy of upper limb disorders showed satisfactory results on the CNN Inception V3 architecture. Using the Batch size and Epoch parameters shown, the best results were obtained at batch size 5 and Epoch 100 on the Inception-V3 architecture with 98% accuracy. This shows that the application of models with minimal dataset quantity with parameters performed by DL-CNN is better in the classification of therapeutic movement videos in this study.

The limitation of this study is that the study used a dataset of motion detection video results of children aged 11-12 years with 4 males and 2 females and did not discuss specifically the method of taking videos because of the development of previous research and also this study only discussed the classification of 4 therapeutic movements applied. Noteworthy for future research suggestions is the need to increase the quantity of more and complete datasets to test the effectiveness of the model for future research. In addition, it is also recommended to test the model with application to the associated demonstration actuator to test the accuracy of the motion classification results in the deep learning model.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

References

[1] A. Gonzalez, L. Garcia, J. Kilby, and P. McNair, "Robotic devices for paediatric rehabilitation: a review of design features," Biomed. Eng. Online, vol. 20, no. 1, p. 89, Sep. 2021.

[2] D. O'Neill and D. E. Forman, "The importance of physical function as a clinical outcome: Assessment and enhancement," Clin. Cardiol., vol. 43, no. 2, pp. 108–117, Feb. 2020.

[3] R. Bouça-Machado et al., "Physical Activity, Exercise, and Physiotherapy in Parkinson's Disease: Defining the Concepts," Mov. Disord. Clin. Pract., vol. 7, no. 1, pp. 7–15, Jan. 2020.

[4] V. Riccio, G. Jahangirova, A. Stocco, N. Humbatova, M. Weiss, and P. Tonella, "Testing machine learning based systems: a systematic mapping," Empir. Softw. Eng., vol. 25, no. 6, pp. 5193–5254, Nov. 2020.

[5] C. Rossa, M. Najafi, M. Tavakoli, and K. Adams, "Robotic Rehabilitation and Assistance for Individuals With Movement Disorders Based on a Kinematic Model of the Upper Limb," IEEE Trans. Med. Robot. Bionics, vol. 3, no. 1, pp. 190–203, Feb. 2021.

[6] C. Marquez-Chin and M. R. Popovic, "Functional electrical stimulation therapy for restoration of motor function after spinal cord injury and stroke: a review," Biomed. Eng. Online, vol. 19, no. 1, p. 34, Dec. 2020.

[7] A. A. Barati, R. Rajabi, S. Shahrbanian, and M. Sedighi, "Investigation of the effect of sensorimotor exercises on proprioceptive perceptions among children with spastic hemiplegic cerebral palsy," J. Hand Ther., vol. 33, no. 3, pp. 411–417, Jul. 2020.

[8] W. Trusaji et al., "Horse Riding Simulator Design to Replicate Human Walking Gait for Hippotherapy in Cerebral Palsy Rehabilitation," Machines, vol. 10, no. 11, p. 1060, Nov. 2022.

[9] R. Llamas-Ramos, J. L. Sánchez-González, and I. Llamas-Ramos, "Robotic Systems for the Physiotherapy Treatment of Children with Cerebral Palsy: A Systematic Review," Int J Env. Res Public Health, 2022.

[10] M. J. Vinolo-Gil, E. Casado-Fernández, V. Perez-Cabezas, G. Gonzalez-Medina, F. J. Martín-Vega, and R. Martín-Valero, "Effects of the Combination of Music Therapy and Physiotherapy in the Improvement of Motor Function in Cerebral Palsy: A Challenge for Research," Children, vol. 8, no. 10, p. 868, Sep. 2021.

[11] A. S. Faradyza et al., "Real Time Gesture Detection Using Kinect in Rehabilitation Therapy for Children with Disability," in 2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Oct. 2021, pp. 452–456.
[12] K. Muhammad et al., "Human action recognition using attention based LSTM network with dilated CNN features," Future Gener. Comput. Syst., vol. 125, pp. 820–830, Dec. 2021.
[13] T. S. Gunawan et al., "Development of video-based emotion recognition using deep learning with Google Colab," TELKOMNIKA Telecommun. Comput. Electron. Control, vol. 18, no. 5, p. 2463, Oct. 2020.
[14] N. Jaouedi, N. Boujnah, and M. S. Bouhlel, "A new hybrid deep learning model for human action recognition," J. King Saud Univ. - Comput. Inf. Sci., vol. 32, no. 4, pp. 447–453, May 2020.
[15] G. Divya and S. Chokkalingam, "Human Action Recognition In Streaming Video Using Deep Learning," Int. J. Adv. Sci. Technol., vol. 29, no. 9, 2020.
[16] D. Chen, S. Shi, X. Gu, and B. Shim, "Robust DoA Estimation Using Denoising Autoencoder and Deep Neural Networks," IEEE Access, vol. 10, pp. 52551–52564, 2022.
[17] S. D. Thepade, M. Dindorkar, P. Chaudhari, and S. Bang, "Face presentation attack identification optimization with adjusting convolution blocks in VGG networks," Intell. Syst. Appl., vol. 16, p. 200107, Nov. 2022.
[18] Y. Abdulazeem, H. M. Balaha, W. M. Bahgat, and M. Badawy, "Human Action Recognition Based on Transfer Learning Approach," IEEE Access, vol. 9, pp. 82058–82069, 2021.
[19] P. Kuppusamy and V. C. Bharathi, "Human abnormal behavior detection using CNNs in crowded and uncrowded surveillance – A survey," Meas. Sens., vol. 24, p. 100510, Dec. 2022.
[20] S. Gupta, "Deep learning based human activity recognition (HAR) using wearable sensor data," Int. J. Inf. Manag. Data Insights, vol. 1, no. 2, p. 100046, Nov. 2021.
[21] V. Ghate and S. H. C, "Hybrid deep learning approaches for smartphone sensor-based human activity recognition," Multimed. Tools Appl., vol. 80, no. 28–29, pp. 35585–35604, Nov. 2021.
[22] V. B. Semwal, A. Gupta, and P. Lalwani, "An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition," J. Supercomput., vol. 77, no. 11, pp. 12256–12279, Nov. 2021.
[23] K. D. McCay, E. S. L. Ho, H. P. H. Shum, G. Fehringer, C. Marcroft, and N. D. Embleton, "Abnormal Infant Movements Classification With Deep Learning on Pose-Based Features," IEEE Access, vol. 8, pp. 51582–51592, 2020.
[24] S. Hechmi, "An Accurate Real-Time Method for Face Mask Detection using CNN and SVM," Knowl. Eng. Data Sci., vol. 5, no. 2, p. 129, Dec. 2022.
[25] Y. Adiwinata, A. Sasaoka, I. P. Agung Bayupati, and O. Sudana, "Fish Species Recognition with Faster R-CNN Inception-v2 using QUT FISH Dataset," Lontar Komput. J. Ilm. Teknol. Inf., vol. 11, no. 3, p. 144, Dec. 2020.
[26] G. Chen, D. Ye, Z. Xing, J. Chen, and E. Cambria, "Ensemble application of convolutional and recurrent neural networks for multi-label text categorization," in 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA: IEEE, May 2017, pp. 2377–2383.
[27] I. P. Kanani and M. Padole, "Deep Learning to Detect Skin Cancer using Google Colab," Int. J. Eng. Adv. Technol., vol. 8, no. 6, pp. 2176–2183, Aug. 2019.
[28] K. Joshi, V. Tripathi, C. Bose, and C. Bhardwaj, "Robust Sports Image Classification Using InceptionV3 and Neural Networks," Procedia Comput. Sci., vol. 167, pp. 2374–2381, 2020.
[29] Y. Li and L. Wang, "Human Activity Recognition Based on Residual Network and BiLSTM," Sensors, vol. 22, no. 2, p. 635, Jan. 2022.
[30] S. Raziani and M. Azimbagirad, "Deep CNN hyperparameter optimization algorithms for sensor-based human activity recognition," Neurosci. Inform., vol. 2, no. 3, p. 100078, Sep. 2022.
[31] H. Yu, X. Miao, and H. Wang, "Bearing Fault Reconstruction Diagnosis Method Based on ResNet-152 with Multi-Scale Stacked Receptive Field," Sensors, vol. 22, no. 5, p. 1705, Feb. 2022.
[32] T. B. Abdallah, I. Elleuch, and R. Guermazi, "Student Behavior Recognition in Classroom using Deep Transfer Learning with VGG-16," Procedia Comput. Sci., vol. 192, pp. 951–960, 2021.