Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3433~3447
e-ISSN: 2548-6861

Human Body Posture Classification from Digital Images Using the InceptionV3 Architecture

Munandar Rahmat Prayogi [1]*, Christy Atika Sari [2], Eko Hari Rachmawanto [3]***
Department of Informatics Engineering, Universitas Dian Nuswantoro, Indonesia
111202315125@mhs.dinus.ac.id [1], christy.atika.sari@dsn.dinus.ac.id [2], eko.hari@dsn.dinus.ac.id [3]

Article Info

Article history:

Received 2026-07-04
Revised 2026-07-24
Accepted 2026-08-08

Keyword:

Human Body Posture,
Image Classification,
InceptionV3,
Transfer Learning,
MPII Human Pose Dataset.

ABSTRACT

This study aims to develop a human body posture classification model based on digital images using the InceptionV3 architecture. The dataset used in this study is the MPII Human Pose Dataset, which contains a wide variety of human activities and body postures. The research began with label extraction from a metadata file in .mat format, followed by the selection of the 20 activity classes with the largest number of samples. Subsequently, the dataset was balanced using an undersampling technique, resulting in 140 images for each class. The images were then resized to 224×224 pixels, normalized using the preprocess_input function, and enhanced through data augmentation applied to the training set. The model was developed using a transfer learning approach with InceptionV3 as the base model. Additional layers, including Global Average Pooling, Dropout, and Dense layers, were added to perform multi-class classification. The experimental results showed that the proposed model achieved a test accuracy of 0.8893 with a test loss of 0.5587. Furthermore, the macro-average metrics obtained from the classification report were 0.9018 for precision, 0.8893 for recall, and 0.8877 for F1-score. These results indicate that the model was able to classify most activity classes effectively. However, several classes with similar visual characteristics still caused misclassification, indicating opportunities for further improvement in human posture recognition performance.

I. INTRODUCTION

Human Activity Recognition (HAR) is an important field of artificial intelligence that enables systems to automatically recognize, understand, and classify human movements and activities. This technology has been widely applied in various real-world domains, particularly in healthcare, elderly monitoring, workplace ergonomics, medical rehabilitation, sports analysis, and intelligent surveillance systems. Previous studies have demonstrated that HAR also plays an important role in physical activity monitoring, behavioral analysis, and the development of intelligent systems for real-world environments [1], [2].

In recent years, image-based approaches have become increasingly popular for human body posture classification because they are more practical than sensor-based methods that require users to wear dedicated devices. Image-based systems do not require additional hardware to be attached to the human body, making them more convenient, easier to deploy, and more suitable for real-world applications. This approach is particularly relevant for visual monitoring in homes, workplaces, and healthcare facilities because it utilizes cameras or digital images as the primary source of information. Furthermore, image-based methods facilitate the development of more flexible systems that are independent of wearable devices [1], [3], [4].

Several previous studies have demonstrated that image-based human body posture classification can be effectively performed using deep learning techniques. Ogundokun *et al.* applied transfer learning and hyperparameter optimization to the MPII Human Pose Dataset and achieved promising classification performance using the AlexNet and VGG16 architectures. Their findings indicate that transfer learning models can effectively handle human posture data, particularly when the amount of training data is relatively limited. The study also emphasized that image augmentation and appropriate hyperparameter optimization play important roles in improving model performance while reducing the risk of overfitting [5].

Another study demonstrated that Convolutional Neural Network (CNN)-based models are capable of achieving high performance in human body posture classification tasks. Rababaah employed indoor RGB images to classify ten different body posture variations and obtained high classification accuracy. These results indicate that CNN models can automatically extract discriminative visual features from images without requiring manual feature engineering. Furthermore, the study reported that the number of extracted features and convolutional kernels significantly influences the final classification performance of the model [6], [7].

Another study further demonstrated that CNN models can achieve satisfactory performance even on relatively small datasets when the network architecture and training parameters are carefully designed. Research on CNN Architecture for Posture Classification on Small Data showed that a relatively simple CNN architecture was able to achieve good performance in patient posture classification using bed pressure data. This finding indicates that model performance is determined not only by dataset size but also by the selection of an appropriate network architecture, effective training strategies, and proper parameter optimization [8].

In addition, several studies have employed pose estimation and keypoint extraction techniques to recognize human body postures and movements. The study Detection of Sitting Posture Using Hierarchical Image Composition and Deep Learning demonstrated that posture-related information can be extracted from visual representations of the human body and effectively utilized for sitting posture classification. Meanwhile, A Hybrid Neural Network-Based Intelligent Body Posture Estimation System in Sports Scenes showed that the combination of CNN and Long Short-Term Memory (LSTM) architectures effectively captures both spatial and temporal information from human body movements. These findings demonstrate that body-related visual information can be effectively utilized for posture recognition using either image or video data [3], [9]. In the field of yoga and sports activities, several studies have also compared various modern CNN architectures. Aydın evaluated several CNN-based approaches for yoga pose classification and demonstrated that deep learning methods outperform simpler models, particularly when the dataset is enriched through image transformation and data augmentation techniques. Similarly, Ashraf et al. developed YoNet for yoga pose classification and compared its performance with several existing CNN architectures. Their proposed model achieved excellent classification performance and further confirmed that selecting an appropriate CNN architecture plays a crucial role in determining the effectiveness of image-based posture classification systems[10], [11].

Another study in posture analysis and healthcare also demonstrated that body-image data can be effectively utilized for human body position detection with high accuracy. Ogundokun *et al.* proposed a hybrid InceptionV3-SVM framework for human posture detection and achieved promising performance. Their study demonstrated that InceptionV3 is capable of extracting important visual features from human body images and can produce strong classification performance when combined with an additional classification algorithm. In another study, InceptionV3 was also adopted as one of the benchmark models for yoga pose classification, further demonstrating its relevance and effectiveness for image-based human posture recognition tasks [10], [12].

Based on previous studies, image-based human body posture classification has shown promising performance across various application domains. Nevertheless, several research gaps remain. First, most previous studies focused only on specific activities or postures, such as yoga poses, sitting postures, sleeping positions, or particular sports activities, leaving the capability of models to recognize a broader range of human activities largely unexplored. Second, many existing studies rely on pose estimation, keypoint-based, skeleton-based, or sensor-based approaches, meaning that original digital images are not directly utilized as the primary input for classification [2], [13], [14]. These approaches generally require additional feature extraction processes or specific body representations, making their implementation more complex. Third, although the MPII Human Pose Dataset has been widely used for human pose estimation research, the activity labels provided within the dataset have received relatively limited attention for image-based human activity classification. Although the dataset was originally developed for human pose estimation tasks, each image is also annotated with a corresponding human activity label, making it suitable for image-based human activity classification. Therefore, a more generalized image-based classification approach capable of recognizing a wider variety of human activities and applicable to diverse real-world scenarios is still needed.

Based on these research gaps, this study proposes an image-based human body posture classification framework using the InceptionV3 architecture. The activity labels available in the MPII Human Pose Dataset were utilized as classification targets, and the 20 activity classes with the largest number of samples were selected to ensure sufficient training data for each class. The selection of these 20 activity classes was motivated by the fact that they contain an adequate number of samples for CNN training, whereas most of the remaining classes contain only a limited number of images, potentially resulting in severe class imbalance and unstable training.

To address class imbalance, an undersampling strategy was applied so that each activity class contained the same number of samples. This strategy was adopted to reduce the dominance of majority classes during model training and to provide each class with a more balanced learning opportunity. Furthermore, image preprocessing, data augmentation, and transfer learning were integrated into a unified framework to improve the model's capability to recognize variations in human body postures. Unlike previous studies, which generally focused on pose estimation or specific activity domains, this study integrates activity label extraction from the MPII Human Pose Dataset, dataset balancing through undersampling, image preprocessing, data augmentation, and InceptionV3-based transfer learning into a single classification framework for evaluating 20 human activity classes using digital images. This approach provides an alternative utilization of the MPII Human Pose Dataset beyond conventional human pose estimation tasks by enabling image-based human activity classification. The selection of the InceptionV3 architecture was motivated by its ability to extract hierarchical visual features, perform effectively within a transfer learning framework, and achieve strong performance in previous image classification studies [5], [12], [15], [16].

This study is expected to contribute to the development of practical and reliable computer vision-based human body posture classification systems by proposing an image-based classification framework that integrates activity label extraction, dataset balancing, image preprocessing, data augmentation, and transfer learning using the InceptionV3 architecture. Furthermore, the findings of this study are expected to provide a foundation for developing applications in healthcare monitoring, medical rehabilitation, sports analysis, intelligent surveillance, and automated human activity monitoring. Nevertheless, the findings remain limited to evaluations conducted on the MPII Human Pose Dataset, and further research is required to validate the generalization capability of the proposed model using additional datasets and more diverse real-world environments [9], [17].

II. METHOD

A. Research Stages

This study was systematically designed to develop a human body posture classification model based on digital images using a Convolutional Neural Network (CNN) with the InceptionV3 architecture. The research process consisted of several stages, including dataset collection, data preprocessing, model development, model training, and performance evaluation. This sequence of stages is consistent with the deep learning-based approaches commonly employed in previous studies on posture classification and pose recognition, particularly in the fields of human pose estimation and human activity recognition [1], [14] as in Figure 1.

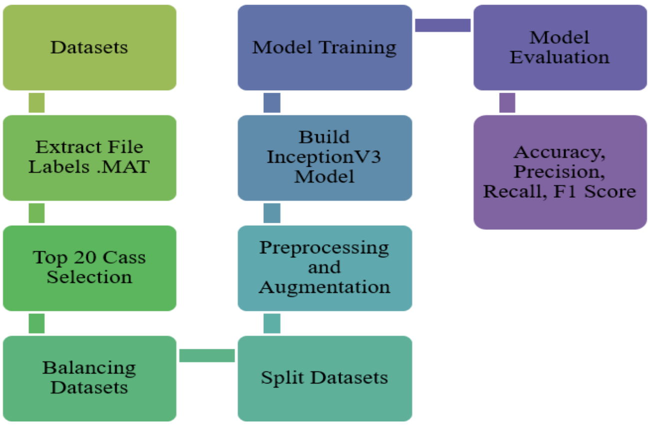

Figure 1. Research Stage

The study began with the collection of data from the MPII Human Pose Dataset, which consists of a diverse range of images representing human activities and body postures https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/software-and-datasets/mpii-human-pose-dataset/download. The dataset was selected because it contains a diverse range of human activities and has been widely used in image-based human posture research. At this stage, activity labels were extracted from a metadata file in .mat format using the Python programming language and the SciPy library. The extracted labels were then used to define the activity categories employed in the classification process. The use of the MPII Human Pose Dataset is also consistent with previous posture classification studies that have utilized similar datasets [5], [12] for transfer learning and posture image analysis.

After the data collection process was completed, the study proceeded with dataset selection and data balancing. In this study, the 20 activity classes with the largest number of samples were selected. Following the selection process, each class was balanced using an undersampling technique, resulting in an equal number of samples for every class, with 140 images per class. Data balancing was performed to reduce class imbalance, which could otherwise introduce bias into the classification model [5], [8], [16]. This step is important because image classification studies involving limited data and imbalanced class distributions often require data balancing strategies to ensure model stability and reliable learning during the training process [4].

The balanced dataset was then divided into three subsets: 70% for training, 20% for validation, and 10% for testing. Subsequently, an image preprocessing stage was performed by resizing all images to 224×224 pixels and applying normalization using the preprocess_input function provided by TensorFlow Keras. In addition, data augmentation techniques, including rotation, zoom, and horizontal flipping, were applied through the ImageDataGenerator framework to increase the diversity of the training data and reduce the risk of overfitting [3], [4], [5], [15], [16]. This strategy is consistent with previous studies demonstrating that image augmentation can expand data variability and improve the generalization capability of deep learning models in human posture classification tasks [4].

The next stage involved the development of a classification model based on the InceptionV3 architecture using a transfer learning approach. Several additional layers, including Global Average Pooling, Dropout, and Dense layers, were added to the final part of the network to adapt the model to the number of activity classes used in this study. InceptionV3 was selected due to its ability to extract hierarchical visual features and effectively handle variations in image characteristics. This approach is consistent with previous studies demonstrating that InceptionV3 is effective for human posture classification tasks, particularly when combined with transfer learning techniques and additional regularization mechanisms to reduce the risk of overfitting [5], [12], [16].

After the model architecture had been established, the training process was conducted using the Adam optimizer with a learning rate of 0.0001 for 100 epochs. The training stage aimed to enable the model to learn the visual patterns associated with each human activity class from the training data that had undergone preprocessing and data augmentation. The use of Adam is also consistent with common practices in deep learning model training, as it provides stable and efficient convergence during the optimization process [5], [8]. This training stage is essential to ensure that the model can effectively distinguish the visual characteristics of different activity classes and achieve optimal classification performance.

The final stage involved evaluating the performance of the proposed model using several evaluation metrics, including accuracy, precision, recall, and F1-score. In addition, a confusion matrix was employed to analyze the model's ability to classify each human activity class correctly. These evaluation metrics provide a more comprehensive understanding of the model's performance on unseen data, particularly when certain classes exhibit highly similar visual characteristics. This evaluation approach is widely adopted in human posture classification and human activity recognition studies because it enables a thorough assessment of model performance, both at the overall level and for individual classes [9], [16], [17].

Based on the research stages described above, the InceptionV3-based CNN model is expected to achieve strong performance in image-based human body posture classification. The combination of class selection, undersampling, preprocessing, data augmentation, transfer learning, and appropriate evaluation techniques is expected to enhance the model's ability to recognize human posture patterns in a more stable, reliable, and measurable manner [4], [5], [14], [16].

B. *Dataset*

The dataset used in this study is the MPII Human Pose Dataset, which contains a wide variety of images representing human activities and body postures. This dataset was selected because it encompasses diverse activity categories, making it suitable for image-based human body posture classification research. The use of image-based datasets is also consistent with previous studies in the field of human pose estimation, which emphasize that images and video data serve as important sources of information for human posture recognition, particularly in deep learning-based approaches [3], [4], [6], [14], [16]. Although the MPII Human Pose Dataset was originally developed for human pose estimation tasks, each image is also accompanied by an activity annotation. The availability of these activity labels enables the dataset to be utilized for image-based human activity classification by treating the activity labels as classification targets. In this study, the activity labels were extracted from the metadata file in *.mat* format using the Python programming language and the SciPy library. From the original 24,984 images, the label extraction process produced 18,030 valid samples containing 397 unique human activity labels.

TABLE I
SUMMARY OF DATASET CHARACTERISTICS

Dataset Processing Stage	**Number of Samples**
Original MPII Human Pose Dataset	24.984
Valid samples after label extraction	18.030
Number of unique activity labels	397
Number of selected activity classes	20
Number of images after undersampling	2.800

Table I summarizes the changes in the number of samples throughout the dataset preparation process, beginning with activity label extraction and ending with the construction of the final dataset used in this study. Owing to the imbalance in the number of samples across activity classes, only the 20 activity classes containing the largest number of samples were selected to maintain the stability of the model training process. These classes were selected because they contained a sufficient number of images for CNN training, whereas most of the remaining activity classes contained only a limited number of samples, which could potentially lead to severe class imbalance and reduce both training stability and model learning performance[4].

The selected activity classes include aerobic step, ballet, workout training, softball, soccer, skiing, skateboarding, rowing, rope skipping, rock climbing, resistance training, golf, forestry, cooking, circuit training, carpentry, bicycling, ballroom, and yoga, as illustrated in Figure 2.

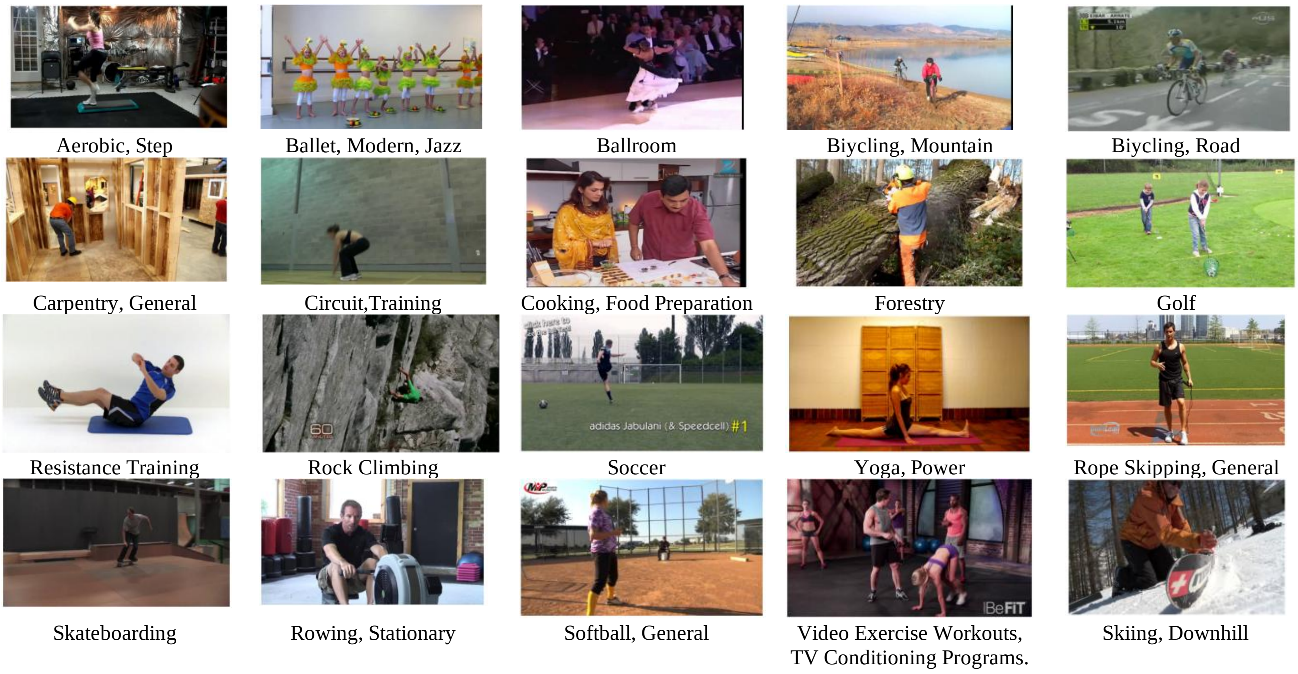

Figure 2. Sample dataset for each class

The class selection strategy is important because posture classification studies involving limited or imbalanced datasets generally require careful class selection and appropriate data distribution management to prevent the model from becoming biased toward particular classes [5], [8], [16], [18]. After the class selection stage, each activity class was balanced using an undersampling technique so that every class contained an equal number of samples, namely 140 images per class. This balancing process was performed to reduce class imbalance, which could otherwise cause the model to become biased toward majority classes while exhibiting lower performance on minority classes.

Although the undersampling technique reduced the total number of samples available for training, this approach was selected because it produces a balanced class distribution without introducing synthetic data. Consequently, each activity class was provided with an equal opportunity to contribute during the training process, thereby reducing the bias toward majority classes. This strategy is consistent with findings reported in previous posture classification studies [4], [5], [8], [18], which demonstrate that appropriate data distribution and data augmentation play important roles in maintaining training stability and improving classification performance.

Alternative approaches, such as oversampling, class weighting, and focal loss, can also be employed to address class imbalance. However, this study adopted undersampling because it provides a simple and balanced class distribution, allowing the performance of the InceptionV3 model to be evaluated more objectively. A comprehensive comparison between undersampling and other data balancing strategies remains an important direction for future research. Therefore, undersampling was considered the most appropriate approach for this study, as its primary objective was to construct a balanced dataset and objectively evaluate the baseline classification performance of the InceptionV3 architecture.

The balanced dataset was subsequently divided into three subsets: 70% for training, 20% for validation, and 10% for testing, as presented in Table II. The data splitting process was performed after the undersampling stage had been completed, ensuring that the training, validation, and testing datasets were all derived from a balanced dataset with an identical class distribution. This strategy was adopted to prevent data leakage during model development and evaluation. The training dataset was used to optimize the model parameters, the validation dataset was employed to monitor model performance throughout the training process, and the testing dataset was reserved exclusively for evaluating the final performance of the trained model on previously unseen data [16].

Following the data splitting stage, image preprocessing was performed by resizing all images to 224 × 224 pixels and normalizing the pixel values using the preprocess_input function provided by TensorFlow Keras. This standardized preprocessing procedure is widely adopted in transfer learning and image-based posture classification studies because it ensures that the input images conform to the format required by the pre-trained model [4], [5], [6], [12], [16].

In addition to normalization, this study employed several data augmentation techniques, including rotation, zoom, and horizontal flipping, using the ImageDataGenerator framework to increase the diversity of the training images and reduce the risk of overfitting. Data augmentation is particularly important for posture classification tasks because human body postures may appear under different viewpoints, scales, and orientations. Previous studies have also demonstrated that image augmentation improves the robustness of deep learning models against data variability while enhancing their generalization capability when encountering previously unseen samples [3], [4], [5], [15]. The undersampling process produced a final dataset consisting of 2,800 images, which served as the basis for training and evaluating the InceptionV3 model for the classification of 20 human activity classes. By employing a balanced data distribution, the proposed model was expected to learn all activity classes more uniformly while reducing the tendency to become biased toward classes containing a larger number of samples[16].

TABLE II
DATASET DISTRIBUTION

Dataset Type	Number of Samples
Training Data	1,960
Validation Data	560
Testing Data	280
Total Samples	2,800

C. Data Preprocessing

The data preprocessing stage was conducted to prepare the images before they were used for training the Convolutional Neural Network (CNN) model. This process aimed to improve the quality of the input data, adapt the image format to the requirements of the model architecture, and reduce the risk of overfitting during training [4]. In image-based human posture and activity classification studies, data preprocessing is considered a crucial component because the quality of the input data significantly influences the model's ability to extract meaningful visual features [16], [19], [20]. The first step in the preprocessing stage was image resizing. All images in the dataset were resized to 224 × 224 pixels to match the input dimensions required by the InceptionV3 architecture. This process was performed to ensure that all images had a uniform size, allowing the CNN model to process the data consistently. The use of standardized input dimensions is also widely adopted in human posture classification and CNN-based human activity recognition studies, as it facilitates more effective and efficient feature extraction [4], [16], [21], [22]. The next step involved image normalization using the preprocess_input function provided by InceptionV3. The purpose of normalization was to adjust the distribution of pixel values to match the input format expected by the pre-trained model, thereby enabling more effective visual feature extraction during the training process [4], [5], [15]. In addition, normalization helps accelerate model convergence and maintain learning stability within deep neural networks [16].

In addition to normalization, this study applied data augmentation techniques to the training dataset. The augmentation process was performed using the ImageDataGenerator framework with several image transformations, including rotation, zoom, and horizontal flipping. The purpose of data augmentation was to increase the diversity of the training data, enabling the model to generalize more effectively to unseen samples [4], [5], [10], [11]. The use of data augmentation is particularly important because variations in human body poses, camera viewpoints, and environmental conditions can lead to significant differences in the visual characteristics of the images. The application of data augmentation has also been proven effective in reducing the risk of overfitting in various human posture classification studies. Research in the field of yoga pose classification has demonstrated that image transformations can help models recognize the same pose despite changes in body orientation or positioning[4]. Similar findings have been reported in studies on sports activity classification and human posture detection, where data augmentation improved the model's ability to handle variations in data that were not encountered during the training process [19], [21], [22].

In this study, the validation and testing datasets underwent only the normalization process without any additional data augmentation. This approach was adopted to ensure that model evaluation was conducted using original images that had not been subjected to artificial transformations. Consequently, the evaluation results provide a more objective assessment of the model's ability to generalize to previously unseen data. This methodology is also widely employed in human posture classification studies, as it enables a more realistic measurement of model performance under conditions that closely resemble real-world scenarios [23], [24]. In addition to improving data quality, the preprocessing stage plays an important role in maintaining the consistency of the visual representations received by the model [25], [26], [27]. In studies related to human pose estimation and human activity recognition, high-quality input images facilitate the model's ability to identify movement patterns, body shapes, and the relationships among different parts of the human body. Therefore, preprocessing is considered a critical step before images are used for training CNN-based models.

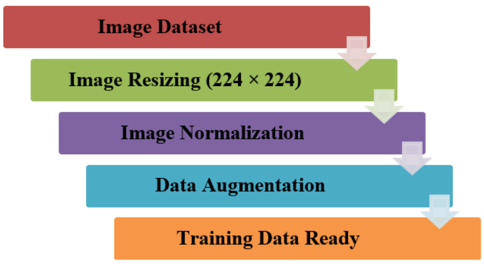

Figure 3. Data Preprocessing Workflow

The output of the preprocessing stage as illustrated in Figure 3 consisted of a collection of images that had undergone resizing, normalization, and data augmentation, making them ready for use in the InceptionV3 training process. The combination of these preprocessing techniques is expected to enhance training stability, improve the model's generalization capability, and support more effective image-based human body posture classification.

D. *InceptionV3 Model Architecture*

In this study, human body posture classification was performed using the InceptionV3 deep learning architecture based on Convolutional Neural Networks (CNNs). The InceptionV3 architecture [4], [5], [12], [16] was selected because of its ability to effectively extract visual features and its proven relevance in image-based human posture classification research. Several studies have demonstrated that InceptionV3 achieves strong performance in human posture detection tasks, particularly when combined with transfer learning techniques and model regularization strategies.

The InceptionV3 model used in this study was a pre-trained model originally trained on the ImageNet dataset and implemented using a transfer learning approach. The purpose of transfer learning was to leverage the weights learned from previous training so that the model could perform more effective visual feature extraction [4], [5], [15], [16], even when the available research dataset was relatively limited in size. This approach is consistent with previous studies demonstrating that transfer learning can improve training efficiency and enhance the generalization capability of models when applied to posture image classification tasks[16]. The original top classification layer of InceptionV3 was removed (include_top=False) to adapt the model to the specific requirements of human activity classification. After utilizing the main feature extraction component of the network (base model), several additional layers were added, including Global Average Pooling, Dropout, and Dense layers, to improve classification performance. This architectural design is consistent with previous posture classification studies that modify pre-trained models to better accommodate the target number of classes and the limitations of relatively small datasets [4], [5], [8].

The Global Average Pooling layer was employed to reduce the dimensionality of the extracted feature maps without discarding important information contained in the images[16]. Subsequently, a Dropout layer with a rate of 0.2 was applied to reduce the risk of overfitting during the model training process [4], [5], [15]. A Dense layer with the ReLU activation function was then added to facilitate the learning of more complex feature representations. In the output layer, the Softmax activation function was employed because this study involves a multi-class classification problem consisting of 20 human activity categories. The Softmax function transforms the output values (logits) into a probability distribution across all classes, ensuring that the sum of the predicted probabilities equals one. The mathematical formulation of the Softmax function is presented in Equation (1), where (P(y_i)) represents the probability of the (i)-th class, (z_i) denotes the output value (logit) of the (i)-th neuron, and (K) represents the total number of classes involved in the classification process. The class with the highest probability value is selected as the final prediction produced by the model.

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

The model architecture employed in this study accepts input images with a resolution of 224 × 224 pixels and three color channels (RGB). The entire model development process was implemented using the TensorFlow Keras framework as shown in Figure 4. The resulting model was subsequently utilized for training and evaluation to assess its performance in image-based human activity classification. The use of a CNN-based architecture for image-based posture classification [6], [7], [8] is also consistent with previous studies demonstrating that CNN models remain effective for human posture classification when the network architecture and training parameters are appropriately adapted to the characteristics of the dataset.

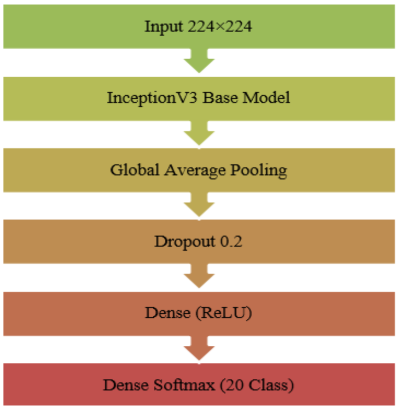

Figure 4. InceptionV3 Model Architecture

E. *Training Model*

The model training stage was conducted after the completion of data preprocessing and model architecture development. In this study, the training process employed a Convolutional Neural Network (CNN) with the InceptionV3 architecture to perform image-based human body posture classification. It was trained using the training dataset, which had previously undergone preprocessing and data augmentation. The training data enabled the model to learn the visual patterns and characteristics associated with each human activity class. In addition, a validation dataset was used to monitor the model's performance throughout the training process. The use of validation data is essential for assessing the model's ability to generalize to data that are not directly involved in the learning process [5], [16]. Model training was performed using the Adam optimizer with a learning rate of 0.0001. The Adam optimizer was selected because it provides a stable and efficient convergence process during deep neural network training. Furthermore, Adam has been widely adopted in deep learning-based human posture and activity classification studies due to its ability to accelerate model learning while requiring relatively minimal hyperparameter tuning [4], [5], [19].

Here, the categorical crossentropy loss function was employed because the task involves multi-class classification with 20 human activity categories. Meanwhile, the Softmax activation function was applied to the output layer to generate classification probabilities for each activity class. The combination of categorical crossentropy and Softmax is a widely adopted approach in multi-class classification problems, as it effectively optimizes the learning process based on the predicted probability distribution across all classes [6]. The categorical crossentropy loss function is used to measure the discrepancy between the true class labels and the predicted probabilities generated by the model. The mathematical formulation of the categorical crossentropy loss function is presented in Equation (2), where (L) represents the loss value, (y_i) denotes the true label of the (i)-th class, ({y}_i) represents the probability predicted by the model for the (i)-th class, and (C) denotes the total number of classes involved in the classification process. A lower categorical crossentropy value indicates that the predicted probabilities are closer to the true class labels, reflecting better model performance.

$$L = -\sum_{i=1}^{C} y_i \, log\ (\hat{y}_i) \tag{2}$$

Based on Table II, the training process was conducted for 100 epochs with a batch size of 16. Throughout the training process, the system recorded the accuracy and loss values at each epoch using the CSVLogger callback. In addition, the ModelCheckpoint callback was employed to save the best-performing model based on the highest validation accuracy achieved during training. Recording the training history and preserving the best model facilitate performance analysis and help prevent the use of models that exhibit inferior performance on the validation dataset.

The entire training process was carried out using the TensorFlow Keras library within the JupyterLab environment. After training was completed, the final model was saved in the .h5 format, allowing it to be reused for evaluation and future implementation purposes. The outcome of the training stage was an InceptionV3 model that had learned the visual characteristics of 20 human activity classes and was ready to be utilized for testing and performance evaluation.

TABLE II
MODEL TRAINING PARAMETERS

Parameter	Value
Model Architecture	InceptionV3
Number of Epochs	100
Batch Size	16
Learning Rate Optimizer	0.001 Adam
Loss Function	Categorical Crossentropy
Output Activation Function	Softmax
Dropout Rate	0.2

F. Model Evaluation

The evaluation stage was conducted to assess the capability of the Convolutional Neural Network (CNN) model based on the InceptionV3 architecture in performing image-based human body posture classification. The evaluation process utilized the testing dataset, which had not been used during the model training phase. The purpose of using a testing dataset was to objectively measure the model's generalization ability when applied to previously unseen data. This evaluation approach is commonly adopted in human posture classification studies because the testing dataset should accurately represent the model's performance under conditions that were not encountered during training [12], [16], [21].

Here, our model was evaluated using several evaluation metrics, namely accuracy, precision, recall, and F1-score. Accuracy was used to measure the overall correctness of the model in classifying the testing data. Precision was employed to assess the accuracy of the model's predictions for a particular activity class, while recall was used to evaluate the model's ability to identify all instances belonging to the actual class. Furthermore, the F1-score was utilized to measure the balance between precision and recall, providing a more comprehensive assessment of classification performance. The combination of these metrics is widely used in posture classification studies because it offers a more complete representation of model performance than relying solely on accuracy [4], [12], [23]. The mathematical formulations of the evaluation metrics used in this study are presented in Equations (3) through Equations (6). Although other evaluation metrics, such as Top-5 Accuracy and Balanced Accuracy, are also commonly employed in multi-class classification problems, this study focuses on Accuracy, Precision, Recall, and F1-score because these metrics provide a comprehensive evaluation of the overall classification performance as well as the performance of each activity class. In particular, Precision, Recall, and F1-score are considered appropriate for assessing class-wise performance in datasets with class imbalance or varying classification difficulty. Furthermore, these metrics are among the most frequently reported evaluation measures in image-based human posture classification studies, thereby facilitating comparison between the proposed model and previous research [4].

In Equations (3) through (6), TP (True Positive) represents the number of samples that are correctly classified into their respective classes, TN (True Negative) represents the number of samples that are correctly identified as not belonging to a particular class, FP (False Positive) represents the number of samples that are incorrectly classified as belonging to a specific class, and FN (False Negative) represents the number of samples that fail to be recognized as belonging to their actual class. Accuracy is used to measure the overall correctness of the classification process, precision evaluates the accuracy of predictions for a specific class, recall measures the model's ability to identify all samples belonging to the true class, and F1-score assesses the balance between precision and recall.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

In addition to the evaluation metrics, this study also employed a confusion matrix to analyze the classification results for each human activity class. The confusion matrix was used to determine the number of correct and incorrect predictions made by the model for each activity category. Through the confusion matrix, classification performance can be examined in greater detail, allowing misclassifications between classes to be identified more clearly. The use of a confusion matrix is particularly important in posture classification tasks because some classes exhibit highly similar visual characteristics, which can lead to prediction errors between activities with comparable postures or movements [12], [16], [21], [23]. The evaluation process was carried out using the Scikit-learn library in the Python programming language. The evaluation results were presented in the form of accuracy and loss curves, a classification report, and a confusion matrix. In addition to classification performance, computational efficiency can also be considered an important aspect in practical deployment. However, this study primarily focuses on evaluating classification performance, while computational metrics such as training time, inference time, and computational cost are identified as potential directions for future work. The accuracy and loss curves were used to monitor the stability of the model training process over 100 epochs. The classification report was employed to present the precision, recall, and F1-score values for each human activity class. This performance reporting approach has also been widely adopted in other posture recognition studies to evaluate classification results at the class level and to identify the classes that are the easiest and the most challenging for the model to recognize [4], [9], [12].

III. Result and Discussion

This section presents the implementation, experimental results, and discussion of the image-based human body posture classification model using the InceptionV3 architecture. The discussion begins with the results of model training on the balanced dataset, followed by the evaluation results obtained from the testing dataset, and an analysis of the model's performance in classifying 20 human activity classes. Furthermore, this section examines the effects of preprocessing, undersampling, data augmentation, and transfer learning on the overall performance of the proposed model.

A. Training Results

The model used in this study was developed using a transfer learning approach with the InceptionV3 architecture as the base model. The main InceptionV3 network was utilized with the parameter include_top=False, followed by the addition of a Global Average Pooling layer, Batch Normalization layer, Dropout layer, a Dense layer with 256 neurons, another Batch Normalization layer, and a final Dense output layer containing 20 neurons to represent the 20 human activity classes. Based on the model summary generated in the notebook, the proposed architecture consisted of a total of 22,341,684 parameters, including 11,075,284 trainable parameters and 11,266,400 non-trainable parameters. This architecture demonstrates that the model not only leverages the feature extraction capabilities of InceptionV3 but also adapts them to the requirements of multi-class classification for the dataset used in this study.

The model training process was conducted using 1,960 training images, 560 validation images, and 280 testing images. Training was performed for 100 epochs with a batch size of 16, a learning rate of 0.0001, the Adam optimizer, and the categorical crossentropy loss function. Throughout the training process, the model recorded accuracy and loss values using the CSVLogger callback, while the best-performing model was saved using the ModelCheckpoint callback based on the highest validation accuracy achieved during training. This training configuration is well suited for CNN-based multi-class classification tasks and supports a more systematic and reliable evaluation process.

The training results indicate that the model was able to learn the visual patterns present in the training data effectively. The training accuracy curve increased rapidly during the early epochs and gradually approached a high value toward the end of the training process. The validation accuracy also improved over time; however, it remained lower than the training accuracy and tended to stabilize within a lower range. This pattern suggests that the model successfully learned meaningful representations from the training data, although a performance gap between the training and validation datasets was still observed. Regarding the loss curves, the training loss decreased sharply during the initial epochs and continued to approach a low value as training progressed. In contrast, the validation loss decreased more gradually and exhibited fluctuations during several epochs. These results indicate that the model achieved convergence during training, while still exhibiting a reasonable generalization gap, which is commonly observed in datasets balanced through undersampling techniques.

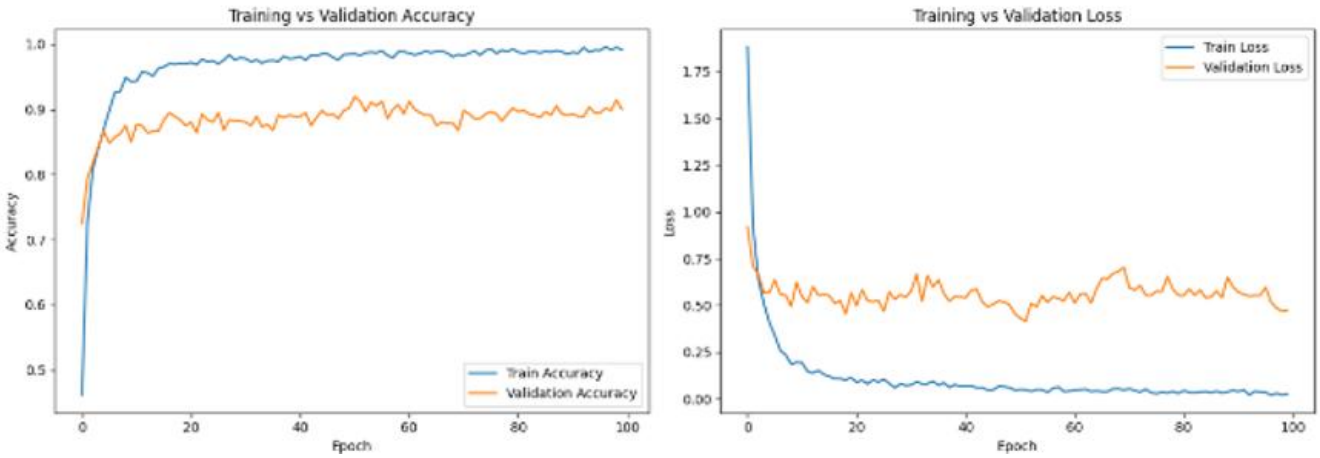

Figure 5. Training Accuracy and Loss Curves

At the end of the training process, the curves indicate that the model achieved better performance on the training dataset than on the validation dataset as shown in Figure 5. This condition suggests that the model was able to learn visual features from posture images effectively, although there is still room for improvement in terms of generalization to previously unseen data. The use of a dropout rate of 0.2 and data augmentation techniques on the training dataset helped reduce the risk of overfitting, preventing the model from relying excessively on highly specific patterns present in the training data. This training behavior is consistent with transfer learning-based posture recognition studies, which emphasize the importance of data augmentation and regularization techniques in maintaining training stability and improving model generalization.

B. Testing Implementation

The model was evaluated using the testing dataset, which was not utilized during the training process. This stage aimed to objectively assess the model's ability to classify human body postures from previously unseen images. The evaluation results demonstrate that the InceptionV3 model achieved satisfactory performance in recognizing the 20 human activity classes. Model performance was assessed using several evaluation metrics, including accuracy, precision, recall, and F1-score. In addition, a confusion matrix was employed to analyze the distribution of classification results and to examine the model's performance for each activity class in greater detail.

During the testing stage, the model was evaluated using the testing dataset, which had not been used during the training process. The evaluation results indicate that the InceptionV3 model was able to perform human body posture classification with satisfactory performance across the 20 activity classes. Based on the testing output, the model achieved a test accuracy of 0.8893 and a test loss of 0.5587. These results demonstrate that the model was capable of effectively recognizing visual patterns from the testing images, although several activity classes were still not classified perfectly.

The classification report also indicates that the model's performance varied across different activity classes. Several classes as illustrated in Figure 6 achieved high precision, recall, and F1-score values, including aerobic, step, golf, rope skipping, general, and soccer. These results suggest that the images belonging to these classes possess distinctive visual characteristics that are easier for the model to recognize. The high performance achieved by these activity classes is likely related to their distinctive body configurations, characteristic movement patterns, and relatively consistent visual context within the dataset. Such visual characteristics enable the model to extract more discriminative features, resulting in more accurate classification. In contrast, classes such as ballet, modern, or jazz, resistance training, and video exercise workouts, TV conditioning programs exhibited lower recall values. The lower recall observed in these classes may be attributed not only to similarities in body posture but also to variations in viewing angle, background complexity, body orientation, and overlapping visual characteristics among different activities. These factors make feature discrimination more challenging for the CNN model. This finding indicates that the model still encounters difficulties in distinguishing between classes that share similar visual features or involve more complex variations in human poses

35/35 [==============================] - 6s 139ms/step

	precision	recall	f1-score	support
aerobic, step	1.0000	1.0000	1.0000	14
ballet, modern, or jazz	1.0000	0.6429	0.7826	14
ballroom	0.8571	0.8571	0.8571	14
bicycling, mountain	0.8667	0.9286	0.8966	14
bicycling, racing and road	0.8667	0.9286	0.8966	14
carpentry, general	0.9286	0.9286	0.9286	14
circuit training	0.9286	0.9286	0.9286	14
cooking or food preparation	0.8667	0.9286	0.8966	14
forestry	1.0000	0.9286	0.9630	14
golf	0.9333	1.0000	0.9655	14
resistance training	1.0000	0.5714	0.7273	14
rock climbing	1.0000	0.7857	0.8800	14
rope skipping, general	0.9333	1.0000	0.9655	14
rowing, stationary	0.9231	0.8571	0.8889	14
skateboarding	0.8462	0.7857	0.8148	14
skiing, downhill	1.0000	1.0000	1.0000	14
soccer	0.8235	1.0000	0.9032	14
softball, general	0.8000	0.8571	0.8276	14
video exercise workouts, TV conditioning programs	0.6500	0.9286	0.7647	14
yoga, Power	0.8125	0.9286	0.8667	14
accuracy			0.8893	280
macro avg	0.9018	0.8893	0.8877	280
weighted avg	0.9018	0.8893	0.8877	280

Figure 6. Classification Report Results

The testing results are further illustrated through the confusion matrix. This matrix shows that the majority of the model's predictions are concentrated along the main diagonal, indicating that a large number of images were classified correctly. This pattern indicates that the proposed model achieved correct predictions for most testing samples, reflecting its overall capability to learn discriminative visual representations of human body postures. However, several misclassifications between classes are still observed, particularly among classes that share similar body movements or activity contexts. For example, several ballroom images were misclassified as ballet, modern, or jazz, while resistance training and skateboarding also exhibited misclassifications into other activity classes with similar body poses and movement characteristics. In contrast, activity classes such as aerobic step, golf, rope skipping, skiing, downhill, and soccer showed predictions that were predominantly concentrated along the main diagonal, indicating that these activities were recognized consistently by the proposed model. Such errors are expected in image-based multi-class classification tasks, as the model must distinguish between subtle variations in human poses across different activities.

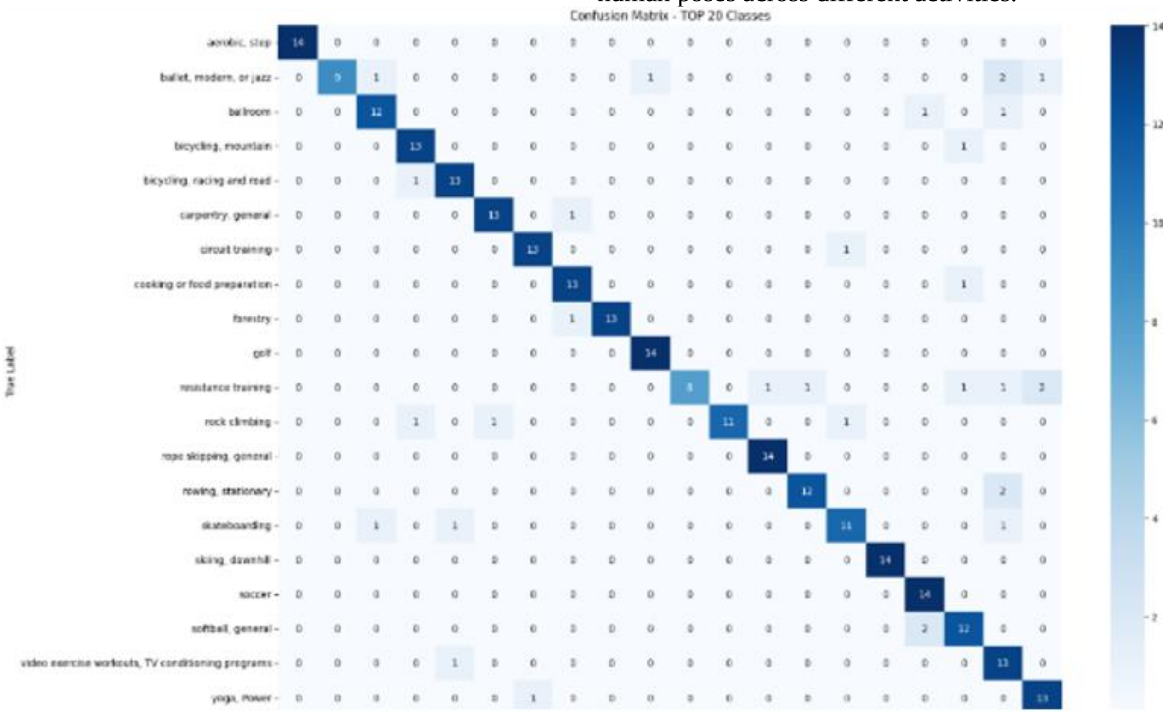

Figure 7. Confusion Matrix

These misclassifications suggest that activities involving similar body poses, body orientation, or movement patterns are inherently more difficult to distinguish using only RGB images. In addition, variations in camera viewpoint, background clutter, partial body occlusion, and differences in image composition within the MPII Human Pose Dataset may further reduce the discriminative visual information available to the model, thereby increasing the likelihood of incorrect predictions. Overall, the testing results demonstrate that the InceptionV3 model trained on the MPII Human Pose Dataset using undersampling and data augmentation techniques was able to achieve strong performance on previously unseen data as shown in Figure 7. Nevertheless, the results also indicate that further improvements are required for certain activity classes to enable the model to produce more consistent and reliable predictions across all activity categories. These findings indicate that the proposed model performs reliably for most activity classes but still exhibits limitations when distinguishing activities with highly similar visual characteristics. Therefore, the model should be interpreted as an effective baseline for image-based human posture classification rather than as a universally optimal solution for all human activity recognition scenarios. The obtained results indicate that the proposed model has the potential to support practical applications such as activity monitoring, sports analysis, rehabilitation assistance, and intelligent surveillance systems. However, further validation using more diverse datasets and real-world environments is still required before practical deployment can be fully realized.

C. Discussion

The evaluation results indicate that the InceptionV3 model employed in this study was able to perform human body posture classification with satisfactory performance. A test accuracy of 0.8893 and a test loss of 0.5587 demonstrate that the model correctly classified the majority of the testing images. These findings are further supported by the macro-average values reported in the classification report, which achieved a precision of 0.9018, a recall of 0.8893, and an F1-score of 0.8877. These metrics suggest that the model maintained a relatively balanced performance across multiple activity classes, although variations in classification performance were still observed among individual classes.

The satisfactory performance achieved by the model can be partly attributed to the undersampling strategy applied during dataset preparation. Balancing the dataset to 140 images per class resulted in a more uniform data distribution, preventing the model from becoming overly biased toward specific classes. This condition is particularly important in multi-class classification tasks, as imbalanced datasets may cause the model to favor classes with a larger number of samples. By applying undersampling, the model was provided with a more equitable learning opportunity across all activity classes. However, this technique also reduced the overall number of available training samples. Therefore, data augmentation played a crucial role in maintaining image diversity and enabling the model to learn a broader range of visual representations.

The preprocessing techniques employed in this study, including image resizing to 224 × 224 pixels and normalization using the preprocess_input function, also contributed to adapting the input images to the requirements of the InceptionV3 architecture. Standardized input dimensions ensured that all images were processed in a consistent format, while normalization adjusted the pixel value distribution to better match the expectations of the pre-trained model. Furthermore, data augmentation techniques such as rotation, zoom, and horizontal flipping improved the model's robustness to variations in image position, orientation, and viewpoint. This is particularly important because human body postures in the MPII dataset may appear under diverse visual conditions. By combining preprocessing and data augmentation, the model was not only able to learn the patterns present in the training images but also to recognize a broader range of body posture variations.

The InceptionV3 architecture employed in this study also played a significant role in the achieved results. InceptionV3 is well known for its ability to extract visual features effectively through its multi-scale convolutional structure. In this study, the original top classification layer was removed (include_top=False) and replaced with additional layers, including Global Average Pooling, Dropout, and Dense layers. This configuration made the model more adaptable to the classification of 20 human activity classes. The Global Average Pooling layer helped reduce feature dimensionality, while the Dropout layer with a rate of 0.2 served to minimize the risk of overfitting. The training results, which showed very high training accuracy and stable validation accuracy, indicate that the model learned meaningful visual representations effectively, although a performance gap between the training and validation datasets remained observable. Such behavior is common in image-based deep learning models, particularly when the number of samples available for each class is relatively limited.

When the performance of individual classes is examined, several activities demonstrate excellent classification results. The aerobic, step and skiing, downhill classes achieved perfect precision, recall, and F1-score values. Similarly, the golf and rope skipping, general classes also exhibited very high performance metrics. These results suggest that images belonging to these classes possess distinctive visual characteristics, making them easier for the model to recognize.In contrast, classes such as ballet, modern, or jazz, resistance training, and video exercise workouts, TV conditioning programs still exhibited lower recall values. This finding indicates that the model continues to face challenges in distinguishing activities that share similar movement patterns or image compositions. Such classification errors are also reflected in the confusion matrix, where some predictions are distributed across visually similar classes, indicating overlap in the visual features learned by the model.

Based on the examples of test image classification results shown in Figure 8, the InceptionV3 model successfully recognized the rock climbing and aerobic, step activities correctly. These results indicate that the model was able to capture the distinctive visual characteristics of both classes, including body posture, movement patterns, and the activity context visible in the images. This finding is consistent with the per-class evaluation results, which showed that several activity classes achieved excellent precision, recall, and F1-score values.

In addition to the correctly classified examples, Figure 9 also presents instances of classification errors made by the model. The ballroom activity was incorrectly predicted as ballet, modern, or jazz, while the skateboarding activity was misclassified as resistance training. These misclassifications indicate that the model still encounters difficulties when different activity classes share similar body poses, movement patterns, or visual contexts. This observation is also consistent with the confusion matrix results, which reveal prediction overlaps among several activity classes that exhibit similar visual characteristics.The confusion matrix reveals that the majority of predictions are concentrated along the main diagonal, indicating that a large proportion of the images were classified correctly. However, several misclassifications are still observed in certain activity classes. These errors generally occur among activities that share similar body postures, camera viewpoints, or movement contexts. This finding highlights that the primary challenge in image-based posture classification is not only recognizing the human body itself but also distinguishing subtle movement details between different activity classes. Therefore, confusion matrix analysis plays an important role in identifying which classes require further improvement in future research.When compared with previous studies, the results of this research can be considered both competitive and relevant. A study conducted by Ogundokun et al. demonstrated that a transfer learning approach combining InceptionV3 with complementary classification methods can achieve high performance in human posture detection tasks. Another study by Ogundokun et al. further showed that the application of data augmentation and hyperparameter optimization can significantly improve human posture classification performance on the MPII Human Pose Dataset. The findings of the present study are consistent with these previous results, as the combination of undersampling, data augmentation, and transfer learning yielded satisfactory performance across 20 human activity classes. Therefore, the use of InceptionV3 in this study can be considered an appropriate approach for image-based human body posture classification.

Despite these promising results, several limitations remain. First, the use of undersampling reduced the total number of available samples compared with the original dataset, thereby limiting the diversity of training examples available to the model. Second, several classes within the MPII Human Pose Dataset exhibit highly similar visual characteristics, making it challenging for the model to distinguish between certain activities accurately. Third, this study employed only a single CNN architecture, namely InceptionV3, and did not include a direct comparison with other state-of-the-art architectures such as ResNet or EfficientNet. Therefore, future research may focus on comparing InceptionV3 with alternative CNN architectures, increasing the amount of training data, exploring different data balancing techniques, or developing more advanced approaches for real-world human activity monitoring and posture recognition applications [28].

Consequently, the findings of this study should not be interpreted as evidence that InceptionV3 is superior to other modern architectures, since no direct comparative experiments were conducted. A comprehensive comparison involving architectures such as ResNet50, EfficientNet, ConvNeXt, or Vision Transformer would provide a more objective assessment of the relative strengths and weaknesses of each model. Another limitation of this study is that no ablation study was performed to quantify the individual contribution of preprocessing, data augmentation, transfer learning, and undersampling to the final classification performance. Therefore, the relative impact of each component could not be evaluated independently and remains an important direction for future research. Furthermore, the proposed model was evaluated only on the MPII Human Pose Dataset. Consequently, its generalization capability across different datasets or real-world scenarios cannot yet be fully established. Future studies should validate the proposed approach on additional public datasets collected under different environmental conditions to assess its robustness and generalization performance.

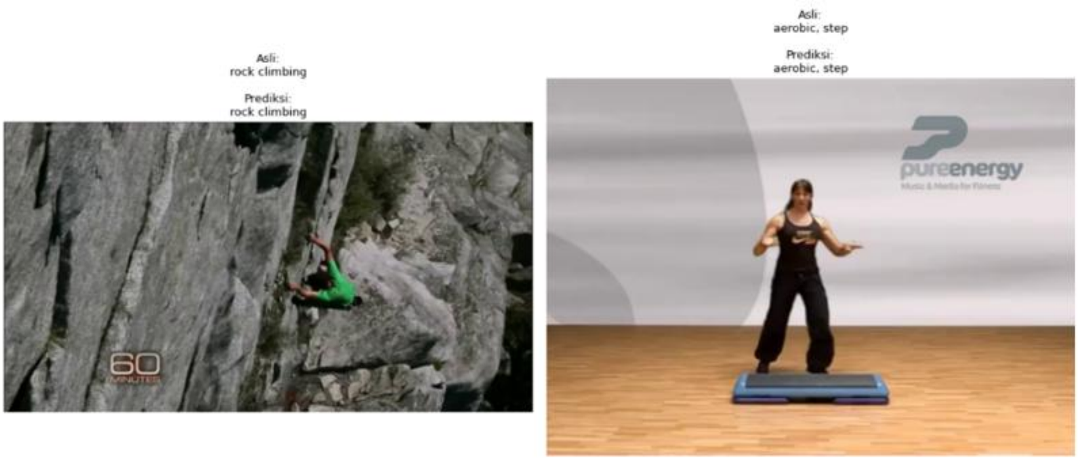

Figure 8. Correctly Classified Test Images

Figure 9. Correctly Classified Test Images

IV. CONCLUSION

Based on the results of this study, the proposed image-based human body posture classification model using the InceptionV3 architecture achieved satisfactory performance on the MPII Human Pose Dataset. The proposed framework involved activity label extraction, selection of the 20 activity classes with the largest number of samples, dataset balancing through undersampling, image preprocessing, data augmentation, and transfer learning. The model, developed using a transfer learning approach with additional layers appended to the final part of the network, achieved satisfactory classification performance on the testing dataset, with a test accuracy of 0.8893 and a test loss of 0.5587. Furthermore, the classification report demonstrated relatively balanced macro-average metrics, with a precision of 0.9018, a recall of 0.8893, and an F1-score of 0.8877, indicating that the model maintained stable classification performance across most activity classes. Nevertheless, the confusion matrix revealed that several classes were still frequently misclassified, particularly those with similar visual characteristics or complex pose variations. This finding suggests that the model still faces challenges in distinguishing certain activities at a more detailed level. Furthermore, the findings of this study are limited to the MPII Human Pose Dataset and should not yet be generalized to other datasets or real-world environments without further validation under more diverse conditions. Another limitation of this study is the reduction in the total number of samples caused by the undersampling process, which consequently limited the diversity of images available for model training. Therefore, future research should focus on comparing InceptionV3 with other CNN architectures, increasing the amount of training data, exploring alternative data balancing techniques, performing validation on additional public datasets, and developing ensemble-based approaches to further improve the robustness and generalization capability of image-based human body posture classification models.

REFERENCES

[1] V. T. Le, K. Tran-Trung, and V. T. Hoang, "A Comprehensive Review of Recent Deep Learning Techniques for Human Activity Recognition," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/8323962.

[2] S. Zhang *et al.*, "Deep Learning in Human Activity Recognition withWearable Sensors: A Review on Advances," *Sensors*, vol. 22, no. 4, pp. 1–43, Feb. 2022, doi: 10.3390/s22041476.

[3] A. Kulikajevas, R. Maskeliunas, and R. Damaševičius, "Detection of sitting posture using hierarchical image composition and deep learning," *PeerJ Comput. Sci.*, vol. 7, pp. 1–20, Mar. 2021, doi: 10.7717/peerj-cs.442.

[4] M. D. Meitantya, C. A. Sari, E. H. Rachmawanto, and R. R. Ali, "VGG-16 ARCHITECTURE ON CNN FOR AMERICAN SIGN LANGUAGE CLASSIFICATION," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 4, pp. 1165–1171, Jul. 2024, doi: 10.52436/1.jutif.2024.5.4.2160.

[5] R. O. Ogundokun, R. Maskeliūnas, and R. Damaševičius, "Human Posture Detection Using Image Augmentation and Hyperparameter-Optimized Transfer Learning Algorithms," *Applied Sciences (Switzerland)*, vol. 12, no. 19, pp. 1–23, Oct. 2022, doi: 10.3390/app121910156.

[6] A. R. Rababaah, "Deep Learning of Human Posture Image Classification using Convolutional Neural Networks," *International Journal of Computer Science and Mathematics*, vol. X, pp. 1–11.

[7] A. Susanto, Y. Kusumawati, E. Dhimas Niagara, and C. A. Sari, "CONVOLUTIONAL NEURAL NETWORK DALAM SISTEM DETEKSI HELM PADA PENGENDARA MOTOR," *2 st Proceeding STEKOM*, vol. 2022, 2022.

[8] M. Mesárošová, O. Mihálik, and M. Jirgl, "CNN Architecture for Posture Classification on Small Data," in *IFAC-PapersOnLine*, Elsevier B.V., Jun. 2024, pp. 299–304. doi: 10.1016/j.ifacol.2024.07.413.

[9] L. Zhang, L. Zhao, and Y. Yan, "A hybrid neural network-based intelligent body posture estimation system in sports scenes," *Mathematical Biosciences and Engineering*, vol. 21, no. 1, pp. 1017–1037, 2024, doi: 10.3934/mbe.2024042.

[10] V. A. Aydın, "Comparison of CNN-based methods for yoga pose classification," *Turkish Journal of Engineering*, vol. 8, no. 1, pp. 65–75, 2024, doi: 10.31127/tuje.1275826.

[11] F. Bin Ashraf, M. U. Islam, M. R. Kabir, and J. Uddin, "YoNet: A Neural Network for Yoga Pose Classification," *SN Comput. Sci.*, vol. 4, no. 2, pp. 1–9, Mar. 2023, doi: 10.1007/s42979-022-01618-8.

[12] R. O. Ogundokun, R. Maskeliūnas, S. Misra, and R. Damasevicius, "Hybrid InceptionV3-SVM-Based Approach for Human Posture Detection in Health Monitoring Systems," *Algorithms*, vol. 15, no. 11, Nov. 2022, doi: 10.3390/a15110410.

[13] H. C. Nguyen, T. H. Nguyen, R. Scherer, and V. H. Le, "Deep Learning for Human Activity Recognition on 3D Human Skeleton: Survey and Comparative Study," *Sensors*, vol. 23, no. 11, pp. 1–33, Jun. 2023, doi: 10.3390/s23115121.

[14] E. Samkari, M. Arif, M. Alghamdi, and M. A. Al Ghamdi, “Human Pose Estimation Using Deep Learning: A Systematic Literature Review,” *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1612–1659, Dec. 2023, doi: 10.3390/make5040081.

[15] R. O. Ogundokun, R. Maskeliūnas, S. Misra, and R. Damasevicius, “A Novel Deep Transfer Learning Approach Based on Depth-Wise Separable CNN for Human Posture Detection,” *Information (Switzerland)*, vol. 13, no. 11, pp. 1–22, Nov. 2022, doi: 10.3390/info13110520.

[16] A. A. Istiqomah, C. A. Sari, A. Susanto, and H. Rachmawanto, “Facial Expression Recognition using Convolutional Neural Networks with Transfer Learning Resnet-50,” 2024. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[17] I. Kareem, S. F. Ali, M. Bilal, and M. S. Hanif, “Exploiting the features of deep residual network with SVM classifier for human posture recognition,” *PLoS One*, vol. 19, no. 12 December, pp. 1–26, Dec. 2024, doi: 10.1371/journal.pone.0314959.

[18] L. Stern and A. R. Fekr, “In-Bed Posture Classification Using Deep Neural Network,” *Sensors*, vol. 23, no. 5, pp. 1–10, Mar. 2023, doi: 10.3390/s23052430.

[19] H. Chen and R. Fan, “Improved convolutional neural network for precise exercise posture recognition and intelligent health indicator prediction,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–16, Dec. 2025, doi: 10.1038/s41598-025-01854-x.

[20] B. A. Khan and J. W. Jung, “Deep Learning-Based Human Activity Recognition Using Dilated CNN and LSTM on Video Sequences of Various Actions Dataset,” *Applied Sciences (Switzerland)*, vol. 15, no. 22, pp. 1–18, Nov. 2025, doi: 10.3390/app152212173.

[21] G. S. Bang and S. B. Park, “Workout Classification Using a Convolutional Neural Network in Ensemble Learning,” *Sensors*, vol. 24, no. 10, pp. 1–15, May 2024, doi: 10.3390/s24103133.

[22] A. S. Talaat, “Novel deep learning models for yoga pose estimator,” *SN Appl. Sci.*, vol. 5, no. 12, pp. 1–13, Dec. 2023, doi: 10.1007/s42452-023-05581-8.

[23] A. Iazzi, M. Rziza, and R. O. H. Thami, “Fall detection system-based posture-recognition for indoor environments,” *J. Imaging*, vol. 7, no. 3, pp. 1–24, Mar. 2021, doi: 10.3390/jimaging7030042.

[24] J. Estrada and L. Vea, “Computer Users Sitting Posture Classification Using Distinct Feature Points and Small Scale Convolutional Neural Network for Humana Computer Intelligent Interactive System During COVID-19,” *Journal of Computer Science*, vol. 19, no. 4, pp. 493–513, 2023, doi: 10.3844/jcssp.2023.493.513.

[25] H. Yoo and K. Chung, “Classification of Multi-Frame Human Motion Using CNN-based Skeleton Extraction,” *Intelligent Automation and Soft Computing*, vol. 34, no. 1, pp. 1–13, 2022, doi: 10.32604/iasc.2022.024890.

[26] M. H. de la Cruz, U. Solache, A. Luna-Álvarez, S. R. Zagal-Barrera, D. A. M. López, and D. Mujica-Vargas, “CNN 1D: A Robust Model for Human Pose Estimation,” *Information (Switzerland)*, vol. 16, no. 2, pp. 1–20, Feb. 2025, doi: 10.3390/info16020129.

[27] X. Cao, X. Wang, X. Geng, D. Wu, and H. An, “An Approach for Human Posture Recognition Based on the Fusion PSE-CNN-BiGRU Model,” *CMES - Computer Modeling in Engineering and Sciences*, vol. 140, no. 1, pp. 385–408, 2024, doi: 10.32604/cmes.2024.046752.

[28] Z. Xu, W. Qu, H. Cao, M. Dong, D. Li, and Z. Qiu, “An Adaptive Human Posture Detection Algorithm Based on Generative Adversarial Network,” *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, 2022, doi: 10.1155/2022/7193234.