IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY SECTION

Received 20 September 2024, accepted 3 October 2024, date of publication 11 October 2024, date of current version 25 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3479165

RESEARCH ARTICLE

PhaseMix: A Periodic Motion Fusion Method for Adult Spinal Deformity Classification

KAIXU CHEN[1], (Student Member, IEEE), JIAYI XU[2], (Member, IEEE), TOMOYUKI ASADA[3,4], KOUSEI MIURA[3], KOTARO SAKASHITA[3], TAKAHIRO SUNAMI[3], HIDEKI KADONE[3,5], (Member, IEEE), MASASHI YAMAZAKI[3], NAOTO IENAGA[6], (Member, IEEE), AND YOSHIHIRO KURODA[6], (Member, IEEE)

[1]Degree Programs in Systems and Information Engineering, University of Tsukuba, Tsukuba 305-8577, Japan
[2]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 113-8654, Japan
[3]Department of Orthopaedic Surgery, Institute of Medicine, University of Tsukuba, Tsukuba 305-8577, Japan
[4]Hospital for Special Surgery, New York, NY 10021, USA
[5]Center for Cybernics Research, University of Tsukuba, Tsukuba 305-8577, Japan
[6]Institute of Systems and Information Engineering, University of Tsukuba, Tsukuba 305-8577, Japan

Corresponding author: Kaixu Chen (chenkaixusan@gmail.com)

This work was supported by the Japan Agency for Medical Research and Development (AMED) under Grant JP23YM0126803.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of the University of Tsukuba Hospital under Application No. H30-087, and performed in with the Declaration of Helsinki.

ABSTRACT Human gait plays a crucial role in medical diagnostics, particularly in the context of adult spinal deformities (ASD). Recent research has applied deep-learning techniques to process video data for ASD diagnosis. Although this method successfully captured dynamic postural information, it does not account for the periodicity and symmetry of motion, particularly the periodicity of gait and the inherent symmetry of locomotor postures in the gait cycle. This omission may have led to the loss of critical information necessary for an accurate diagnosis. To resolve this issue, we present a method that can effectively capture cycle and action symmetry information in motion and exploit the characteristics of gait motion. The proposed method is subsequently trained using the fused data within the deep learning model. Our experiments, performed on a video dataset consisting of 81 patients, showed that our method outperformed baseline approaches. The proposed method achieved an accuracy of 71.43, precision of 72.80, and F1 score of 71.15. The complete set of codes, models, and results can be accessed from the GitHub repository: https://github.com/ChenKaiXuSan/Skeleton_ASD_PyTorch. In this study, we present an innovative video-based approach to support clinical diagnoses of gait disorders, with a focus on periodic motion and postural symmetry. The application of deep learning in diagnostic methods has been enhanced by utilizing the kinematic properties of walking movements. The experimental results demonstrate that the proposed method can provide more accurate diagnostic results when using limited video data.

INDEX TERMS
Adult spinal deformity, deep learning, human action recognition, healthcare, gait posture, periodic motion, motion symmetry.

I. INTRODUCTION

Gait is a fundamental and ubiquitous form of human locomotion that involves complex coordination and dynamic stability, both of which are reflected in the gait cycle. Gait analysis techniques hold clinical significance, as they allow for the assessment of walking characteristics. The gait cycle, which is characterized by phases such as heel strike, midstance, and toe-off, embodies a rhythmic pattern that reflects the coordination and balance of human locomotion. These cyclic movements are not merely mechanical; they encapsulate essential biomechanical information crucial for

The associate editor coordinating the review of this manuscript and approving it for publication was Chulhong Kim.

the diagnosis and treatment of various musculoskeletal and neurological conditions. Analysis of the periodicity and symmetry of gait cycles in video-based classification tasks is of paramount importance in the field of deep learning-based classification systems. The use of deep-learning techniques, which can recognize subtle changes in gait dynamics, is a key objective in the development of robust classification models for identifying gait abnormalities and disorders. This presentation highlights the importance of integrating periodicity and symmetry features in the development of such models, which can improve classification accuracy.

In medicine, spinal disorders have the potential to alter a patient's gait. The accurate diagnosis of spinal diseases and related deformities is critical because of their ability to cause significant health issues. These conditions carry notable physical and functional implications, leading to discomfort, limited mobility, nerve impairment, paralysis, and constraints on daily activities. As spinal diseases advance, they can result in more severe complications, such as persistent pain and neurological dysfunction. Adult spinal deformity (ASD) is defined as an abnormal curvature or malalignment of the spine that occurs during adulthood. Moreover, ASD can exert a significant impact on health-related quality of life (HRQoL) [1], [2], [3]. These deformities can cause symptoms such as back pain, stiffness, numbness or weakness in the legs, difficulty walking, and changes in posture or appearance. In severe cases, ASD can affect organ function and quality of life. Presently, research is being conducted to identify novel methods for studying and diagnosing diseases by analyzing the walking posture of patients captured on video [4] or using sensors [5], [6], [7], [8]. The aforementioned studies were conducted using walking posture, but did not address the kinematic properties of the human gait, such as the periodicity of motion and symmetry of locomotor postures.

Periodic motion is used to describe body movements during gait. Human gait encompasses a continuous cycle of movements in which various muscles contract and relax in a coordinated manner to propel the body forward while maintaining balance and stability. This complex interplay among neural signals, muscular contractions, and skeletal movements ensures efficient and fluid locomotion, allowing individuals to navigate their environment with ease and precision. Multiple studies have provided evidence to support the pivotal role of gait analysis in the diagnosis of spinal disorders [9], [10], [11]. These studies employed gait analysis to establish and elaborate on the clinically significant elements of typical human walking patterns, evaluate gait characteristics in individuals affected by spinal conditions, and identify the aspects of human gait that are correlated with pre- and post-surgical patient function and results.

Gait posture symmetry refers to the uniformity and balance between the movements of the left and right limbs. This metric is important for several reasons. First, symmetry is often used as an indicator of normal and healthy gait patterns. Deviations from a symmetrical gait can signify underlying pathologies, injuries, or biomechanical inefficiencies. In video-based diagnostics, the analysis and capture of movement symmetry are of paramount importance. For instance, gait asymmetries are commonly observed in individuals with stroke, cerebral palsy, or lower-limb amputations. If these asymmetric features can be captured, disease classification can be performed more accurately. Consequently, we posit that the integration of symmetry-related features may facilitate more accurate disease diagnosis.

In this study, we present a video-based fusion technique that aims to improve our analysis of periodic motion and posture symmetry using a single-camera setup. Our method provides a comprehensive and accessible solution for gait analysis by focusing on periodicity in motion and symmetry in postures. Unlike traditional methods that require multiple sensors or costly hardware, our approach employs advanced computer vision and machine learning techniques to analyze motion patterns, as illustrated in Figure 1. The major contributions of this study are as follows:

- We propose a method for image-level fusion, which considers periodicity and symmetry features.
- We propose a video-based method for detecting the walking phase to accurately perform feature fusion.
- We conducted a comparative analysis of predictions generated by the proposed method using different backbone architectures. This analysis aimed to demonstrate the importance of periodicity and symmetry in periodic motion in model detection.

The remainder of this paper is structured as follows. Section II reviews the related work, covering advancements in data fusion techniques and their application to current diagnostic methods in medicine. Section III introduces our proposed cyclic-motion-based approach, detailing its core components and design. In Section IV, we outline the experimental setup and datasets used for evaluation, followed by the presentation of experimental results and an ablation study. Section V discusses the findings, emphasizing performance improvements and identifying limitations. Lastly, Section VI concludes the paper, summarizing key contributions and suggesting potential directions for future research.

II. RELATED WORK

In this section, we first introduce the basics of data fusion and then extend the content to current diagnostic methods in medicine. We will also introduce related research on feature fusion under the deep learning framework, review the research on gait analysis for spinal diseases, and introduce data on periodicity and motion symmetry in periodic motion.

A. DATA FUSION METHOD

The fusion method at the data level involves combining information from multiple sources or modalities with the objective of enhancing the discriminative power of the recognition system. By fusing information from different

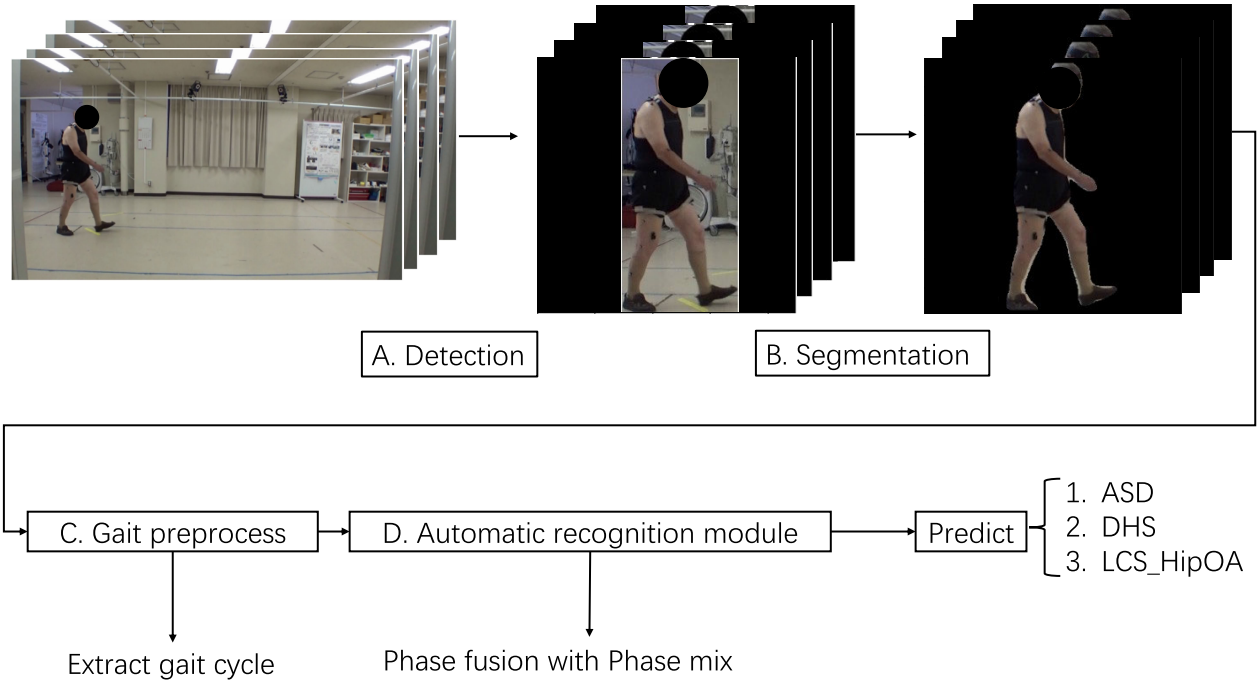


FIGURE 1. Flowchart of the proposed approach. A: Determining the patient's position from the original image. B: Isolating the patient's body from the background. C: Delineating the walking cycle and harmonizing the dataset, ensuring fair representation of various classes. D: Fusing the subsequent walking cycle and feeding it into the model for learning. The patient's head was mosaiced.

TABLE 1. Comparison of current diagnostic methods. We propose a method that employs a single camera, does not necessitate contact with the patient's body, does not require physician review, exhibits low device dependence, and considers periodicity and movement symmetry in walking locomotion.

Method	Diagnostic Process	Contact Body	Physician Review	Facility Dependency
Quantitative analyses [1]	The patient completes the requisite paperwork, after which the physician formulates a diagnosis based on the information provided.	No	Yes	Questionnaires
Radiographs [12]	Patient diagnosed with radiographs and reviewed by doctor	Yes	Yes	X-ray machine
Dynamic methods [5], [13]	Sensors placed on patient's body, followed by diagnosis based on captured results	Yes	Yes	Dynamic capture system
Machine-learning method [4]	The use of RGB video, with a focus on full body features.	No	No	Camera
Proposed method (Ours)	The RGB video was analyzed in order to ascertain the periodicity of the motion and the symmetry of actions within the motion.	No	No	Camera

sources, the system can capture complementary details and improve recognition robustness under various conditions. The following section presents an overview of the various data fusion methodologies.

Feature-level fusion extracts features independently from each modality and concatenates or combines them before feeding them into the algorithm. Reference [14] demonstrates the feasibility of such fusion and emphasizes the importance of further research in this direction. They discussed feature-level fusion in three different scenarios: fusion of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) coefficients of a face; fusion of LDA coefficients corresponding to the R, G, and B channels of a face image; and fusion of face and hand modalities. Reference [15] presents a framework for simultaneously protecting multiple biometric templates of a user as a single secure sketch. The practical implementation of the proposed feature-level fusion framework was demonstrated using two well-known biometric cryptosystems: fuzzy vault and fuzzy commitment. A detailed analysis of the trade-off between matching accuracy and security in the proposed multibiometric cryptosystems was presented, which used two different databases (a real and virtual multimodal database) containing the three most popular biometric modalities: fingerprint, iris, and face. Reference [16] recorded electroencephalogram (EEG) signals from 86 patients with depression and 92 normal controls while they were exposed to various audio stimuli. Linear and nonlinear features were then extracted and selected from each modality to obtain the distinctive features. A linear combination technique was employed to fuse these EEG features into a global feature vector and identify several powerful features for depression recognition. Genetic algorithms were used for feature weighting to further improve the performance of the recognition framework.

Reference [17] presented a robust approach for human activity recognition (HAR) using multimodal feature-level fusion. The proposed method integrates features extracted from different modalities, such as accelerometer data, gyroscope data, and video frames, at the feature level to improve the accuracy and robustness of HAR systems. Reference [18] presented a method for improving HAR by fusing data from depth cameras and inertial sensors. The proposed method integrates information from both modalities to capture the complementary aspects of human motion, thereby improving the accuracy and robustness of action recognition systems.

In addition, there are specific image-based data fusion methods, namely, pixel-based methods. In pixel-level fusion, the pixel intensities or representations from different modalities are combined using mathematical operations. Reference [19] concluded that image fusion utilizing wavelets with higher levels of decomposition performed better according to some metrics, while PCA showed superior performance in other metrics. These findings highlight the importance of selecting appropriate fusion techniques based on the desired performance criteria. Reference [20] presented a novel approach for pixel-level fusion using the signal sparse representation theory. This method focuses on effectively integrating complementary information from multiple source images to produce a single fused image that retains significant features from all inputs. Reference [21] addressed the limitations of existing image fusion techniques, which include low efficiency, long running times, lack of detailed image information, and poor image fusion quality. To overcome these issues, the authors proposed a pixel-level image-fusion algorithm based on edge detection. Reference [22] proposed a novel regularization strategy for training strong classifiers with localizable features. This strategy replaces image patches with patches from other images to augment the training data, thereby encouraging the model to learn discriminative features from both the original and mixed images. Reference [23] presented a simple yet effective data augmentation technique that extends beyond traditional empirical risk minimization methods. It can generate augmented training examples by linearly interpolating between pairs of examples and their labels. By encouraging the model to learn linear combinations of input data and labels, mixup regularizes the model and improves its generalization performance. Reference [24] presented a tobacco leaf classification method using a fuzzy clustering-based neural network combined with multiple histogram analyses across different color spaces. They utilize PCA for dimensionality reduction and employ fuzzy C-means clustering for training, with a Softmax function to generate class probabilities. The method has been validated on real industrial images, demonstrating competitive performance compared to traditional classifiers. Reference [25] introduced fractional Riesz potentials based on chirp functions, exploring their relationship with fractional Fourier transforms and fractional Laplace operators. They establish the boundedness of these potentials on rotation-invariant spaces and provide numerical simulations illustrating their applications in image processing. A key contribution of their work is a novel image encryption method using double phase coding, which enhances security by offering greater degrees of freedom. Reference [26] proposed a dimensionality reduction methodology for reinforced fuzzy rule-based neural networks (FRNNs), leveraging a correlation coefficient-based feature selection strategy and fuzzy clustering to manage high-dimensional data. This approach reduces computational complexity and streamlines the design of radial basis function neural networks (RBFNNs) by utilizing fuzzy clustering to determine receptive field parameters. In contrast to conventional RBFNNs, FRNNs integrate dimensionality reduction and L2-norm regularization to improve generalization. The effectiveness of the method has been demonstrated on 28 real-world benchmark datasets, highlighting its superiority over traditional models.

The objective of the present study was to examine image-based pixel-level fusion methods, with a particular focus on improving the performance and reliability of human gait recognition systems by leveraging complementary information obtained from multiple sources. To achieve this, we propose a method called *PhaseMix*, which is designed to facilitate the efficient fusion of period information in periodic motion and symmetry information in posture.

B. CURRENT METHODS USED IN DIAGNOSTICS

The clinical diagnosis of ASD involves the recognition and assessment of atypical spinal curvature and alignment in adult patients. ASD encompasses a spectrum of spinal conditions that can lead to deformities such as scoliosis (sideways curvature), kyphosis (forward rounding of the upper back), or a combination of these. The following section presents the current methods employed in the diagnosis of ASD. These methods can be categorized into three main groups: static, dynamic, and deep learning.

1) STATIC METHOD

Quantitative analyses are commonly used in medical practice. Lateral standing full-spine radiographs currently serve as the gold standard for diagnosing ASD and other spinal diseases [12], [27], [28], [29]. However, these radiographic assessments are static evaluations conducted in a standing position, which can mask underlying issues due to compensatory mechanisms in other joints. The objective of the quantitative analysis of digital images [30], [31] is to evaluate the reliability of a new computer-assisted measurement method which utilizes digital reconstruction images to assess scoliotic spines. Computer-assisted digital analysis offers the potential for rapid, precise, and comprehensive evaluations. Despite their benefits, these techniques face significant accessibility challenges in medical facilities. Consequently, currently available static evaluation methods are insufficient for the comprehensive diagnosis of ASD. Therefore, there is a pressing need for more accessible and dynamic diagnostic methods.

2) DYNAMIC METHOD

Several studies [6], [7], [8], [32] have highlighted the importance of dynamic motion systems in the diagnosis of ASD. A central tenet of this dynamic approach is capturing alterations in joint movements during walking. Reference [32] utilized motion analysis to evaluate postural changes during standing and walking in ASD patients. They conducted reassessments in ten patients following corrective spinal surgery, analyzing continuous kinematic and kinetic data using statistical parametric mapping. References [7] and [8] proposed a diagnostic method by analyzing changes in the joints during walking using the Vicon MX system. By placing reflective markers on specific anatomical landmarks and strategically employing high-speed cameras around a walkway or treadmill, the system captures and analyzes a subject's movements. This information is then used to assess normal gait patterns, identify abnormalities, and evaluate treatment effectiveness. Although 3D motion analysis systems provide invaluable insights into understanding human gait, they are not without limitations. The acquisition and installation of these systems can be financially onerous, requiring the use of specialized equipment and software. An alternative and more cost-effective approach involves the use of electromyography (EMG) sensors in lieu of a 3D motion gait system. EMG has been employed to investigate the relationship between dynamic and static spinal gait postures [7], [33]. EMG sensor systems are used to quantify and record the electrical activity of skeletal muscles. Researchers have placed electrodes on specific muscles to analyze muscle activation patterns and assess the involvement of different muscle groups during walking. Typically, the system comprises 12 EMG sensors attached to a patient's trunk and lower limb muscles before each experiment [7]. Reference [33] compared postoperative changes in trunk and lower extremity muscle activity between ASD patients and age-matched controls using surface EMG. Significant differences in muscle activation patterns were observed, highlighting the impact of surgical correction on trunk and lower-extremity muscle function in ASD patients. Although the use of EMG sensors has provided valuable contributions to gait posture recognition, it has some limitations. There are also temporal and infrastructural constraints, and there is a lack of consensus on an optimized evaluation method.

In contrast, our proposed video-based diagnostic approach focuses on capturing motion characteristics between video frames and integrating key movement features such as motion periodicity and posture symmetry. Compared to traditional sensor-based methods, our approach provides several advantages, such as being non-contact and eliminating the need for physical interaction with the patient. The video-based method provides convenience and broad applicability, allowing for easier implementation in various settings. It also enables comprehensive body information capture, monitoring full-body dynamics rather than focusing on localized measurements. Furthermore, video-based diagnostics reduce maintenance and usage costs, as they rely on standard camera equipment rather than specialized hardware.

3) MACHINE-LEARNING METHOD

Given the dynamic nature of the spine and alterations in gait and posture, researchers have investigated the possibility of directly capturing changes in individual joints from videos using machine-learning techniques. Reference [34] aimed to predict surgical complications in patients undergoing elective ASD procedures using machine learning techniques. Advanced computational methods were applied for preoperative risk assessment and to improve patient outcomes after spinal surgery. Reference [35] presented a novel approach for predicting blood transfusion requirements after ASD surgery, utilizing machine learning techniques and a large, multi-institutional dataset. The results of this study have the potential to improve preoperative risk assessment and patient management in ASD surgery, thereby enhancing patient outcomes and reducing the need for blood transfusions. Reference [4] introduced a novel approach for classifying ASD utilizing two-stage video-based convolutional neural networks (CNNs). This method aims to enhance ASD diagnosis accuracy by analyzing video data. The study demonstrated the promising potential of CNNs in medical image analysis and disease classification for ASD diagnosis.

However, no existing studies have considered periodicity in periodic motion and posture symmetry, which is a significant limitation when analyzing actions based on periodic motion. Periodic motion and posture symmetry are fundamental aspects of biomechanics, particularly in the context of gait analysis and rehabilitation. They serve as cornerstones for understanding the biomechanics, kinematics, and dynamics of locomotion.

A gait cycle encompasses a sequence of coordinated movements involving multiple joints, muscles, and skeletal structures. Understanding the nuances of the gait cycle offers insights into various aspects of human movement, including balance, stability, efficiency, and functionality [36]. Periodicity is a fundamental concept in the analysis of biomechanical motion, particularly in gait analysis. The term ''periodic motion'' refers to movements that repeat at regular intervals, such as walking or running. A significant body of research has been dedicated to the development of methodologies for detecting and analyzing periodicity in human motion. [37] introduced a novel approach for detecting periodic patterns in gait using wearable sensors. This approach has been demonstrated to be effective in the clinical setting for monitoring patients with gait abnormalities. The term ''posture symmetry,'' particularly in the context of periodic motions such as gait, is of paramount importance for maintaining balance and efficiency in human movements. Reference [38] introduced the symmetry function as a novel and effective tool for evaluating gait symmetry in trans-femoral amputees. The study demonstrated that this function could accurately identify asymmetries in gait patterns, correlating well with

subjective clinical assessments. Recent advances in sensor technology and machine learning have significantly enhanced the capabilities of periodic motion analysis tools. Video-based gait diagnostic systems commonly employ either generalized Eigenvector-based image methods or pose estimation techniques [39], [40]. We created a comparative table illustrating the existing diagnostic methods for ASD and other spinal disorders, as shown in Table 1.

III. METHODS

This section outlines the proposed cyclic-motion-based approach for differentiating ASD from other spinal disorders. Figure 1 provides an overview of the proposed method, which can be divided into five parts. Figure 1A depicts the process of locating the patient's position from the original image, while Figure 1B focuses on separating the patient's body from the background. Figure 1C defines the walking cycle and balances the dataset to ensure equitable representation of different classes. Finally, Figure 1D shows the mixing of periodic motion features from the walking cycle and inputting them into the model for learning purposes. The following section examines the definition of the gait cycle definition module and feature mixing methodology.

A. GAIT-DEFINED MODULE

The definition of the gait cycle is a crucial step in the analysis of human gait because it allows for the identification of pivotal phases or components within the cycle. These phases encompass the heel strike, toe-off, and swing phases, each of which exerts a significant influence on the overall walking pattern. An array of techniques can be employed to precisely delineate these phases. In this study, we focused on pattern recognition methodologies that leverage predefined templates or patterns to identify specific events in RGB video footage. The conventional method entails the identification of the gait cycle within a video sequence through the determination of the maximum distance between the legs. In this study, we introduce a novel gait cycle recognition approach that integrates object detection and pose estimation to effectively identify a patient's gait cycle index from a video. The outcomes of the bounding box (bbox) and pose recognition were amalgamated to conclusively determine the gait cycle, as shown in Figure 2. Figure 2A depicts the processing flow of the algorithm. We employed YOLOv8 [41] to obtain the bounding box and pose estimation results. To determine the gait cycle frame index, it is necessary to combine the width of the bounding box with the distance between the legs, which is obtained from pose estimation. When only the bounding box was used for the determination, it was observed that the width measurement was subject to inaccuracies owing to hand movements (Figure 2B). Similarly, relying solely on the distance between the two feet from pose estimation resulted in misjudgments when the foot key points were inaccurate (Figure 2C). Therefore, we integrated both the pose estimation and bounding box methods to derive a more precise index for the gait cycle. The entire algorithm training process is described in Algorithm 1.

Algorithm 1 The Gait-Defined Module for the Proposed Method

Input: one person video frames
Input: max_distance ← determined from all frames.
Output: defined gait cycle index

```
function FIND_GAIT_CYCLE_INDEX(video frames)
    defined_gait_cycle_index ← []
    for each frame in video frames do
        bbox_width ← bbox                ▷ get from detection
        foot_distance ← foot key points  ▷ get from pose estimation
        if |foot_distance - max_distance| < bias then
            if bbox_width <= foot_distance then
                defined_gait_cycle_index ← frame number
            else if bbox_width > foot_distance then
                next frame   ▷ this is not the optimal distance.
            end if
        end if
    end for
    Return defined_gait_cycle_index
```

B. PHASE-MIXING METHOD

The *PhaseMix* approach combines pertinent features extracted from disparate phases of the cycle during periodic motion. We utilized a pixel-level fusion method to integrate features derived from distinct cycle stages into a unified image representation. The integration of features from disparate cycle stages enabled the acquisition of more comprehensive motion information. The fused image contained the spatial and temporal aspects of the gait cycle, including posture symmetry and periodic motion, while enhancing the learning ability of the model.

Information regarding the period of the periodic motion and posture symmetry is shown in Figure 3. In the figure, A and B are combined to represent a cycle of motion, designated in our study as the gait cycle. Figure 3A shows the initial phase of the cycle, specifically the first half of the walking cycle. Figure 3B depicts the subsequent phase of the cycle, which is the second half of the gait cycle. In the figure, orange represents the movements on the right side of the body, including the right hand, right arm, right leg, etc. Green represents the movements on the left side, including changes in the movements on the left side of the body (left hand, left arm, left leg, etc.). As shown in the figure, the first and second halves of the periodic motion are symmetrical. The methodology employed is indicated by the red dotted line in the figure. To obtain the periodic motion and posture symmetry of movements in the gait cycle, images of different phases were fused. To ensure that the images resulting from the fusion process retained the posture symmetry, it was necessary to select images A and B with precisely the same position and orientation. This ensures that the image resulting

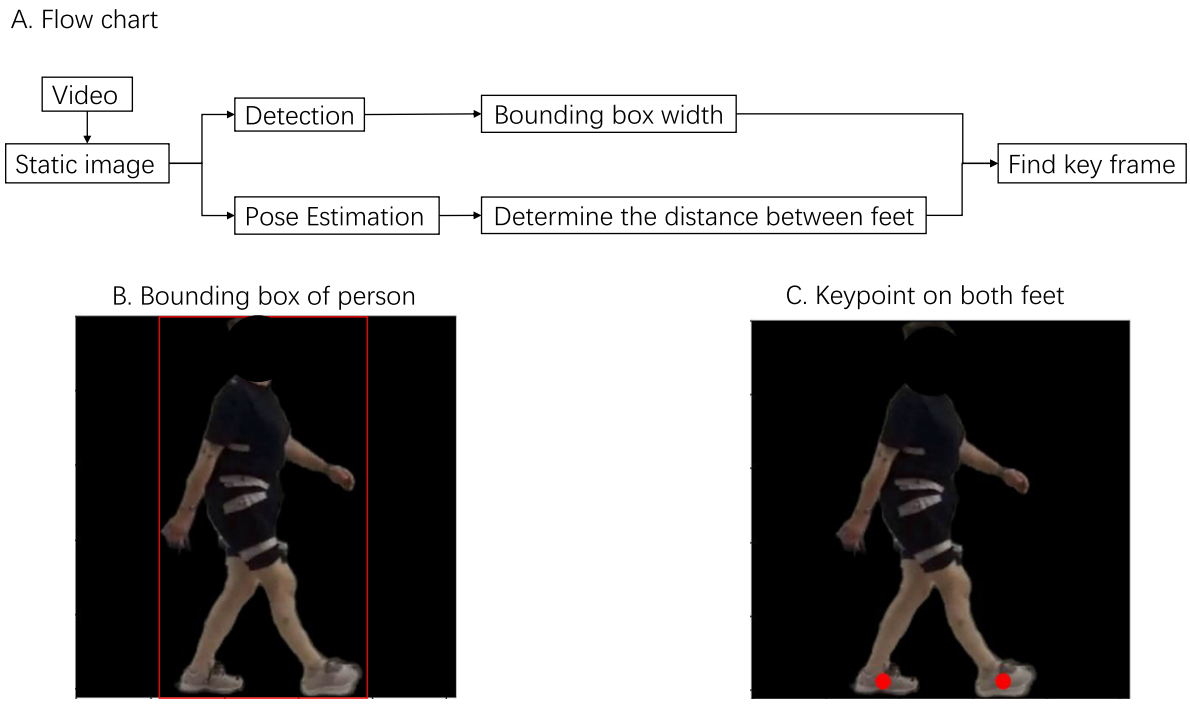


FIGURE 2. Flow chart of the gait-defined module. Panel A illustrates the processing flow of the algorithm, while Panels B and C show the results of the bounding box of the person and key point on both feet, respectively. The patient's head was mosaiced.

from the fusion process (Figure 3C) contains the posture symmetry inherent to cyclic motion.

A comprehensive explanation is presented in Figure 4 and Algorithms 2. Initially, the gait cycle index derived from the gait-defined module (Section III-A) was employed to differentiate between the left swing (initial half of the walking cycle) and right swing (subsequent half of the walking cycle) phases. Figure 4A involves ascertaining the maximum width within a single phase. Figure 4B standardizes the remaining frames based on the maximum width identified in Figure 4A to ensure size uniformity across all frames. In this approach, the cropping method is employed to standardize the size of all frames based on the maximum bounding box width in the current gait cycle, which ensures that subsequent fusion can occur. Figure 4C shows the calculation of the average number of frames required for the synthesis from various phases. Figure 4D shows the pixel-level fusion process. To achieve fusion, two normalized images were combined along the width dimension, and the images were then centered at the same position. In the fused images, the left side indicates the left swing phase, while the right side indicates the right swing phase. In the context of the temporal dimension, given that the time series was not destroyed, the fused images were maintained in the same sequence within the time series.

Algorithm 2 The PhaseMix for the Proposed Method

Input: one person video frames, gait cycle index, bbox
Output: phase mixing frames

```
funstion PHASE_MIX(video_frames, gait_cycle_index, bbox)
    left_swing_frames ← defined by gait cycle index
    right_swing_frames ← defined by gait cycle index

    for frames in (left_swing_frames, right_swing_frames) do
        max_bbox_width ← max(frames, bbox)
                ▷ find max bbox width of the frames
        normalized_frames ← crop (frames, max_bbox_width)
                ▷ crop by max_bbox_width_frame

    end for
    for frames in gait cycle do
        fused_image ← (normalized_left_swing_frame, normalized_right_swing_frame)
            ▷ merge different frames by width dimension
    end for
    Return fused_image
```

IV. EVALUATION

This section begins with a presentation of the gait video dataset utilized in this study, followed by a comprehensive description of the experimental configuration. This is followed by a quantitative analysis of the accuracy of the gait-defined module. The subsequent sections will analyze and compare the proposed method with an existing baseline of action-recognition systems. Finally, two distinct ablation experiments will be presented. The first ablation experiment examined the significance of periodic motion and posture

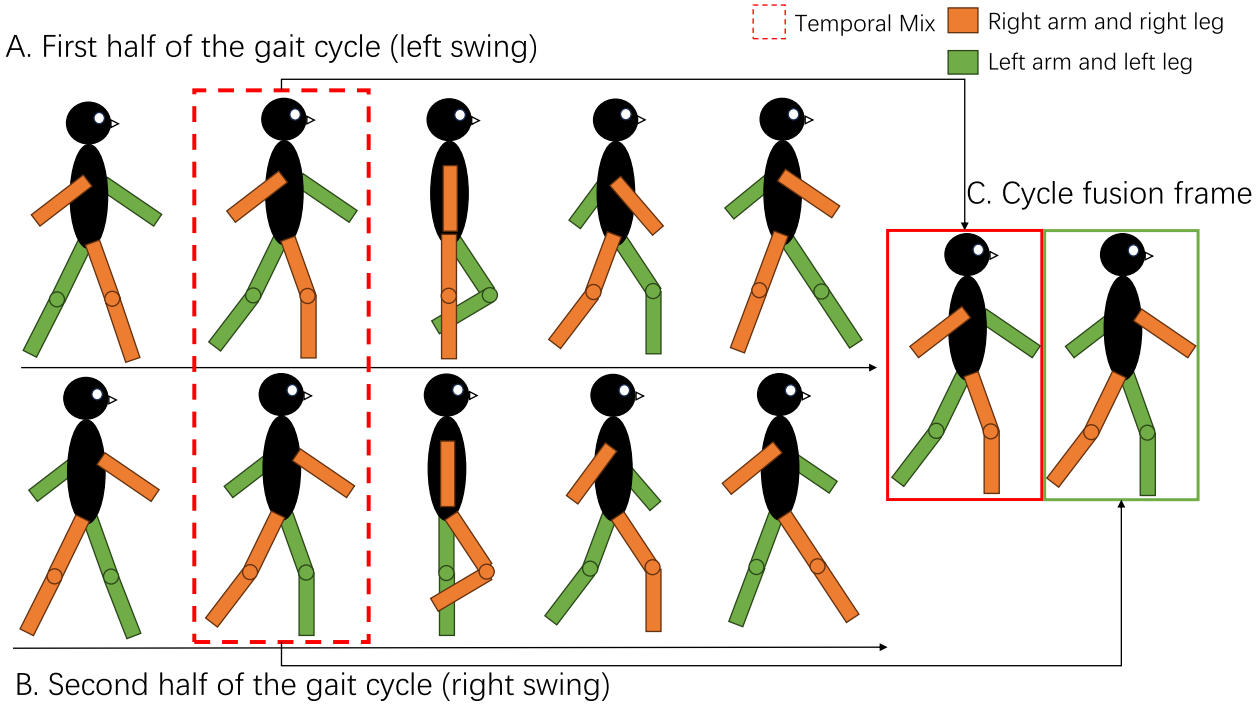


FIGURE 3. The introductory image of *PhaseMix*. The combination of (A) and (B) represents a walk cycle. (A) The initial portion of the cycle. (B) The subsequent portion. To achieve a symmetrical image, it is necessary to select images from (A) and (B) that are identical in terms of their position and orientation. This is illustrated in (C). This ensures that the resulting image (C) after fusion includes the symmetric movements inherent to the cyclic motion.

TABLE 2. Dataset details.

Label	Abbreviations	Diseases detail name	Patient	Clipped video	Total
ASD	ASD	Adult Spinal Deformity	54	1046	1046
DHS	DHS	Dropped Head Syndrome	16	587	587
LCS_HipOA	LCS	Lumbar Canal Stenosis	9	260	324
	HipOA	Hip Osteoarthritis	2	64	

symmetry in the video data. In the second experiment, several training strategies were compared.

A. DATASET

To advance the research on the use of gait and spinal posture data for medical analysis, we prepared a dataset comprising videos utilized in [4], [5], [7], [8], and [13]. This study was approved by the Ethics Committee of the University of Tsukuba Hospital (H30-087) and was conducted in compliance with contemporary amendments to the Declaration of Helsinki and ethical guidelines.

The dataset encompassed video recordings of individuals with ASD and a control group, capturing their walking patterns and spinal postures. These videos aimed to offer valuable insights into the correlation between gait abnormalities and spinal posture. Participants were recruited for this gait posture video dataset from ASD clinics or research centers. Informed consent was obtained. The participants were instructed to walk naturally while being recorded using high-quality video cameras fixed at specific angles. An oval-shaped pathway was designated for walking within an indoor space, spanning a distance of 10 meters. The patients traversed the pathway from left to right, or vice versa, with their walking posture captured from a side view using a camera stationed at a fixed location. Figure 5 illustrates the walking patterns observed in individuals with various diseases.

Each video was recorded at a frame rate of 30 frames per second and a resolution of 1920 × 1080 (high-definition resolution), with the video quality set to high. Videos were recorded in MP4 format and encoded using H.264. The participants were instructed to record a video from the start of their walk until they were no longer able to continue. The duration of the original videos ranged from 60 to 300 seconds. Consecutive frames from each original video were extracted to remove those that did not contain the individuals. This resulted in 1,957 individual video clips, each ranging from 2 to 10 seconds in duration. The gait posture label assigned to each video was determined based on the diagnosis provided by spine surgeons, which was derived from diagnostic radiographic assessments using standing whole-spine X-ray images and the clinical symptoms exhibited by the patients. Table 2 provides detailed information regarding the dataset, which comprises videos of 81 patients

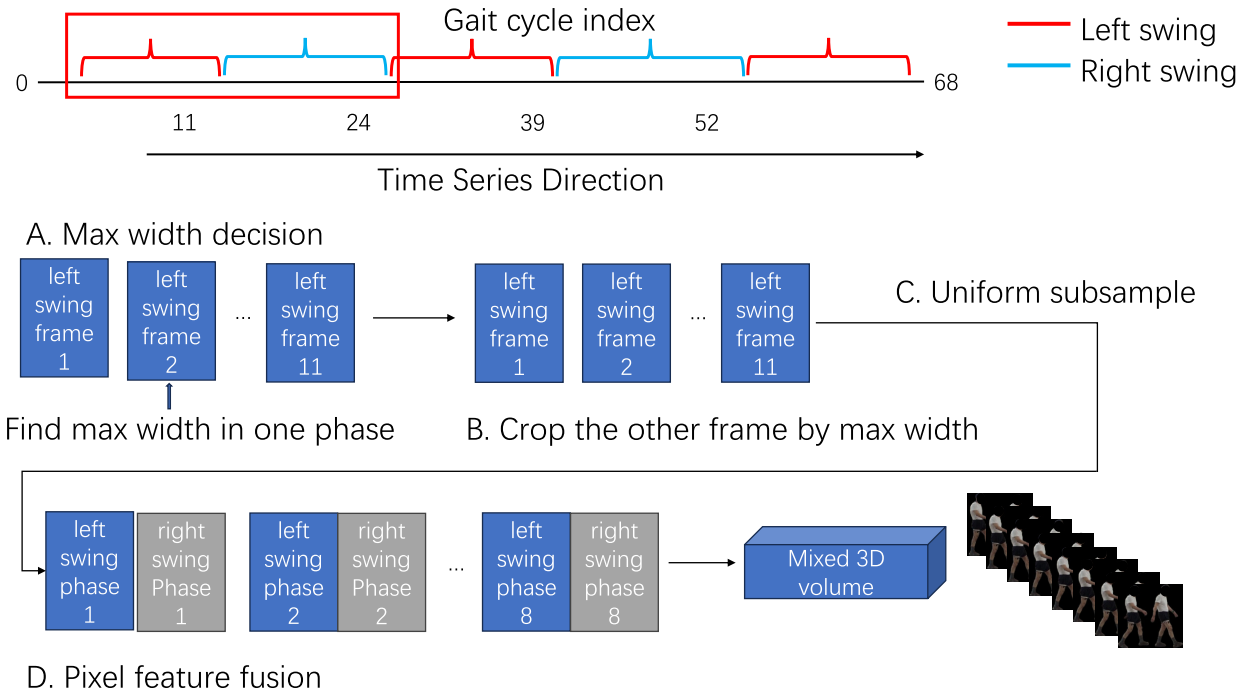


FIGURE 4. Flow chart of the *PhaseMix* method. Initially, the gait cycle index acquired from the gait-defined module (Section III-A) is employed to identify the left and right swing phases. (A) The maximum width is determined within a single phase. (B) The other frames are normalized with the maximum width determined in (A) to ensure uniform size across all frames. (C) The mean number of frames is calculated from the different phases. (D) Pixel-level fusion is executed. The left side represents the first half of the gait cycle (left swing), while the right side signifies the second half of the gait cycle (right swing), ensuring that the time series of the distinct phases remains unchanged during the composition. The patient's head was mosaiced.

(61 males and 20 females) aged between 22 and 84 years, with a mean age of 70 years.

The facility specializes in gait analysis for a range of musculoskeletal pathologies, including ASD, lumbar canal stenosis (LCS), dropped head syndrome (DHS), and hip osteoarthritis (HipOA). In a study conducted by [4], the authors differentiated between ASD and non-ASD cases. However, in our study, we extended the classification categories from two to three for several reasons:

TABLE 3. Quantitative analysis of the gait-defined module.

Disease	Bounding box	pose estimation	proposed method
ASD	0.14	13.16	0.0
DHS	0.15	13.33	0.0
LCS	15.25	15.25	0.17
HipOA	0.14	14.82	0.0

- Enhanced categorization provides more detailed insights, which aids in the accurate diagnosis of various conditions.
- Through our observations, we noted that the features exhibited by LCS and HipOA are more closely related and distinct from the other classifications.
- Considering the challenge of maintaining balanced training data, dividing the conditions into three distinct classes facilitates more effective training of the model.

B. GAIT-DEFINED MODULE

The objective of the gait-defined module was to evaluate the precision of the gait cycle by examining diverse analytical techniques. Accurate acquisition of the gait cycle index is of utmost importance to ensure the accuracy of subsequent feature fusion. In this study, we evaluated the efficacy of three methods for determining the gait cycle index: bounding box, pose estimation (bipedal), and the proposed method. The Euclidean distance was employed as a means of comparison, with smaller numbers indicating that the comparison method was closer to the ground truth and thus more accurate. The calculation was performed by comparing the index of the walking cycle predicted using the method in question with the index of the ground truth. To ensure that the lengths of the prediction and target lists are equal, it is necessary to perform the operation of complementing zero at the nearest position to the ground truth. This discrepancy occurs because the result of the prediction using bounding box or pose estimation may indeed have some indices, which could otherwise result in unequal lengths between the two lists. For the ground truth, a frame-by-frame observation method was employed to identify the gait cycle index position. To facilitate the quantitative analysis, representative videos from the four diseases included in the dataset were selected.

Table 3 presents the results of the quantitative analyses of the various methods. As shown in the table, the proposed method demonstrates satisfactory prediction outcomes, with

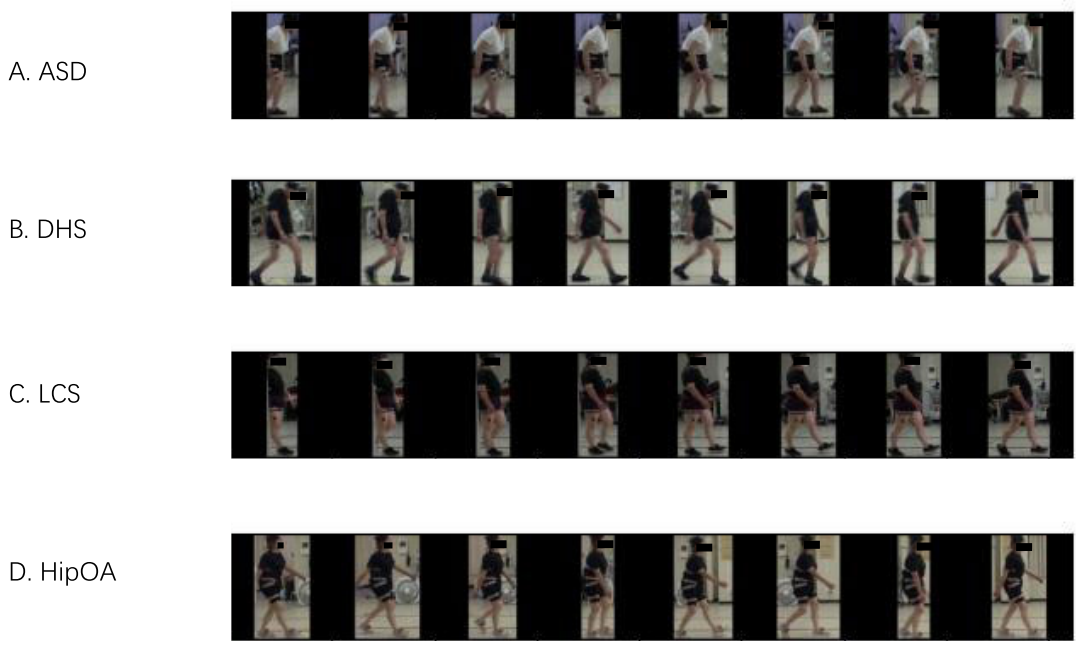


FIGURE 5. Demonstration of walking maneuvers for different diseases. (A) ASD, (B) DHS, (C) LCS, and (D) HipOA. The patients' faces were mosaiced.

a value of zero signifying precise alignment between the predicted gait cycle index and the ground truth. Values exceeding 10 are caused by an excess of missing positions within the prediction list. Our methodology has the potential to mitigate the issues of index loss and non-detection, thereby facilitating the generation of a more precise index of the gait cycle.

C. EXPERIMENTAL CONFIGURATION

Training the model (depicted in Figure 1D) involved the utilization of various machine-learning models. The initial training of the proposed method was conducted using a pretrained model of the Kinetics-400 dataset [42]. All experiments were conducted in identical hardware and software environments. Video clips were segmented into 1-second shots, each comprising 30 frames. The training process for all experiments was conducted over a period of approximately 20 hours, utilizing a single Nvidia A5000 GPU (24 GB) and the PyTorch framework version 1.13.1. We employed an early stopping technique to expedite the training process. Training was terminated, and the training checkpoint was saved if the prediction loss on the validation data did not decrease within five epochs. In all the experiments, the model was trained on the gait dataset using the Adam optimizer [43] with a learning rate of 10^{-5}. If the validation loss did not improve over three epochs, the learning rate was halved. A threefold cross-validation procedure was employed for both the training and testing phases to guarantee the fairness and reliability of the distribution of the dataset. The model's performance was evaluated during testing using several metrics, including the accuracy, precision, and F1 score. The evaluation metrics presented in this paper are all averages derived from the three-fold cross-validation.

TABLE 4. Evaluation metrics for our baseline experiments. We undertake a comparison between the proposed method and the baseline methods.

Models	Accuracy	Precision	F1 score
Without PhaseMix			
CNN [44]	52.56%	**81.11%**	54.01%
CNN LSTM [45]	38.83%	45.41%	34.90%
Two stream [47]	54.89%	76.49%	56.46%
STGCN [46]	60.42%	60.22%	59.80%
3D CNN [4], [48]	66.30%	73.26%	68.55%
With PhaseMix			
PhaseMix + CNN	38.08%	50.21%	30.67%
PhaseMix + CNN LSTM	47.66%	52.22%	47.86%
PhaseMix + Two stream	50.41%	46.16%	39.66%
PhaseMix + 3D CNN (Ours)	**71.43%**	72.80%	**71.15%**

D. COMPARATIVE STUDY

To demonstrate the effectiveness of the proposed method, we compared it with current mainstream action recognition methods. A comparative analysis was conducted to evaluate the efficacy of various deep learning architectures, including the normal CNN (ResNet style [44]), CNN LSTM [45], skeleton-based method (STGCN [46]), optical flow-based method (two-stream [47]), and 3D CNN [48]. To enhance the reliability of the results, we combined the proposed PhaseMix fusion methodology with various backbone architectures and also assessed the performance of these backbones without applying the proposed fusion methodology. The results of the comparison are presented in Table 4 and Figure 6. The most notable results have been highlighted in bold text for ease of identification. Figure 6 shows the confusion matrix for our experiments. In this figure, lighter colors indicate a

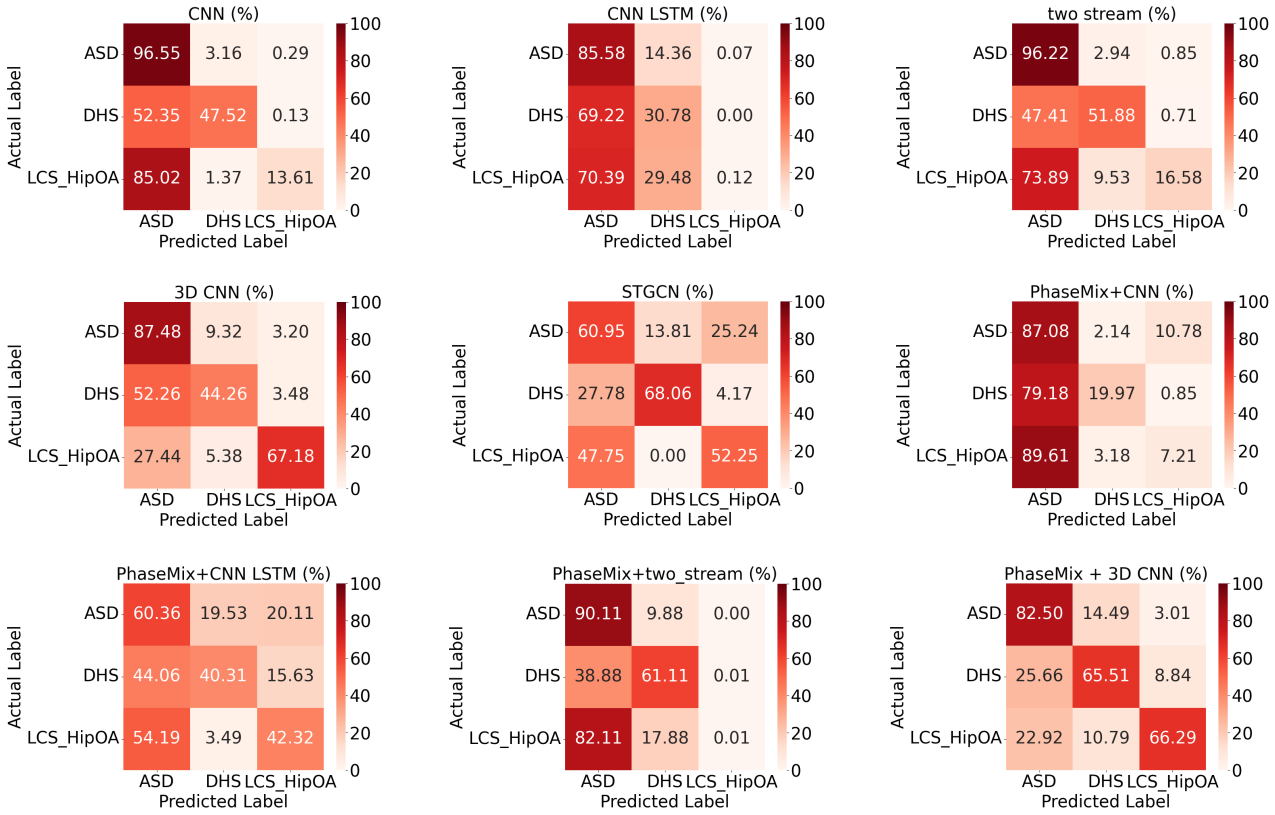


FIGURE 6. Confusion matrix of different models.

higher ratio of video clips among the target clips, whereas darker colors indicate a lower ratio. The rows represent the videos labeled by the doctor (ground truth), and the columns represent the results predicted by the model. For presentation purposes, the numbers in the graph represent the ratios normalized to the targets.

E. ABLATION STUDY

To further examine the effectiveness of *PhaseMix* in the classification stage, we employed the same backbone network (3D CNN) architecture as that of the baseline for the ablation study. Two sets of ablation experiments were performed for demonstration. The first set demonstrates the importance of periodicity and symmetry in the proposed *PhaseMix* method. This was achieved by modifying the periodicity of motion and posture symmetry in the training data. The results of this comparison are presented in Table 5 and Figure 8. The second set of ablation experiments compared the performance of the proposed method under different training strategies. Specifically, the efficacy of methods that utilized only the left swing (first half of the gait cycle), only the right swing (second half of the gait cycle), random selection, *PhaseMix*, and late fusion was evaluated. The results of this comparison are presented in Table 5 and Figure 9.

1) PERIODICITY AND SYMMETRY

The first ablation experiment demonstrated that the periodicity of motion and posture symmetry played important roles in the proposed method. We prepared manual training data by artificially modifying periodicity and symmetry. Figure 7 shows how to break periodic motion and posture symmetry to create data. To demonstrate the relationship between periodic motion and posture symmetry, we illustrated the conversion of the gait cycle to a circumference. Figure 7A illustrates the correspondence between the gait cycle and the circumference. The numerical values displayed in the figure represent the number of frames in the video. The red boxes represent the initial gait cycle of the individual in the video, beginning at frame 0 and concluding at frame 24. The subsequent gait cycle is represented by the blue boxes, beginning at frame 24 and concluding at frame 52. The upper half of the circumference (red line segment) represents the left swing phase, whereas the lower half of the circumference (blue line segment) represents the right swing phase.

As illustrated in Figure 7B, non-periodic data were created by randomly deleting frames (illustrated by the blue area in the figure). The deleted frames had the same proportion but were at different angular positions between the cycles. Thus, the sampled frames for the fusion images can be alternated in either the left or right swing phases, or both. Based on this operation, the periodic motion of the gait cycle can be adjusted manually. The deletion rate λ is calculated as follows:

$$\lambda = \begin{cases} \frac{D_L + D_R}{N_L + N_R} & \text{if non-periodicity} \\ \frac{D_R}{N_R} & \text{if asymmetry} \end{cases} \quad (1)$$

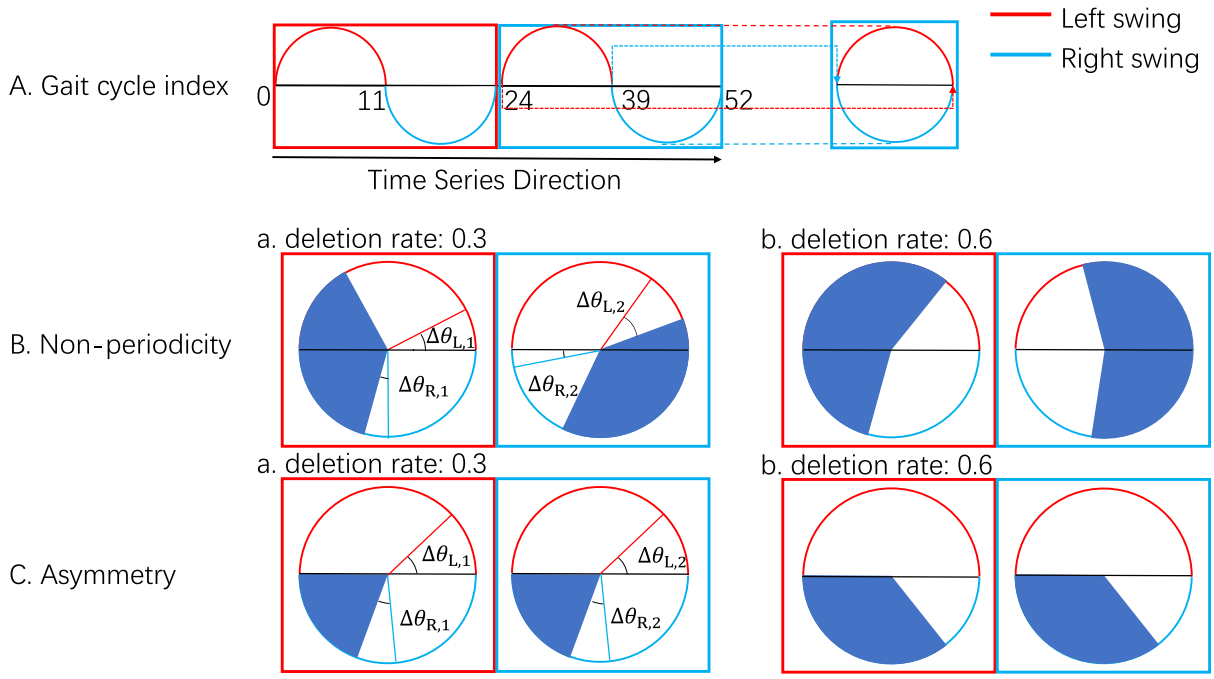


FIGURE 7. Flowchart of the modified periodicity and symmetry approach. (A) The utilization of the circle's circumference to represent the gait cycle. (B) The method for breaking the periodicity of the motion by deleting frames proportional to the deselected position. (C) The method of breaking the posture symmetry. The images processed in (B) and (C) are used as training data for the ablation study.

where N_L and N_R are the number of frames for the original left and right swing phases of a cycle, respectively, D_L and D_R are the number of deleted frames for the left and right swing phases, respectively. The sampling intervals $\Delta\theta_{\mathrm{L}}$ and $\Delta\theta_{\mathrm{R}}$ are calculated as follows:

$$\Delta\theta_{\mathrm{L}} = \frac{N_L - D_L}{N_S} \tag{2}$$

$$\Delta\theta_{\mathrm{R}} = \frac{N_R - D_R}{N_S} \tag{3}$$

where N_S is the number of sampling frames used to create fused images. For example, in Figure 7B.a, the concluding portion of the left swing phase and initial portion of the right swing phase in the first gait cycle were excised. In the second gait cycle, the beginning of the left swing phase and end of the right swing phase were removed. For the fused image, the frames extracted from the processed left and right swing phases were sampled by $\Delta\theta_{\mathrm{L},1}$ and $\Delta\theta_{\mathrm{R},1}$ intervals, respectively. The same methodology was employed for images fused during the second gait cycle.

Consequently, the remaining frames in one cycle and the subsequent cycles had different walking phases. As illustrated in Figure 7C, asymmetric data were created by randomly deleting only the right swing phase frames and setting D_L to 0 in Equation (2). The deleted frames had the same proportions and angular positions between cycles. This resulted in different angular positions between the left and right swing phases, but the same proportions between cycles. The training frame selection strategy was similar to that for non-periodic data.

Table 8 presents the results of disrupting the periodic motion and posture symmetry under varying scaling conditions. For the frame, we selected 0.3 and 0.6 as the scaling conditions. As shown in the table, the accuracy of the model declined significantly as the percentage of corrupted posture symmetry increased. Regarding the motion periodicity of the destruction, the accuracy of the model significantly decreased as the percentage of destruction increased. 0.3 and 0.6 rate frames were randomly selected for deletion for the model. Figure 8 shows the confusion matrix for the first ablation study. To conduct the experiments on motion periodicity disruption, we randomly deleted 0.3 scale and 0.6 scale frames to disrupt motion periodicity. The deletion of the 0.3 scale frames resulted in a decline in the model's ability to classify DHS, whereas its ability to classify LCS_HipOA increased. Upon deletion of the 0.6 scale frames, the model exhibited a decline in its ability to classify LCS_HipOA. With regard to posture symmetry, the model demonstrated an accuracy of 66.32 in classifying ASD after deleting 0.3 frames in the right swing phase (Figure 8 asymmetric 0.3), whereas the accuracy was originally 82.50 without breaking posture symmetry (*PhaseMix*). When the corrupted values reached 0.6 (Figure 8 asymmetric 0.6), we observed a significant decrease in the model's ability to classify LCS_HipOA, from 66.29 to 42.10.

2) DIFFERENT TRAINING STRATEGIES

In the second ablation experiment, we demonstrated the performance of the proposed method under different training strategies. Table 5 illustrates that the model's performance

TABLE 5. Evaluation metrics for our ablation study. The first ablation study shows that the periodicity of motion and posture symmetry play an important role in the proposed method. The second ablation study compares the proposed method with alternative training strategies using the same backbone.

Models	Accuracy	Precision	F1 score
First ablation study			
Non-periodicity $\lambda = 0.3$	68.29%	70.36%	67.97%
Non-periodicity $\lambda = 0.6$	68.75%	72.04%	67.92%
Asymmetric $\lambda = 0.3$	70.28%	71.02%	69.95%
Asymmetric $\lambda = 0.6$	67.06%	69.76%	64.94%
Second ablation study			
Left swing	58.29%	**73.38%**	61.12%
Right swing	63.02%	73.27%	66.00%
Random (3D CNN)	66.30%	73.26%	68.55%
Late fusion	61.16%	70.30%	60.90%
Proposed method			
PhaseMix + 3D CNN	**71.43%**	72.80%	**71.15%**

was enhanced when the features of both the left and right swing phases were combined, exceeding the outcomes of training alone, random selection, and late fusion training. The term ''random'' denotes a training strategy where the model is not trained with consideration of the walk period. Instead, a specified number of frames is extracted from one second in a predetermined order, which is a common approach employed in contemporary general character recognition methods. The term ''late fusion'' indicates that the walking cycle is initially divided into the left and right swing phases. Subsequently, the two networks are trained using the information from the left and right swings, respectively. Finally, the features are fused at the final feature level of the network. In the context of training based on the left or right swing phase, a gait-defined module was employed to divide the walking cycles in the video. Subsequently, only information pertaining to the left or right swing phase was utilized for training in accordance with the walking index. The *PhaseMix* method employs our pixel-level fusion technique, which combines the features of the left and right swing phases. This approach allows for the learning of features for the entire gait cycle, including periodicity in the gait cycle and symmetry in the gait action, within a single training step while reducing the loss of key features.

The results demonstrate that the proposed method has superior performance compared to other methods. Figure 9 shows the confusion matrix for the ablation study. The left swing phase is an effective classifier for ASD; however, two other diseases (DHS and LCS_HipOA) were misclassified as ASD. This misclassification likely occurred because the left swing phase lacks the features necessary to accurately classify DHS and LCS_HipOA, leading to this behavior of the classifier. Compared to the left swing phase, the right swing phase showed decreased ability to classify ASD but improved ability to classify DHS and LCS_HipOA. It is hypothesized that the action features incorporated into the right swing phase provide more useful information for classification. The categories classified using the random method exhibited greater uniformity. However, the exclusion of the walking cycle during the learning process results in the loss of crucial information, which in turn reduces the classification ability for specific cases (LCS_HipOA). The late fusion method demonstrated satisfactory classification ability for ASD and some ability for DHS. However, it performed poorly in specific cases (LCS_HipOA). We hypothesized that this was due to the inability of the network to effectively localize disease-specific feature information during the pre-extraction phase. For instance, the feature changes of the pelvis are more pronounced in LCS and HipOA; however, the final post-fusion classification result is poor owing to the lack of relevant information. In contrast, the phase mix approach accounts for the periodicity of motion and symmetry of body movements during the gait cycle, integrating and blending different phase characteristics for a more effective learning of disease-specific features.

V. DISCUSSION

This paper presents an image fusion method for the remote diagnosis of ASD and other spinal gait diseases using video analysis. The proposed method integrates the periodicity and symmetry of walking motion through a fusion approach that considers periodic motion features. In the proposed method, the gait cycle was initially analyzed to more precisely determine the specific phases of the cycle. Subsequently, a fusion method that incorporated gait cycle features was employed to enhance the diagnosis of various diseases. The proposed method is open-source and is available at https://github.com/ChenKaiXuSan/Skeleton_ASD_PyTorch. Additionally, a structured experimental code was provided to facilitate the replication of the experiments.

A. COMPARATIVE STUDY

We propose a pixel-level fusion approach that considers the periodicity of motions and the inherent symmetry of posture. The study commenced with a comparison experiment in which the proposed method was compared with several prevalent action recognition techniques. Table 4 provides a comprehensive comparison of the various methods.

As indicated in the table, the CNN method achieved a better precision score. We believe that this can be attributed to the CNN model's capacity to classify ASD with particular efficacy, although it does not necessarily extend to other classes. Figure 6 shows the confusion matrix of the various methods. The accuracy of the set of tables and confusion matrix demonstrates that the CNN method utilizing static images is less effective in classification, with a greater propensity for misclassifying DHS and LCS_HipOA as ASD. It has been hypothesized that this phenomenon can be attributed to the absence of temporal information, which may be a contributing factor. Contrary to expectations, the performance of the PhaseMix method combined with CNN did not show significant improvement. The goal of PhaseMix is to fuse periodic motion and action symmetry features while emphasizing their temporal evolution. However, CNNs, even those with deep architectures, are not inherently designed

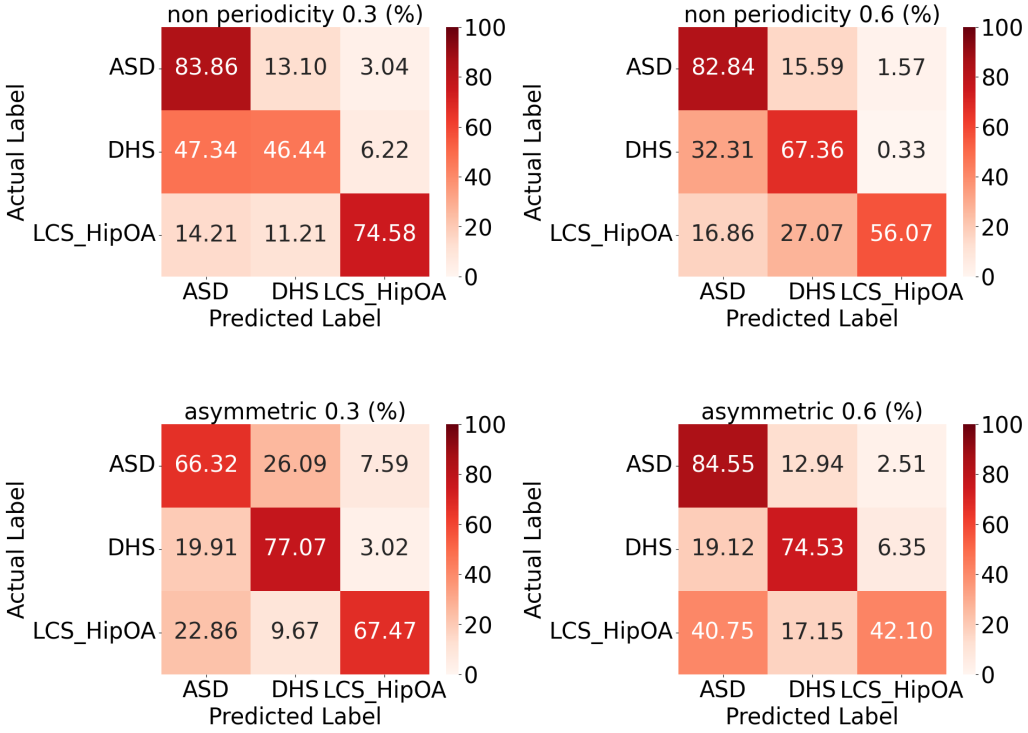


FIGURE 8. Confusion matrix for the second ablation study. The representative confusion matrix with Non-periodic 0.3 and 0.6, and Asymmetry 0.3 and 0.6.

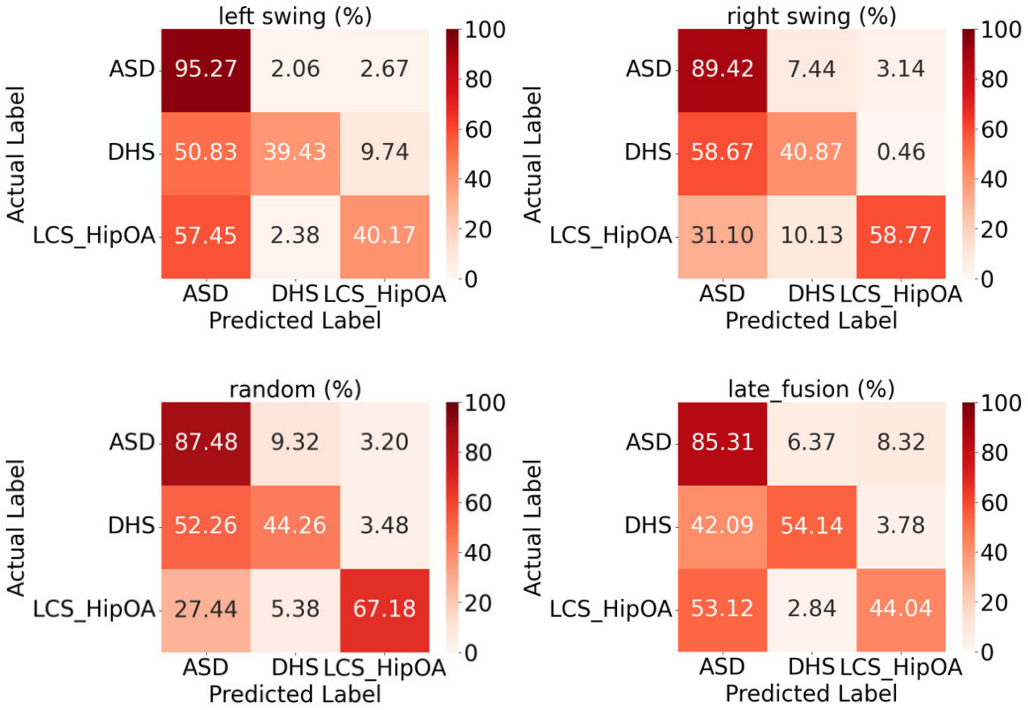


FIGURE 9. Confusion matrix for the second ablation study, with different training strategies.

to capture temporal dynamics. As a result, incorporating PhaseMix features may introduce redundancy, potentially diluting essential information and negatively impacting overall performance.

The CNN LSTM method demonstrated the least promising performance. It is hypothesized that the incorporation of temporal information into static image-based methods is an ineffective approach that confounds the features that can be

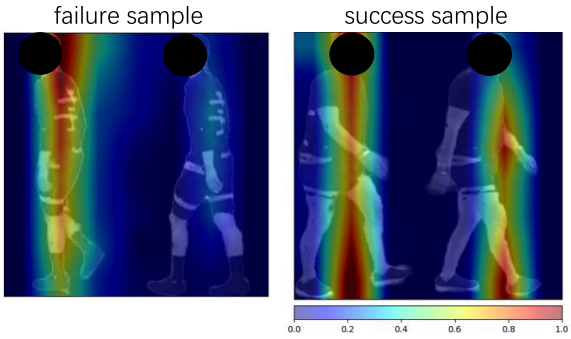


FIGURE 10. The class attention map is used for the proposed method. In the resulting visualizations, red areas represent regions that provide more positive information about the final result, while blue areas represent regions that provide less positive information. The patient's head was mosaiced.

learned from static images. The combination of PhaseMix with CNN LSTM resulted in improved model performance because the PhaseMix module effectively extracts periodic and symmetry-related features that are essential for understanding complex motion patterns. When integrated with CNN LSTM, these features are not only spatially encoded by the CNN but also sequentially processed by the LSTM. This dual representation enhances the model's ability to capture subtle variations in gait patterns, which is crucial for accurate diagnosis.

The two-stream method demonstrated superior performance to both the CNN and CNN LSTM methods, yet it exhibited suboptimal results for a select number of samples in the DHS and LCS_HipOA datasets. It is believed that the optical flow-based method requires a substantial number of reference samples in the context of special evidence extraction, which is rendered ineffective in the case of sparse samples (LCS_HipOA). The performance decline when combining the PhaseMix method with two-stream architecture is likely due to the limitations of the optical flow stream. Optical flow captures motion between consecutive frames by analyzing pixel displacements, but it struggles to effectively utilize the complex features generated by PhaseMix, which already encapsulate periodic motion and action symmetry. While optical flow focuses on short-term motion dynamics, PhaseMix features provide a more structured and comprehensive representation of motion patterns. This mismatch prevents the optical flow stream from fully leveraging the enriched information provided by PhaseMix, resulting in suboptimal performance.

The 3D CNN methods are classifiers that do not utilize the proposed method. The results demonstrated that the model's capacity for classifying ASD and LCS_HipOA was satisfactory, whereas the ability to classify DHS was inadequate. It is postulated that this is attributable to the loss of learned features owing to the periodicity of walking and the motion symmetry, which were not considered in the classification. The combination of PhaseMix and 3D CNN achieved the best performance due to several key advantages. 3D CNNs are designed to process spatiotemporal data, capturing both spatial and temporal information within video frames. When integrated with PhaseMix, which emphasizes periodic motion and action symmetry, the 3D CNN can effectively leverage these enriched features. This results in a more detailed representation of gait dynamics, enhancing the model's ability to distinguish subtle motion patterns. Additionally, PhaseMix provides a structured fusion of motion features, boosting the discriminative power of the input data. This synergy maximizes the impact of both spatial and temporal information, leading to a more accurate understanding of underlying motion dynamics.

The application of skeletal-based methods during classifier training results in a loss of accuracy because of the inability to guarantee the accuracy of the skeletal data extracted from the video. Furthermore, the conversion of skeletal data from videos is a time-consuming process. Our method obtains a relatively balanced classification result that considers the periodicity and symmetry of movements during walking. Additionally, it achieves good classification performance for sparse samples.

B. ABLATION STUDY

Two distinct ablation experiments are presented in different sections of the proposed method. The first ablation experiment illustrated the pivotal role of the symmetry of movement and the periodicity of walking in the proposed method. The second ablation experiment demonstrated the efficacy of the proposed method when applied to diverse training strategies.

The first experiment examined the impact of the symmetry of movement and the periodicity of walking on the proposed method. As illustrated in Table 5 and Figure 8, the evaluation index of the model declined when the symmetry of the action was manually broken. Furthermore, the classification capacity of the model deteriorated as the proportion of symmetry breaking increased. This evidence suggests that the symmetry information of the action is important in the proposed method. Furthermore, the results of the action-periodic destruction experiments indicated a notable decline in the model's performance when different proportions of destruction were implemented. Despite the stability of the model evaluation metrics across varying proportions of destruction, we hypothesized that the proportion of destruction may not exert a significant influence because of the randomly selected locations for cyclic destruction.

Subsequently, the second ablation experiment was conducted. The efficacy of the proposed pixel-fusion method was evaluated by analyzing the performance of the classifiers. A comparison between the phase-mixing method and other approaches, including left swing, right swing, random selection, and late feature fusion, is presented. Table 5 lists the evaluation metrics of the classifiers. We can observe that the proposed phase-mixing method outperforms the other methods. The application of either the left swing phase or the right swing phase for training the classifier resulted in inferior outcomes, possibly because it fails to encompass the information of the entire walk cycle. In particular, the

using the left swing phase alone yielded the least favorable results. The movements occurring during the stance phase presented a significant challenge for the classifier in terms of classification. The random selection method yielded satisfactory results. However, it has the potential to disrupt the continuity of actions because it does not consider the walking cycle. In other words, the sample selection was performed without considering the properties of periodic motion, periodicity of motion, and symmetry of action. Maintaining coherence in actions (temporal sequentiality) is of paramount importance in time series-based classifiers. The performance of the late-fusion method is intermediate between those of the left and right swing methods and is inferior to that of the random-selection method. It appears that although periodic motion is preprocessed in the learning samples, the periodicity and symmetry information of the action in the periodic motion is still lost because late-fusion learning does not share the network parameters during the learning process. The proposed method yielded the most optimal results, and the rationale behind this outcome is as follows:

- Capture movement symmetry. Merging left and right movements enables the model to learn patterns of symmetry. For example, the motion of the left and right feet during walking is symmetrical. By capturing this symmetry, the model can understand the consistency and balance of the movement.
- Emphasize mutual information. The fusion of left and right movements into a single image serves to accentuate the relationship between them. This makes it easier to identify features that are difficult to see from a single swing phase. For example, the position and movement of the right foot when the left foot is in a forward phase provides important information about the overall gait pattern.
- Improved abnormality detection. Comparing left and right movements makes it easier to detect abnormalities. Since the left and right sides are often symmetrical in normal movement, abnormalities on one side are easier to detect. This enables the early detection of disease and injury.
- Data enrichment. By fusing left and right movements, the amount of information contained in a single image increases. This enriches the data that the CNN learns on and improves the performance of the model. In particular, the overlap of left and right motions at different times makes it easier to capture changes in motion over time.

To further analyze the model's classification ability, we also present the results of the confusion matrix. Figure 9 shows the prediction results for different experimental methods. As illustrated in the figure, the proposed method demonstrated superior performance in classifying rare diseases (LCS_HipOA) compared with the comparison method. The results for the second type of disease (DHS label in our dataset) were not as good as those of the other methods. We believe that this is because the features of DHS and ASD are very similar in the fusion, which results in a high misclassification rate. However, the high discriminative ability for rare diseases suggests that the model performs better at classifying periodic movements, considering the periodicity and symmetry of the movements.

C. MODEL VISUALIZATION

In this study, to conduct a comprehensive analysis of the model's internals, we used Grad CAM++ [49] for model visualization. This technique enables the identification of specific regions of the model that should be focused on during training. In addition, identifying the specific parts of a patient's body involved in walking provides more informative insights that can improve the diagnostic process. Visualizations were generated using the implementation provided by [50], which is available in the PyTorch repository, to ensure robust and reproducible results.

Figure 10 illustrates this process for the proposed method. For successful predictions, the model focused more on the patient's body parts, whereas for failures, it tended to focus on irrelevant parts (areas other than the patient). In successful cases, the model emphasized the positions of the head and feet. We believe that this is because head and foot movements are the most pronounced during walking; thus, the model learns more information from these body parts. In cases of failure, the final prediction was incorrect because the model did not adequately focus on the patient's body. Instead, the model often focused on the background and ignored the features of the patient. This focus misalignment resulted in prediction errors. Although the model focuses more on body parts for success, it also requires periodicity of movement and symmetry of motion. Furthermore, the proposed method indicates that the model pays more attention to the effects of the distinct cycles in periodic motion (left and right swings) owing to the enhancement of the effective information in the image and consideration of the symmetry of the motion. It is our contention that this is a crucial consideration when the model makes predictions.

D. LIMITATIONS

While our method demonstrates utility in diagnosing ASDs, several limitations must be acknowledged.

A key limitation is the relatively small dataset used for model training and evaluation. While the results provide a strong foundation, the limited dataset size may hinder the generalizability of our findings to broader populations. Our model demonstrated limited classification capacity for some classes, such as DHS and LCS_HipOA. This is likely due to analogous movements at the video level and the small dataset size. Addressing this will require a larger and more varied dataset to improve classification accuracy. Future research should focus on expanding the dataset through collaborations with clinical institutions to obtain more extensive and diverse data. We also aim to include videos of normal human gait for comparison, as the current dataset is composed mainly

of clinical patients. This will enable a more comprehensive evaluation of the model's generalizability. Dataset imbalance also poses challenges for accurate categorization. Future work should explore improved methods for achieving a more balanced dataset. In the short term, data augmentation techniques can be employed to simulate larger datasets and enhance the model's robustness. Extraneous factors, such as camera angle, lighting conditions, and patient attire, may also influence model performance. Standardizing video capture protocols could mitigate these effects. In addition, the resolution and quality of the videos used may limit the model's ability to detect subtle gait differences. Higher-resolution videos could provide more detailed information and improve the model's discriminatory power. Finally, as the dataset size grows, ensuring the model's computational efficiency and scalability without compromising performance will be crucial for broader clinical applications.

Another limitation is the lack of a comprehensive analysis of the method's computational scalability, particularly with larger datasets. While this study demonstrates the feasibility of our method on a moderate dataset, future work should investigate scalability by optimizing computational efficiency and applying the approach to larger, more diverse datasets. This will help to better understand the method's practical utility in real-world diagnostic scenarios.

VI. CONCLUSION

In conclusion, we propose a periodic movement fusion method for diagnosing adult spinal deformities using a video format. The proposed method is a pixel-level, plug-and-play approach that requires no changes to the network structure. We propose a framework to first identify the indices in periodic motion from the video, followed by fusion at the pixel level, while considering the periodicity of the periodic motion and the symmetry of the motion. The proposed method, *PhaseMix*, exhibited superior performance compared to the other methods. Moreover, our findings suggest that in the context of video-based disease diagnosis, the use of a neural network is essential to account for the periodicity of periodic motion and the symmetry of the motion.

As noted in Section V-D, our method, while promising for diagnosing ASDs, has several limitations. The small dataset may limit generalizability, particularly for classes with similar movements. Dataset imbalance and factors like camera angle, lighting, and video quality may also affect performance, underscoring the need for standardized video capture and higher-resolution footage. Data augmentation can help address these issues in the short term. Finally, as the dataset grows, ensuring the model's scalability and computational efficiency will be crucial for clinical applications.

Future studies will aim to apply the PhaseMix method across a wider population to assess its robustness and potential for broader clinical applications. While the proposed method has demonstrated effectiveness in diagnosing ASD, further investigation is needed to evaluate its performance on non-ASD cases and healthy subjects. Moreover, it would be beneficial to use more accurate and specific joint detection methods to enhance the predictive power of the model. Additionally, it is necessary to address the issue of sparse data to achieve a more balanced dataset, particularly when increasing the size of the dataset is not feasible. Furthermore, it is believed that neural networks are useful for long-term training. Therefore, we plan to use a longer video time to train the network, with the expectation of obtaining better results.

REFERENCES

[1] H. J. Kim, J. H. Yang, D.-G. Chang, S.-I. Suk, S. W. Suh, K.-S. Song, J.-B. Park, and W. Cho, "Adult spinal deformity: Current concepts and decision-making strategies for management," *Asian Spine J.*, vol. 14, no. 6, pp. 886–897, Dec. 2020.

[2] B. G. Diebo, N. V. Shah, O. Boachie-Adjei, F. Zhu, D. A. Rothenfluh, C. B. Paulino, F. J. Schwab, and V. Lafage, "Adult spinal deformity," *Lancet*, vol. 394, no. 10193, pp. 160–172, 2019.

[3] J. Youssef, D. Orndorff, C. Patty, M. Scott, H. Price, L. Hamlin, T. Williams, J. Uribe, and V. Deviren, "Current status of adult spinal deformity," *Global Spine J.*, vol. 3, no. 1, pp. 51–62, 2013.

[4] K. Chen, T. Asada, N. Ienaga, K. Miura, K. Sakashita, T. Sunami, H. Kadone, M. Yamazaki, and Y. Kuroda, "Two-stage video-based convolutional neural networks for adult spinal deformity classification," *Frontiers Neurosci.*, vol. 17, Dec. 2023, Art. no. 1278584.

[5] T. Asada, K. Miura, M. Koda, H. Kadone, T. Funayama, H. Takahashi, H. Noguchi, Y. Shibao, K. Sato, and F. Eto, "Can proximal junctional kyphosis after surgery for adult spinal deformity be predicted by preoperative dynamic sagittal alignment change with 3D gait analysis? A case–control study," *J. Clin. Med.*, vol. 11, no. 19, p. 5871, 2022.

[6] T. Asada, K. Miura, H. Kadone, K. Sakashita, T. Funayama, H. Takahashi, H. Noguchi, K. Sato, F. Eto, H. Gamada, K. Inomata, M. Koda, and M. Yamazaki, "The relationship between spinal alignment and activity of paravertebral muscle during gait in patients with adult spinal deformity: A retrospective study," *BMC Musculoskeletal Disorders*, vol. 24, no. 1, pp. 1–10, Jan. 2023.

[7] K. Miura, H. Kadone, M. Koda, K. Nakayama, H. Kumagai, K. Nagashima, K. Mataki, K. Fujii, H. Noguchi, T. Funayama, T. Abe, K. Suzuki, and M. Yamazaki, "Visualization of walking speed variation-induced synchronized dynamic changes in lower limb joint angles and activity of trunk and lower limb muscles with a newly developed gait analysis system," *J. Orthopaedic Surgery*, vol. 26, no. 3, Sep. 2018, Art. no. 2309499018806688.

[8] K. Miura, H. Kadone, M. Koda, T. Abe, T. Funayama, H. Noguchi, K. Mataki, K. Nagashima, H. Kumagai, Y. Shibao, K. Suzuki, and M. Yamazaki, "Thoracic kyphosis and pelvic anteversion in patients with adult spinal deformity increase while walking: Analyses of dynamic alignment change using a three-dimensional gait motion analysis system," *Eur. Spine J.*, vol. 29, no. 4, pp. 840–848, Apr. 2020.

[9] J. Crosbie, R. Vachalathiti, and R. Smith, "Patterns of spinal motion during walking," *Gait Posture*, vol. 5, no. 1, pp. 6–12, Feb. 1997.

[10] M. W. Whittle and D. Levine, "Three-dimensional relationships between the movements of the pelvis and lumbar spine during normal gait," *Hum. Movement Sci.*, vol. 18, no. 5, pp. 681–692, Oct. 1999.

[11] R. Haddas, K. L. Ju, T. Belanger, and I. H. Lieberman, "The use of gait analysis in the assessment of patients afflicted with spinal disorders," *Eur. Spine J.*, vol. 27, no. 8, pp. 1712–1723, Aug. 2018.

[12] S. D. Glassman, S. Berven, K. Bridwell, W. Horton, and J. R. Dimar, "Correlation of radiographic parameters and clinical symptoms in adult scoliosis," *Spine*, vol. 30, no. 6, pp. 682–688, Mar. 2005.

[13] K. Miura, M. Koda, H. Kadone, T. Abe, H. Kumagai, K. Nagashima, K. Mataki, K. Fujii, H. Noguchi, T. Funayama, K. Suzuki, and M. Yamazaki, "Successful detection of postoperative improvement of dynamic sagittal balance with a newly developed three-dimensional gait motion analysis system in a patient with iatrogenic flatback syndrome: A case report," *J. Clin. Neurosci.*, vol. 53, pp. 241–243, Jul. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967586818303205

[14] A. A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," *Proc. SPIE*, vol. 5779, pp. 196–204, Sep. 2005.
[15] A. Nagar, K. Nandakumar, and A. K. Jain, "Multibiometric cryptosystems based on feature-level fusion," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 1, pp. 255–268, Feb. 2012.
[16] H. Cai, Z. Qu, Z. Li, Y. Zhang, X. Hu, and B. Hu, "Feature-level fusion approaches based on multimodal EEG data for depression recognition," *Inf. Fusion*, vol. 59, pp. 127–138, Jul. 2020.
[17] M. Ehatisham-Ul-Haq, A. Javed, M. A. Azam, H. M. A. Malik, A. Irtaza, I. H. Lee, and M. T. Mahmood, "Robust human activity recognition using multimodal feature-level fusion," *IEEE Access*, vol. 7, pp. 60736–60751, 2019.
[18] C. Chen, R. Jafari, and N. Kehtarnavaz, "Improving human action recognition using fusion of depth camera and inertial sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 1, pp. 51–61, Feb. 2015.
[19] V. Naidu and J. Raol, "Pixel-level image fusion using wavelets and principal component analysis," *Defence Sci. J.*, vol. 58, no. 3, pp. 338–352, May 2008.
[20] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Inf. Fusion*, vol. 13, no. 1, pp. 10–19, Jan. 2012.
[21] J. Chen, L. Chen, and M. Shabaz, "Image fusion algorithm at pixel level based on edge detection," *J. Healthcare Eng.*, vol. 2021, pp. 1–10, Aug. 2021.
[22] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.
[23] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.
[24] E.-H. Kim, Z. Wang, H. Zong, Z. Jiang, Z. Fu, and W. Pedrycz, "Design of tobacco leaves classifier through fuzzy clustering-based neural networks with multiple histogram analyses of images," *IEEE Trans. Ind. Informat.*, vol. 20, no. 3, pp. 4698–4709, Mar. 2024.
[25] Z. Fu, Y. Lin, D. Yang, and S. Yang, "Fractional Fourier transforms meet Riesz potentials and image processing," *SIAM J. Imag. Sci.*, vol. 17, no. 1, pp. 476–500, Mar. 2024.
[26] Z. Wang, E.-H. Kim, S.-K. Oh, W. Pedrycz, Z. Fu, and J. H. Yoon, "Reinforced fuzzy-rule-based neural networks realized through streamlined feature selection strategy and fuzzy clustering with distance variation," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 10, pp. 5674–5686, Oct. 2024.
[27] S. D. Glassman, K. Bridwell, J. R. Dimar, W. Horton, S. Berven, and F. Schwab, "The impact of positive sagittal balance in adult spinal deformity," *Spine*, vol. 30, no. 18, pp. 2024–2029, Sep. 2005.
[28] V. Lafage, F. Schwab, A. Patel, N. Hawkinson, and J.-P. Farcy, "Pelvic tilt and truncal inclination: Two key radiographic parameters in the setting of adults with spinal deformity," *Spine*, vol. 34, no. 17, pp. E599–E606, Aug. 2009.
[29] F. Schwab, A. Patel, B. Ungar, J.-P. Farcy, and V. Lafage, "Adult spinal deformity—Postoperative standing imbalance: How much can you tolerate? An overview of key parameters in assessing alignment and planning corrective surgery," *Spine*, vol. 35, no. 25, pp. 2224–2231, Dec. 2010.
[30] J. Cheung, D. J. Wever, A. G. Veldhuizen, J. P. Klein, B. Verdonck, R. Nijlunsing, J. C. Cool, and J. R. Van Horn, "The reliability of quantitative analysis on digital images of the scoliotic spine," *Eur. Spine J.*, vol. 11, no. 6, pp. 535–542, Dec. 2002.
[31] T. Vrtovec, B. Likar, and F. Pernuš, "Quantitative analysis of spinal curvature in 3D: Application to CT images of normal spine," *Phys. Med. Biol.*, vol. 53, no. 7, pp. 1895–1908, Apr. 2008.
[32] P. Severijns, L. Moke, T. Overbergh, E. Beaucage-Gauvreau, T. Ackermans, K. Desloovere, and L. Scheys, "Dynamic sagittal alignment and compensation strategies in adult spinal deformity during walking," *Spine J.*, vol. 21, no. 7, pp. 1059–1071, Jul. 2021.
[33] T. Banno, Y. Yamato, O. Nojima, T. Hasegawa, G. Yoshida, H. Arima, S. Oe, H. Ushirozako, T. Yamada, K. Ide, Y. Watanabe, K. Yamauchi, and Y. Matsuyama, "Comparison of the postoperative changes in trunk and lower extremity muscle activities between patients with adult spinal deformity and age-matched controls using surface electromyography," *Spine Deformity*, vol. 10, no. 1, pp. 141–149, Jan. 2022.
[34] J. S. Kim, V. Arvind, E. K. Oermann, D. Kaji, W. Ranson, C. Ukogu, A. K. Hussain, J. Caridi, and S. K. Cho, "Predicting surgical complications in patients undergoing elective adult spinal deformity procedures using machine learning," *Spine Deformity*, vol. 6, no. 6, pp. 762–770, Nov. 2018.
[35] W. M. Durand, J. M. DePasse, and A. H. Daniels, "Predictive modeling for blood transfusion after adult spinal deformity surgery: A tree-based machine learning approach," *Spine*, vol. 43, no. 15, pp. 1058–1066, 2018.
[36] M. W. Whittle, "Clinical gait analysis: A review," *Hum. Movement Sci.*, vol. 15, no. 3, pp. 369–387, Jun. 1996.
[37] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, pp. 2255–2283, Feb. 2012.
[38] S. Winiarski, A. Rutkowska-Kucharska, and M. Kowal, "Symmetry function—An effective tool for evaluating the gait symmetry of transfemoral amputees," *Gait Posture*, vol. 90, pp. 9–15, Oct. 2021.
[39] A. Rohan, M. Rabah, T. Hosny, and S.-H. Kim, "Human pose estimation-based real-time gait analysis using convolutional neural network," *IEEE Access*, vol. 8, pp. 191542–191550, 2020.
[40] S. Z. A. Rahman, S. N. H. S. Abdullah, and M. Z. B. A. Nazri, "The analysis for gait energy image based on statistical methods," in *Proc. Int. Conf. Adv. Electr., Electron. Syst. Eng. (ICAEES)*, Nov. 2016, pp. 125–128.
[41] G. Jocher, A. Chaurasia, and J. Qiu. (Jan. 2023). *Ultralytics YOLO*. [Online]. Available: https://github.com/ultralytics/ultralytics
[42] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, and A. Zisserman, "The kinetics human action video dataset," 2017, *arXiv:1705.06950*.
[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[45] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, Feb. 2020, pp. 362–366.
[46] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 7444–7452.
[47] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 568–576.
[48] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.
[49] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N Balasubramanian, "Grad-CAM++: Improved visual explanations for deep convolutional networks," 2017, *arXiv:1710.11063*.
[50] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

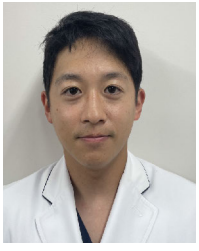
KAIXU CHEN (Student Member, IEEE) received the M.S. degree from Kanazawa University, Kanazawa, Japan, in 2022. He is currently pursuing the Ph.D. degree in engineering with the University of Tsukuba, Tsukuba, Japan. His current research interests include computer vision and human action recognition in the medical field.

JIAYI XU (Member, IEEE) received the Ph.D. degree in engineering from the University of Tsukuba, Tsukuba, Japan, in 2023. Since 2024, she has been a Project Researcher with the Research Center for Advanced Science and Technology (RCAST), The University of Tokyo. Her research interests include haptics, virtual reality, human–computer interactions, and robotics.

TOMOYUKI ASADA received the M.D. degree and the Ph.D. degree in medicine from the University of Tsukuba, Tsukuba, Japan, in 2013 and 2023, respectively. He has become a Board-Certified Spine Surgeon and was approved by the Japanese Society for Spine Surgery and Related Research, in 2022. Since 2022, he has been an Endowed Research Fellow with the Hospital for Special Surgery, New York, NY, USA, where he focuses on translational research related to spinal disorders and surgery. His research also encompasses gait analysis in spinal disorders, a field that was a significant part of his Ph.D. studies.

KOUSEI MIURA received the M.D. degree from the University of Tsukuba, Tsukuba, Japan, in 2008, and the Ph.D. degree in medicine from the Graduate School of Medicine, University of Tsukuba, in 2019. He became a Board-Certified Spine Surgeon and was approved by the Board of the Japanese Society for Spine Surgery and Related Research, in 2020. He has been an Assistant Professor with the Department of Orthopaedic Surgery, Institute of Medicine, University of Tsukuba, since 2021. His research interest includes gait analysis in spinal disorders.

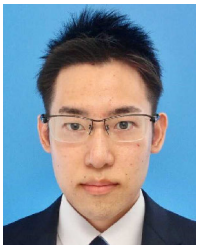
KOTARO SAKASHITA received the M.D. degree from the University of Tsukuba, Tsukuba, Japan, in 2015. He is currently pursuing the degree with the University of Tsukuba, focusing on gait analysis in spinal disorders. He has been a member of the Japanese Society for Spine Surgery and Related Research, since 2020. He received the certification as an orthopedic surgeon from the Japanese Orthopaedic Association, in 2021.



TAKAHIRO SUNAMI received the M.D. degree from the School of Medicine, Hamamatsu University, Hamamatsu, Japan, in 2015. He is currently pursuing the degree with the University of Tsukuba, researching gait analysis in spinal disorders. Since 2020, he has been a member of the Japanese Society for Spine Surgery and Related Research. He received the certification as an orthopedic surgeon from the Japanese Orthopaedic Association, in 2021.

HIDEKI KADONE (Member, IEEE) received the Ph.D. degree in information science and technology from The University of Tokyo, Tokyo, Japan, in 2008. He is currently an Associate Professor with the Institute of Medicine, and Center for Cybernics Research, University of Tsukuba, Tsukuba, Japan. His research interests include clinical motion measurement and analysis and the development of wearable assistive devices to compensate for support and improve the motor function of patients after neurological or mechanical disorders.

MASASHI YAMAZAKI received the M.D. degree from the School of Medicine, Chiba University, Japan, in 1983, and the Ph.D. degree in medicine from the Graduate School of Medicine, Chiba University, Japan, in 1990. Since 2012, he has been a Professor and the Chairperson of the Department of Orthopaedic Surgery, Institute of Medicine, University of Tsukuba. His research interests include medical-engineering collaboration in the field of spine surgery. He was the President of the Japanese Orthopaedic Association, from 2017 to 2019, and the Managing Director of the Japanese Society for Spine Surgery and Related Research, from 2014 to 2018.



NAOTO IENAGA (Member, IEEE) received the Ph.D. degree in engineering from Keio University, Yokohama, Japan, in 2020. He has been an Assistant Professor with the Institute of Systems and Information Engineering, University of Tsukuba, since 2021. He is currently working on the application of machine learning and computer vision to solve practical problems, particularly in fisheries and occupational therapy.

YOSHIHIRO KURODA (Member, IEEE) received the Ph.D. degree in informatics from Kyoto University, Kyoto, Japan, in 2005. He was an Assistant Professor with the Graduate School of Engineering Science, Osaka University, from 2006 to 2013, an Associate Professor with the Cybermedia Center, Osaka University, from 2013 to 2016, and an Associate Professor with the Graduate School of Engineering Science, Osaka University, from 2016 to 2019. Since 2019, he has been a Professor with the Institute of Systems and Information Engineering, University of Tsukuba. His research interests include haptic interaction technologies and biomedical engineering.