Title	Gait Analysis for Healthcare using Small Datasets
Author(s)	廖, 若辰
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/88142
rights	
Note	

Gait Analysis for Healthcare using Small Datasets

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2022

Ruochen LIAO

For my mother.

Abstract

Medical and health management has been highly demanded in this ageing society. To respond to this demand, the medical/healthcare domain makes extensive use of artificial intelligence (AI) technologies to mitigate the shortage on medical/healthcare human resources. Currently, most of the medical/healthcare AI technologies use deep learning approaches and one of such typical technologies is medical image analysis, which is used for the diagnosis of various diseases.

Besides, image-based gait analysis is an important application direction of medical and healthcare AI technologies. Since there are diseases that cause gait symptoms, medical doctors often capture videos of patients' gait for analysis and diagnosis, indicating that these gait videos contain useful information to analyze the patient's condition. Furthermore, as one of the most common motions, many physical conditions are reflected on gait. The information of medical or health can be obtained by analyzing gait; however, it is difficult to achieve such analysis with manual visual observation, and therefore requires the use of AI technology.

Since deep learning is a data-driven approach, its application requires a large amount of training data. Most of the recent studies on gait analysis using deep learning methods rely on existing large-scale gait databases, but these databases lack annotations relevant to the medical/healthcare domain. On the other hand, the datasets in medical/healthcare domain studies are hardly large enough to perform deep learning methods due to various limitations. This makes studies in the medical/healthcare domain must endure small datasets, and therefore it is critical that how to obtain results by analyzing small datasets.

In this paper, the author proposed methods of gait analysis for medical/healthcare applications under small-scale gait databases which make the most of domain-specific knowledge related to target tasks. More specifically, the author proposed the following two strategies for gait analysis through small-scale datasets for different cases of dataset scale and relevance of domain-specific knowledge to the task. In cases where the dataset scale is too small to support any machine learning methods, it is necessary to find direct and exhaustive domain-specific knowledge and manually design analyzable features according to that knowledge. On the other hand, for cases where the dataset scale is small yet still sufficient for fine-tuning the

deep learning network, the authors select primitive information that is relevant to the task and that can be extracted from existing large-scale databases as domain-specific knowledge and utilized it for pre-training primitive networks. The author conducted the following two studies to demonstrate the effectiveness of these strategies.

First, the author proposed a method to support diagnosing idiopathic normal pressure hydrocephalus (iNPH) by video-based gait analysis. Gait is an important factor in the diagnosis of idiopathic normal pressure hydrocephalus. However, except for walking speed tests, existing diagnosis methods only assess gait qualitatively (i.e., manual observation by medical doctors). This study proposed a quantitative and multi-faceted method to assess gait disturbance via video-based analysis. In this study, the dataset containing only 18 patients, whereas the gait symptoms of iNPH have been described exhaustively in the medical domain. The author therefore adopted the first strategy, i.e., manually designed gait features that could be extracted from images according to the description of the symptoms of the disease in medical domain works and developed a method to judge the cerebrospinal fluid (CSF) tap test results of iNPH patients using these features.

Second, the author proposed a method of health indicator estimation using video-based gait analysis. Knowing the health indicators can help with personal health management, and gait analysis is a more convenient and efficient way to estimate health indicators than existing methods. Since this task is targeting the general public, the author conducted experiments for data collection, and obtained gait videos and health indicators from 332 subjects. This dataset is larger than the dataset of iNPH patients, but still not large enough to train a deep learning network from scratch. Therefore, the author extracted gait primitives related to health indicators from an existing large-scale gait database for pre-training, and then fine-tuned the primitive network into a health indicator estimator using the health indicator dataset.

Moreover, the author tried to exchange the methods used in the two aforementioned studies to confirm the suitability of the proposed strategies. As a result, the second strategy yielded worse performance for the iNPH diagnosis support task than the original first strategy, while the first strategy yielded worse performance for the health indicator estimation task than the original second strategy. This indicates the necessity of appropriately switching the strategies depending on the training dataset size as well as the domain-specific knowledge as the author proposed.

List of Publications

A. Journal publications

- A Video-Based Gait Disturbance Assessment Tool for Diagnosing Idiopathic Normal Pressure Hydrocephalus. R. Liao, Y. Makihara, D. Muramatsu, I. Mitsugami, Y. Yagi, K. Yoshiyama, H. Kazui, M. Takeda. IEEJ Transactions on Electrical and Electronic Engineering, Vol. 15, No. 3, pp. 433-441, Feb. 2020.

- Health Indicator Estimation by Video-Based Gait Analysis. R. Liao, K. Moriwaki, Y. Makihara, D. Muramatsu, N. Takemura, and Y. Yagi. IEICE Trans. on Information and Systems, Vol. E104-D, No. 10, pp. 1678-1690, Oct. 2021.

B. International conference and workshop (with review)

- Real-Time Gait-Based Age Estimation and Gender Classification from a Single Image. C. Xu, Y. Makihara, R. Liao, H. Niitsuma, X. Li, Y. Yagi, J. Lu. Proc. of the IEEE Winter Conf. on Applications of Computer Vision 2021 (WACV 2021), online, pp. 3460-3470, Jan. 2021.

C. International conference and workshop (without review)

- Video-based gait analysis in cerebrospinal fluid tap test for idiopathic normal pressure hydrocephalus patients. R. Liao, Y. Makihara, D. Muramatsu, I. Mitsugami, Y. Yagi, K. Yoshiyama, H. Kazui, M. Takeda. In Proc. of The 10th International Workshop on Robust Computer Vision (IWRCV 2015), Beijing, China, Nov. 2015.

D. Domestic conference and workshop (without review)

- Gait Video Analysis for Pre- and Post-CSF Tap Test in Patients with Idiopathic Normal Pressure Hydrocephalus. R. Liao, Y. Makihara, D. Muramatsu, I. Mitsugami, Y. Yagi, K. Yoshiyama, H. Kazui, M. Takeda. In Proc. of the 15th Japanese Conf. of Normal Pressure Hydrocephalus, Osaka, Japan, Feb. 2014.
(特発性正常圧水頭症患者に対する髄液排除試験前後の歩行映像解析.

廖若辰，槇原靖，村松大吾，満上育久，八木康史，吉山顕次，數井裕光，武田雅俊. 第15回日本正常圧水頭症学会，平成26年2月)

- A Study on Body Composition Estimation by Gait Video Analysis. R. Liao, K. Moriwaki, Y. Makihara, D. Muramatsu, N. Takemura, and Y. Yagi. In Proc. of Workshop on Computer Vision and Image Media (CVIM), Okayama, Japan, Sep. 2019.
(歩行映像解析による体組成推定に関する一検討. 廖若辰，守脇幸佑，槇原靖，村松大吾，武村紀子，八木康史. 情報処理学会CVIM研究会，2019年9月)

- Gait Video Analysis for Pre- and Post-CSF Tap Test in Patients with Idiopathic Normal Pressure Hydrocephalus. R. Liao, Y. Makihara, D. Muramatsu, I. Mitsugami, Y. Yagi, K. Yoshiyama, H. Kazui, M. Takeda. In the 29th Symposium of Osaka Advanced Research Collaboration Forum for Information Science and Technology (OACIS), Osaka, Japan, Nov. 2015.
(特発性正常圧水頭症患者に対する髄液排除試験前後の歩行映像解析. 廖若辰，槇原靖，村松大吾，満上育久，八木康史，吉山顕次，數井裕光，武田雅俊. IT連携フォーラムOACIS第29回シンポジウム，平成27年11月)

- A Study on Body Composition Estimation by Gait Video Analysis. R. Liao, K. Moriwaki, Y. Makihara, D. Muramatsu, N. Takemura, and Y. Yagi. In the 75th Academic Conference of the Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan, Nov. 2019.
(歩行映像解析による体組成推定に関する一検討. 廖若辰，守脇幸佑，槇原靖，村松大吾，武村紀子，八木康史. 大阪大学産業科学研究所第75回学術講演会，2019年11月)

- A Study on Body Composition Estimation by Gait Video Analysis. R. Liao, K. Moriwaki, Y. Makihara, D. Muramatsu, N. Takemura, and Y. Yagi. In Dynamic Alliance for Open Innovation Bridging Human, Environment and Materials G3 Subcommittee, Nov. 2019.
(歩行映像解析による体組成推定に関する一検討. 廖若辰，守脇幸佑，槇原靖，村松大吾，武村紀子，八木康史. ダイナミック・アラアンスG3分科会，2019年11月)

- A Study on Body Composition Estimation by Gait Video Analysis. R. Liao, K. Moriwaki, Y. Makihara, D. Muramatsu, N. Takemura, and Y. Yagi. In the 2nd Symposium of Graduate School of Medicine and Institute of Scientific and

Industrial Research, Osaka University, Feb. 2020.
(歩行映像解析による体組成推定に関する一検討. 廖若辰，守脇幸佑，槇原靖，村松大吾，武村紀子，八木康史. 第2回 大阪大学医学系研究科・産業科学研究所懇話会，2020年2月)

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Professor Yasushi Yagi, who provided me many precious opportunities and invaluable advice during my long way towards my Ph.D.

A similar gratitude belongs to my secondary supervisor Professor Yasushi Makihara for his dedicated guidance, continuous support, and patience during my all these years in Yagi Lab.

Next, I would like to thank Professor Ikuhisa Mitsugami, Daigo Muramatsu and Md. Atiqur Rahman Ahad for their treasured support and valuable advice in my studies.

I would also like to thank my reviewers, Professor Fumihiko Ino, Hajime Nagahara, and Katsuro Inoue for their valuable comments that helped me to accomplish this thesis.

Lastly, I would like to thank my dearest family and all my friends and lab mates for their kind help and support in making my study and life in Japan a great time.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AI   Artificial Intelligence

BAS   Backward Arm Swing

BFM_MA Mean of Arms' Body Fat Mass

BFM_ML Mean of Legs' Body Fat Mass

BFM_T Trunk's Body Fat Mass

BFM   Body Fat Mass

BMI   Body Mass Index

BS   Back Straightness

CNN   Convolutional Neural Network

conv   Convolutional Layer

CSF   Cerebrospinal Fluid

DW   Duck-footed Walking

FAS   Forward Arm Swing

fc   Fully-connected Layer

FFM   Fat Free Mass

fps   Frame Per Second

GEI   Gait Energy Image

Nomenclature

GSSR Gait Status Scale-Revised

IBS   InBody Score

iNPH idiopathic Normal Pressure Hydrocephalus

L_LBM Limbs' Lean Body Mass

LBM_MA Mean of Arms' Lean Body Mass

LBM_ML Mean of Legs' Lean Body Mass

LBM_T Trunk's Lean Body Mass

LBM Lean Body Mass

LS   Lateral Sway

MAE Mean Absolute Error

ML   Machine Learning

norm Normalization Layer

NPH Normal Pressure Hydrocephalus

OD   Obesity Degree

PBF   Precent Body Fat

PCA   principal component analysis

PG   Petit-pas Gait

pool Pooling Layer

SLM Soft Lean Mass

SL   Stride Length

SMI   Skeletal Muscle Index

SMM Skeletal Muscle Mass

SVR   Support Vector Regression

TMW Ten Meter Walking

TUG   Timed Up & Go

VFL   Visceral Fat Level

WG   Wide-base Gait

WHR Wasit-Hip Ratio

Chapter 1

Introduction

Human needs are constantly changing with the development of society. In modern society, with the great material and spiritual abundance, the pursuit of human happiness has gradually shifted to health, since the only way to enjoy the material and spiritual happiness is to maintain a healthy body. Especially due to the increase in average life expectancy, many countries and regions around the world are entering or about to enter an aging society, thus increasing the social demand for medical care and health management.

To respond to this demand, the healthcare domain has been actively introducing technologies from other domains for research and development. Among them, artificial intelligence (AI) technology introduced from the domain of information science and technology plays an important role. Earlier, almost all computer-aided diagnosis technologies were referred to as medical AI [1]. In the 1980s, AI approaches to medical imaging referred to technologies that binarize images and extract features using methods such as Hough transform, and then use a set of logical rules to deduce the image content based on the presence or absence of the features [2–6]. Subsequently, statistical models [7–9] and classical machine learning technologies such as Bayesian networks [10, 11], logistic regression [12–14], principal component analysis [15–17], support vector machines [18–20], and decision trees [21–23] have also been applied.

And now, deep learning, the state-of-the-art AI technology, is also being introduced into the healthcare domain. Deep learning is a data-driven approach that discovering intricate structures in high-dimensional data, and has excelled in areas such as image recognition, speech recognition, and natural language processing. Compared to classical machine learning, deep learning requires very little engineering by hand while enabling higher performance, and thus can easily take advantage of increases in the amount of available computation and data [24, 25].

Introduction

One of the most well-known applications of deep learning technology in the healthcare field would be the pedometer function of smartphones. By analyzing the data measured by the built-in accelerometer and gyroscope, the smartphone can count the number of steps taken by the carrier in real time [26]. Furthermore, deep learning methods provide assistance in drug discovery and development for validation of drug targets and optimization of drug structure design [27], and are used in genetic analysis to process large and complex genomic datasets [28].

Deep learning methods are known to excel at image analysis and are therefore often used in clinical diagnosis to analyze medical images, such as computed tomography, magnetic resonance imaging, optical coherence tomography and dermoscopy images. These analyses are used in the diagnosis of various diseases, including cardiovascular disease, retinal disease, tuberculosis, skin cancer, pneumonia [29–33], etc. As practical applications of these technologies, there are already commercial companies offering mature medical AI solutions, such as IBM Watson [34]. In the sudden outbreak of the novel coronavirus 2019, deep learning-based CT image analysis technologies were used for fast screening and enabled large-scale deployment [35].

In addition to the specialized medical imaging above, images captured with common cameras, such as gait images, can also be used in the healthcare-related studies. Since there are diseases that cause gait symptoms, hospitals often capture videos of patients' gait for analysis and diagnosis, indicating that these gait videos contain enough information to analyze the patient's condition [36, 37]. Furthermore, as one of the most common motions, many physical conditions are reflected in gait, such as height, weight, leg length, muscle mass and body fat mass [38–40, 40, 41]. Therefore, it can be assumed that the gait video also includes these health-related information. However, it can be seen through daily life experience that it is very difficult to obtain this information through visual inspection, which requires a long period of training, and the information obtained can be unreliable. Therefore, quantitative analysis of gait to obtain more detailed information has become a natural trend.

In order to obtain quantitative data, early gait studies usually required specialized equipment such as force platforms, wearing goniometers and accelerometers for measurement, and thus the amount of data was very limited. However, with the development of technology in the domain of computer vision, non-contact measurement approaches using cameras have significantly reduced the difficulty of collecting data, allowing deep learning methods to be applied to gait analysis. The latest research can now infer a variety of information from gait such as identity [42–45], age [46–48], gender [48, 49], emotion [50–52] and aesthetic attributes [53, 54].

Table 1.1 Existing major gait databases.

| Name	#Subjects	Covariates	year
CMU Mobo [55]	25	Views, walking speed	2001
Soton small database [56]	12	Views, clothing, carrying, walking speed	2001
Soton multimodal [57]	>400	Views	2011
CASIA database A [58]	20	Views	2002
CASIA database B [59]	124	Views, clothing, carrying	2006
CASIA database C [60]	153	Carrying, walking speed	2006
USF dataset [61]	122	Views, age, carrying, walking surface	2005
AVA [62]	20	Views, body size, walking path	2014
WOSG [63]	155	Views	2013
KY 4D [64]	42	Views, 3D models	2010
OU-ISIR Treadmill Dataset C [65]	200	Views, age	2010
OU-ISIR LP [66]	4016	Views, age	2012
OU-ISIR MVLP [67]	10307	Views	2018
OULP-Age [68]	63846	Age	2017
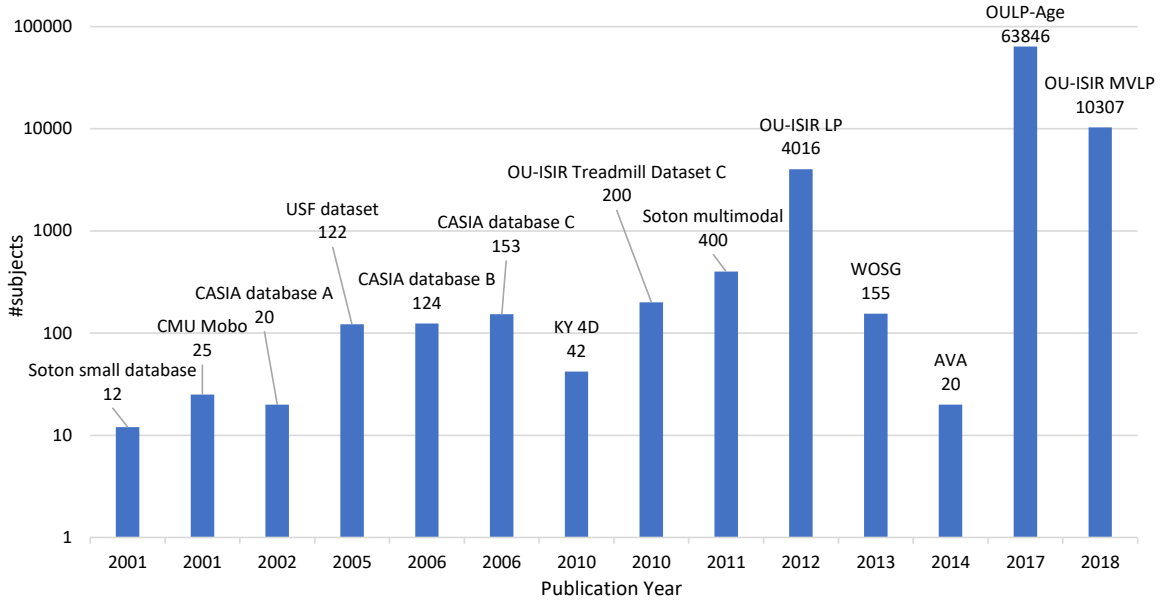
Fig. 1.1 Scale of major gait databases.

Since deep learning requires large amounts of data, those studies of image-based gait analysis rely on large-scale gait databases. Existing major gait databases are summarized in Table 1.1 [69, 68]. It can be seen that the earlier gait databases were smaller in scale, while those released in recent years have become larger (Fig. 1.1). However, these databases focus on factors such as views, age, clothing, carrying, and walking speed, while lacking annotation related to the healthcare domain.

On the other hand, the datasets collected by studies in healthcare domain are hardly sufficient for deep learning. The reason for this is that the annotation of datasets suitable for healthcare domain studies is more specialized than the general gait analysis. For example, in the application of disease diagnosis, only diagnoses made by professional physicians can be taken as the ground truth, and the number of physicians limits the annotation of the dataset. Moreover, the analysis is required to be performed separately according to different diseases, and it is difficult to collect a large enough sample due to the limited number of patients with a specified disease. In some topics targeting general health, the subjects are not restricted by conditions, but data collection still requires specialized personnel and instruments for annotation. Therefore, the scale of the dataset for this type of topics, although it can be larger than the disease diagnosis one's, is still limited. As discussed above, studies in the healthcare domain must endure small datasets, and therefore it is critical that how to obtain results by analyzing small datasets.

The scale of a dataset determines the amount of information it contains. When a dataset is large enough, deep learning can mine the contained information that is valid for achieving
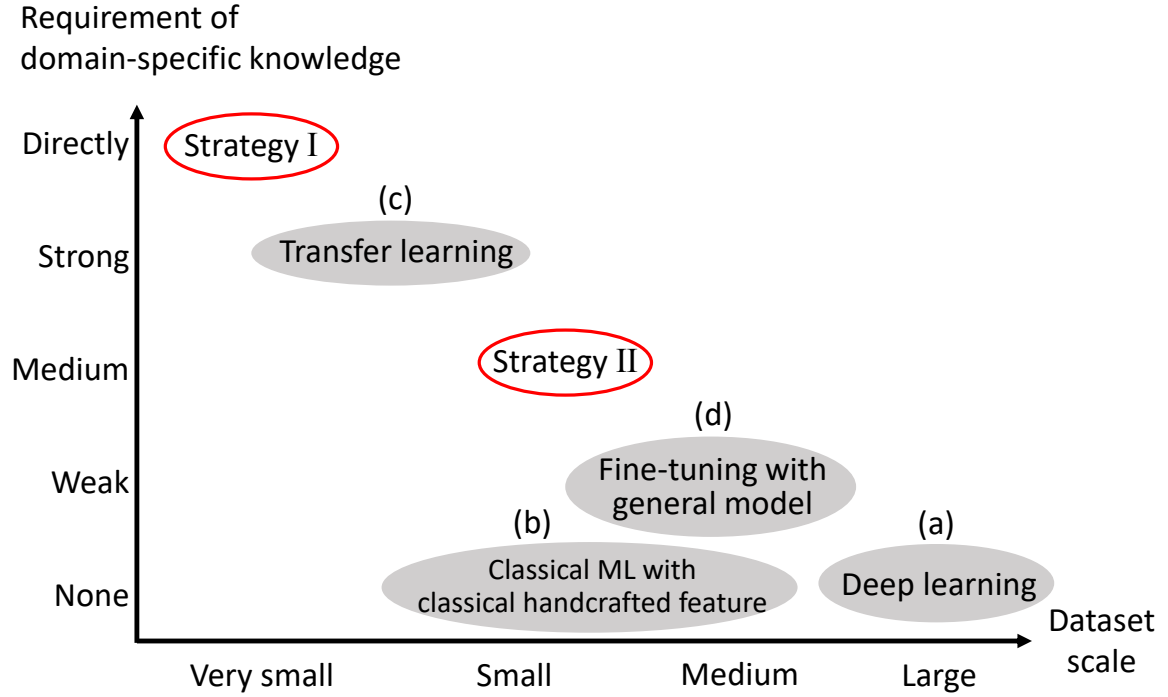
Fig. 1.2 Existing methods for image analysis and the two proposed strategies. The existing methods are shown as solid gray ellipses, and the proposed strategies are enclosed in red ellipses.

a specific task, i.e., domain-specific knowledge, on its own. And when it is undersized, it is necessary to introduce knowledge manually to compensate for the lack of information.

The author summarized the various methods commonly used in image analysis and showed their requirements for external domain-specific knowledge and dataset scale in Fig. 1.2. As shown in the figure, deep learning methods ((a) in the figure) do not need to be supplemented with any external domain-specific knowledge, but can only be used for large-scale datasets.

For the analysis of small-scale datasets, one approach is to extract classical manual image features such as local binary pattern and histogram of oriented gradients, and use classical machine learning (ML) methods such as support vector machines and nearest neighbor algorithm for analysis ((b) in Fig. 1.2) [70, 71]. It is considered that classical ML methods are more suitable for analyzing small-scale datasets than deep learning methods. However, since these methods do not utilize any domain-specific knowledge, these methods, despite being much smaller than deep learning methods, still require datasets of a certain scale. Moreover, such methods have inferior performance compared to deep learning methods with sufficiently scaled datasets, and thus are difficult to achieve complex tasks.

In addition, good performance can also be obtained with small datasets using transfer learning methods ((c) in Fig. 1.2), when the target is the same or similar to the task of an existing trained deep neural network model [72–75]. In this case, models with good performance which trained with large-scale databases (Referred to as the general pre-trained model), e.g., ResNet [76] and VGG [77] models pre-trained with ImageNet [78], are usually used. These general pre-trained models are trained from large-scale databases, and thus it can be considered that the weights of their layers preserve the domain-specific knowledge of their tasks. Those topics that are suitable for the transfer learning approach require approximately the same domain-specific knowledge because their tasks are very similar to those of the pre-trained models. For such topics, only the final output layer of the pre-trained model needs to be retrained, and thus small-scale databases are sufficient.

However, such studies that are fully aligned with the task targets of general pre-trained models are, after all, in the minority, and more studies are on topics that are only weakly related to the task targets of existing trained networks. In such cases, it is still possible to use those general trained models mentioned before. However, since the domain-specific knowledge contained in the general pre-trained model is far removed from those required for the topic, more layers need to be retrained than for transfer learning [79–81]. Such approach is known as fine-tuning ((d) in Fig. 1.2). In order to retrain more layers, more information is needed from the dataset. Therefore, the fine-tuning approach still requires a certain level of dataset scale to obtain adequate accuracy, and are not suitable for healthcare domain studies where only small datasets are available.

As can be seen from the above examples, the need for domain-specific knowledge is in a trade-off with the scale of the topic's dataset, while studies in healthcare domain often have detailed domain-specific knowledge despite the small scale of their datasets. Therefore, the author believes that the dependence on the scale of datasets can be reduced by enhancing the utilization of domain-specific knowledge.

For example, in studies on the diagnosis of specific diseases, the datasets are often extremely small because of the limited number of patients, and are difficult for public sharing due to protection of individual privacy. Such small datasets are not even amenable to analysis using classical ML methods. However, since these diseases have been well studied in the medical domain, their domain-specific knowledge is often very exhaustive. For such cases, the knowledge can then be used to define handcrafted image features to replace classical, general features, e.g., designing extraction methods for features by tracing physicians' approaches to determining symptoms. Also, instead of ML, analysis methods that are specific to the topic and these features may yield better results.

On the other hand, for studies where domain-specific knowledge is less directly related to the topic while the scale of dataset is not adequate to fine-tune a general pre-trained model, it is better to utilize the available domain-specific knowledge to pre-train a basal network for the fine-tuning. In this paper, the author refers to such domain-specific knowledge related to the topic as primitives, and refers to networks pre-trained with the primitives as primitive networks. The fine-tuning of such primitive networks will give better results than using those general pre-trained models, which are trained with larger datasets, yet the domain-specific knowledge contained is unrelated to the topic.

The author has studied topics with different scales of datasets and domain-specific knowledge with a view to proving the idea. The contributions of this work are summarized as follows.

1. Strategy for gait analysis through small-scale datasets

In research in the domain of gait analysis, leveraging known domain-specific knowledge is an effective solution when the scale of the dataset is not sufficient to directly use deep learning methods. In this work, two strategies are proposed to cope with the problem of undersized datasets as follows.

I. In cases where the dataset scale is too small to support any ML methods, it is necessary to find direct and exhaustive domain-specific knowledge and manually design analyzable features according to that knowledge.

II. For cases where the dataset scale is small yet still sufficient for fine-tuning the deep learning network, information that is as relevant to the topic as possible while being extractable from existing large-scale databases can be selected as domain-specific knowledge and utilized by pre-training primitive networks.

The author conducted following two studies to demonstrate the effectiveness of these strategies.

2. Judgement of cerebrospinal fluid (CSF) tap test for idiopathic Normal Pressure Hydrocephalus (iNPH) patients from walking test video.

iNPH is a disease that predisposes people of advanced age and can lead to dementia, gait disturbances and other symptoms. The CSF tap test is used to determine whether a patient can be cured by a specific surgery, and the results are judged primarily by improvements in gait symptoms, which are therefore well-described in the medical domain. However, due to the aforementioned difficulties, the scale of its dataset is very small (less than 20 subjects). The author therefore adopted strategy I, i.e., manually designed gait features that could be extracted from images according to the description of the symptoms of the disease in medical

domain works, and developed a method to judge the CSF tap test results of iNPH patients using these features.

3. Estimation of health indicators related to body composition from a walking video.

Knowing health indicators can help with personal health management, and gait analysis is a more convenient and efficient way to estimate health indicators than existing methods. Since this topic is targeting the general public, the subject is not as restricted as the previous study. However, the scale of the dataset is still limited (300+ subjects) because the measurement of health indicators requires special instruments and has to organize specialized activities to collect data. In addition, there is no existing literature that directly describes how the health indicators affect gait as in the previous study. The author therefore adopted strategy II, i.e., chose some gait characteristics that could be extracted on a large-scale database as the sub-optimal of the domain-specific knowledge, and use them to pre-train the deep learning networks.

Chapter 2

A Video-Based Gait Disturbance
Assessment Tool for Diagnosing
Idiopathic Normal Pressure
Hydrocephalus

2.1 Introduction

Normal pressure hydrocephalus (NPH) is a syndrome first described by Hakim and Adams [82] that is characterized by the classic clinical trial of gait disturbance, cognitive dysfunction, and urinary symptoms. NPH is further classified into secondary and idiopathic NPH (iNPH). iNPH develops without an identifiable cause, and occurs in approximately 1% of older adults [83, 84].

According to the Japanese guidelines for the management of iNPH [85], iNPH is classified into three diagnostic levels: preoperatively "possible", "probable", and postoperatively "definite". The diagnostic flow is provided as follows:

Step 1: Patients with at least one of the symptoms of dementia, gait disturbance, and urinary incontinence (known as the clinical triad) [86, 87], and who meet other criteria, are diagnosed as possible iNPH.

Step 2: The possible iNPH patients who meet the criteria of a cerebrospinal fluid (CSF) examination, and who have one of the investigational features obtained by morphologic brain imaging [88, 89] or a CSF tap test [90, 91], are diagnosed as probable iNPH. The CSF tap test is judged by the improvement of the clinical triad, particularly the gait disturbance [92, 93]. More specifically, a patient undergoes fixed-distance walking tests such as the Timed Up &

Go (TUG) test [94, 13] before and after the lumbar puncture of the CSF tap test. Thereafter, the patient is labeled as tap-positive if they gain >10% improvement in walking time as an objective (quantitative) criterion, or improvement in gait disturbances through visual inspection by a physician as subjective (qualitative) criterion, otherwise, the patient is labeled as tap-negative. The tap-positive patients are then diagnosed as probable iNPH.

Step 3: A surgical procedure (i.e., a shunt surgery) is indicated for the probable iNPH patients, and the patients with improved postoperative symptoms are diagnosed as definite iNPH.

In the present study, we focused on the CSF tap test (Step 2). Since it is difficult to correctly judge all patients using a single objective criterion (i.e., the walking time), additional criteria are required. The criteria associated with qualitative gait disturbance assessment through visual inspection by a physician are useful, however, they are strongly dependent on the subjectivity and degree of proficiency of the individual physicians. Indeed, the physician requires considerable experience to make a correct judgment through visual inspection. Thus, it is necessary to develop a quantitative method to assess gait disturbances from multiple aspects to improve the accuracy of diagnosis.

Other studies assessing gait disturbance to diagnose iNPH have also used qualitative visual inspection-based methods including the dynamic gait index [95], the Tinetti performance-oriented mobility assessment [96], or a customized criterion [97, 98]. Although there are several studies that discuss the quantitative assessment of gait disturbances in iNPH, the techniques require special equipment such as accelerometers [99], footswitches [100, 101], knee goniometers [100], and motion capture systems [101], which are unlikely to be available for routine clinical use.

By contrast, video-based gait disturbance assessment is promising for routine clinical use as patients do not need to wear accelerometers or goniometers, and setup of a commercially available digital video camera is easier than that for aligned footswitches. Indeed, the efficacy of video-based gait analysis is well established in the fields of computer vision, pattern recognition, and biometrics. In particular, holistic silhouette-based gait representation [102, 103] is widely used in video-based gait analysis, with numerous applications including human identification, forensics, gait personality measurement, gender classification, and age estimation [104–109].

The above-mentioned holistic silhouette-based gait representations are, however, not necessarily effective for medical/health applications and hence we often require more directly interpretable gait features (e.g., stride [110], sudden motion variation [111], etc.) in the medical/health applications. The existing methods in medical/health field [112, 111, 110] usually extract the directly interpretable gait features from a side-view image or multi-view

images including side-view images. Since it requires a relatively large space to capture a walking image sequence from the side-view, application scenes are limited for the methods, which require the side-view images. Considering that walking tests are often conducted at narrow corridors in the hospital in clinical practice, it is desirable to extract the gait features from frontal view to save space.

The aim of the present study is to quantitatively assess the gait disturbance of iNPH patients using video-based gait analysis (without the requirement for any special equipment), and to validate its efficacy for judgment of the CSF tap test. More specifically, we select four gait features from the gait status scale-revised (GSSR), which is used in subjective assessment by physicians, and develop a method to quantitatively assess the four gait features to judge the CSF tap test more accurately compared with that using a single quantitative criterion (i.e., the walking time).

2.2 Method

2.2.1 Patients

We recruited eighteen patients (twelve women and six men) who were previously diagnosed with possible iNPH at Osaka University Hospital (Table 2.1). CSF tap tests were performed on the patients, and a set of walking tests were performed twice daily over at least five days just before and after the tap tests. Patients judged as tap-positive in the tap test were indicated for shunt surgery, and another set of walking tests was performed after the surgery to verify the tap test results. Two tests were included in each set of walking tests: the 3 m TUG, and the 10 meter walking test (TMW). The tests are both round-trip walks, where the TUG starts from sitting on a chair, and contains the actions for standing and sitting, whereas the TMW does not. The best time scores for each of the two walking tests were taken from each set performed before and after the CSF tap test. If either or both of the best time scores of the TUG and the TMW for a patient improved by >10% after the tap test, the patient was judged as tap-positive, and vice versa. Nine patients were judged to be tap-positive in the preoperative walking tests, while the others were tap-negative. Of the tap-positive patients, two had no improvement in their walking time (defined here as speed-negative patients) in both the TUG and the TMW, and were diagnosed with tap-positive by gait improvement assessed through visual inspection by physicians. However, the results of the postoperative walking tests proved that all nine patients had definite iNPH, indicating that the tap-positive assessment of the speed-negative patients was correct.

Table 2.1 Patient characteristics and walking test scores.

Patient	Gender	Age (yr.)	Result of tap test	>10% improvement	
				TUG	TMW
P1	F	71	Positive	Yes	No
P2	F	82	Positive	Yes	Yes
P3	M	75	Positive	Yes	No
P4	F	67	Positive	No	Yes
P5	F	80	Positive	Yes	No
P6	F	76	Positive	Yes	No
P7	F	68	Positive	Yes	No
P8	F	78	Positive	No	No
P9	M	75	Positive	No	No
N1	F	72	Negative	No	No
N2	M	78	Negative	No	No
N3	F	82	Negative	No	No
N4	M	69	Negative	No	No
N5	F	81	Negative	No	No
N6	M	75	Negative	No	No
N7	F	87	Negative	No	No
N8	M	72	Negative	No	No
N9	F	83	Negative	No	No

For gait disturbance analysis, we captured walking videos of the patients from the frontal view using a consumer digital video camera (Sony Corporation, Tokyo, Japan), with a resolution of 1920×1080 pixels and a framerate of 30 frames per second. The videos included multiple trials of TUG and TMW, which were performed before and after the lumbar puncture.

2.2.2 Gait feature assessment

Gait features

An example of a typical gait of a patient with iNPH is shown in Fig. 2.1. To select the gait features to be assessed, we referred to an existing scale of gait disturbances of iNPH, termed the GSSR [113]. GSSR is a qualitative scale that relies on visual inspection by a physician. The 10 factors of gait disturbance that the GSSR focuses on and their scoring criteria are defined as follows.

Fig. 2.1 Image sequence of the typical gait of an iNPH patient, with a frame rate of 6 frame per second.

1. Postural stability: Response to sudden, strong posterior displacement produced by a pull on the shoulders.

   0: Stop by oneself with $<$2 steps

   1: Stop by oneself with 3–5 steps

   2: rush $>$6 steps but can stop by oneself

   3: gradually rush and cannot stop by oneself

   4: just fall down without a foot appearing backward

2. Wide-base gait

   0: None

   1: Present

3. Petit-pas gait

   0: None

   1: Present and steps with the distance between two feet $>$1x foot length

   2: The distance between two feet $<$1x foot length

4. Freezing of gait

   0: None

   1: The first steps slower or smaller than the subsequent ones

   2: A walk stops on the way

5. Shuffling gait

   0: None

   1: Present

6. Independence of walking

   0: Normal

   1: Necessity of monitoring

   2: Necessity of assistance

7. Lateral sway

14

0: None

1: Lateral sway of the trunk without any foot corrections

2: Lateral foot corrections during walking are present

8. Festinating gait

    0: None

    1: Accelerate but can stop by oneself

    2: Accelerate and cannot stop by oneself

9. Disturbed tandem walking: >2 foot corrections in 8 steps of open-eye tandem walking

    0: None

    1: Present

10. Duck-footed Walking

    0: None

    1: Present

The scale requires that each item of the GSSR is rated with a discrete value (e.g., 0 and 1 for wide-base gait), and the total score is then used to judge if the patient has a gait disturbance caused by iNPH. In the present study, we focused on four items (lateral sway, petit-pas gait, wide-base gait, and duck-footed walking) because the features associated with these items can be observed from frontal gait image sequences (i.e., videos) of the patient, which can be captured relatively easily during walking tests.

Assessment method

The proposed assessment methods of the four items are described below.

Pre-processing: We manually selected the stable-walking sequences for assessment from the video, beginning from the frame in which the patient moved his/her foot to take the first step, until the frame in which he/she takes a step to change direction before the turn-back point. Because the walking distance was only 3 m in TUGs, the patients' stable-walking sections were too short for assessment. Thus, we selected all image sequences from the TMWs. To prepare for the following assessment, we first detected the patients' faces with a commercial off-the-shelf face detector (OKAO Vision; OMRON Corporation, Kyoto, Japan), and blurred their faces for privacy protection. We then estimated the camera orientation

based on the parallel lines of a corridor in the hospital, and corrected the image rotation/tilt using the camera orientation so that the image plane is perpendicular to both the ground and the wall surface. It allows us to obtain the horizontal or vertical distance of the two points by measuring the coordinates. For the silhouette-based gait analysis, we extracted the whole-body silhouette of a patient. Because gait is periodic, we selected a subsequence composed of two stable gait periods with sufficient image resolution of the patient from the entire image sequence to reduce the time required for silhouette extraction. We normally chose the last two gait periods before turning around that had no step fluctuation. We extracted the silhouette using the grab-cut algorithm [114] for semi-automatic segmentation, and added manual interventions to create a higher-quality silhouette. We also manually set the bounding boxes for the patients' regions throughout the whole sequence.
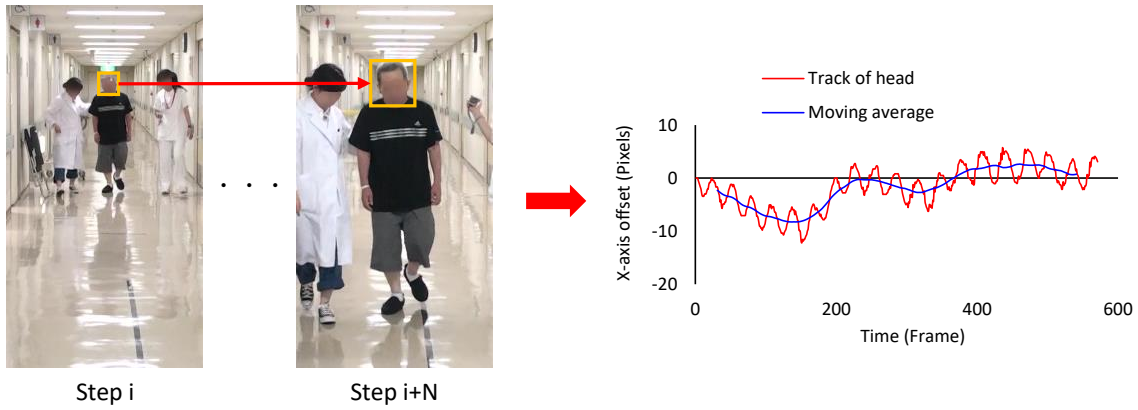

Fig. 2.2 Assessment of lateral sway (LS). The X-axis position of the patient's head is recorded as a trajectory as shown in the red curve on the right, and the moving average of this trajectory is calculated using the gait period as the period (blue curve).

Lateral sway: Lateral sway assesses the trunk's sway during walking. As defined by the GSSR, patients with trunk sway are given a score of 1, and patients with fluctuations in their heel strike positions are given a score of 2. In the present study, we only used the score of 1 to assess trunk sway. As shown in Fig. 2.2, we computed the trajectory of each patient's head position along a horizontal axis provided by face tracking, and subtracted the moving average of the trajectory to calculate the deviation of the head from the midline of the body. We then normalized the deviation by the patient's height (i.e., the height of the bounding box to the patient) because the amount of lateral sway is scaled by the patient's height, and computed the mean of the normalized deviation as the lateral sway score

$$q_{\text{LS}} = \frac{1}{h_i} \frac{1}{n - 2\lfloor p/2 \rfloor} \sum_{i=\lfloor p/2 \rfloor}^{n-\lfloor p/2 \rfloor} (c_i - \bar{c}_i),$$

16

where h_i [pixel] is the height of bounding boxes (i.e., the height of a patient) at the i-th frame, n [frame] is the number of frames in an image sequence, p [frame] is the gait period of patient, $\lfloor \cdot \rfloor$ is a floor function, c_i is the horizontal position of the center of patient's head at the i-th frame, and $\bar{c}_i$ is a moving average of c_i with a period of p, which can be computed as,

$$\bar{c}_i = \frac{1}{p} \sum_{k=i-\lfloor p/2 \rfloor}^{i-\lfloor p/2 \rfloor + p - 1} c_k.$$

Petit-pas gait: Petit-pas gait literally means "gait with a very small stride". We counted the number of steps during the walking test and calculated the stride. Because the stride is scaled by the patient's height, we normalized the stride by the height of the bounding boxes, and then defined the normalized value as the petit-pas gait score

$$q_{\text{PG}} = \frac{1}{h_{\max}} \frac{L}{n_{\text{st}}},$$

where $h_{\max}$ is the maximum height of bounding boxes when patient reach the turn-back point, L is the length of the walking course, and n_{st} is the number of steps.

Silhouette Silhouette of lower body Foot region Lowermost point
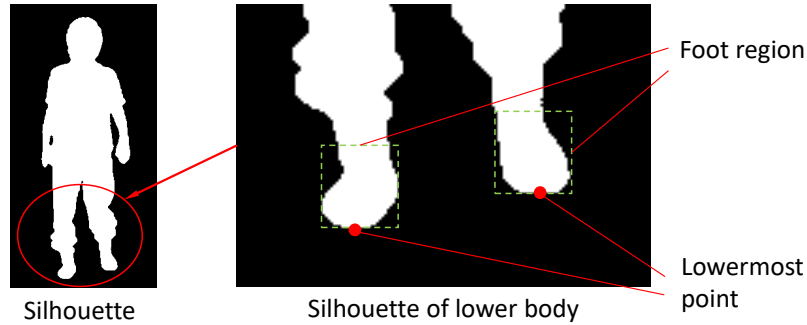
Fig. 2.3 Detection of stance phase for the assessment wide-base gait and duck-footed walking. Since both symptoms are assessed when the foot is in stance phase, we detected both feet in each frame of the gait video and labeled them separately whether it was in stance phase or not.

Wide-base gait and duck-footed walking: A wide-base gait means that the lateral interval between the feet of the patient while the walking is wide. Duck-footed walking means that the patient's large toe is oriented outwardly during the stance phase. Because both features are observable during the stance phase, we try extracting the stance phases from a silhouette sequence. As shown in Fig. 2.3, we first locate feet search window whose bottom, left, and right coincide with the bottom, the left, and the right of the bounding box, respectively, and whose height is set in proportion to the height of the bounding box. We then divide
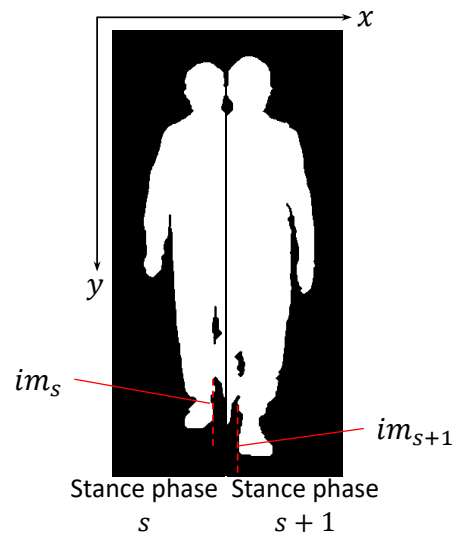
Fig. 2.4 Assessment of wide-base gait. The innermost foreground pixel within the foot region which is in stance phase is recorded as the innermost point of the foot.
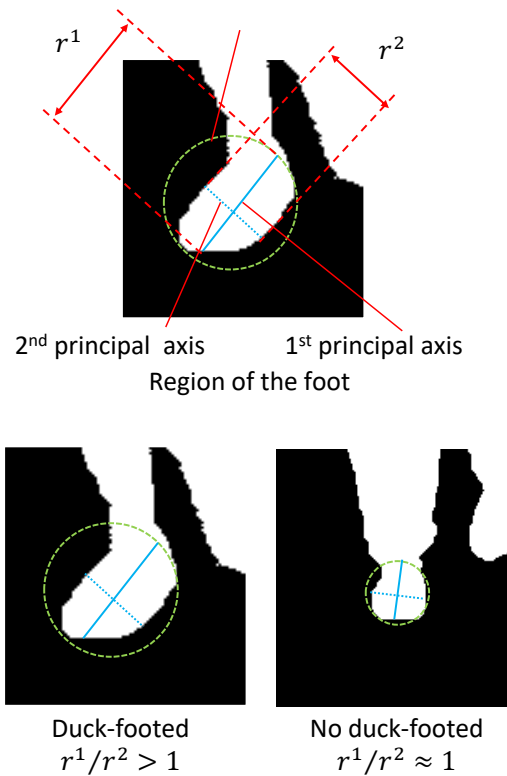

Fig. 2.5 Assessment of duck-footed walking. The area inside the green dashed circles was used for PCA calculations. The two figures below are typical examples of foot silhouettes with and without duck-footed walking, respectively.

the window by its horizontal middle point into left/right foot search sub-windows, and a silhouette region inside the left/right foot search sub-window is regarded as left/right foot region. We then extract the bottom of the left/right foot region as the foot bottom points. Finally, we can estimate a phase ϕ_i (stance phase or non-stance phase) at the i-th frame by analyzing the vertical movement of the foot bottom points based on the assumption that the vertical movement should be small at the stance phase as

$$\phi_i = \begin{cases} \text{Stance phase,} & \Delta y_i^b \leq \alpha \\ \text{Non-stance phase,} & \text{otherwise} \end{cases},$$

$$\Delta y_i^b = y_{i+1}^b - y_i^b,$$

where y_i^b is the vertical position of the foot bottom points, and α is a threshold to judge the phase. The threshold is automatically determined by Otsu's discriminant analysis criterion on a histogram obtained by a set of the vertical movements $\{\Delta y_i^b\}$ for each patient.

We analyzed the silhouettes of the stance phases for evidence of a wide-base gait or duck-footed walking.

In the assessment of wide-base gait (Fig. 2.4), we calculated the horizontal distance between the innermost points of the foot silhouettes during adjacent stance phases, and normalized them by the height of bounding boxes. The mean of these normalized values was defined as the wide-base gait score

$$q_{\text{WG}} = \frac{1}{n_{\text{sp}}} \sum_{s=1}^{n_{\text{sp}}-1} \text{distance}(x_s, x_{s+1}),$$

$$\text{distance}(x_s, x_{s+1})$$
$$= \begin{cases} x_s - x_{s+1}, & \text{Stance phase } s \text{ is left foot} \\ x_{s+1} - x_s, & \text{Stance phase } s+1 \text{ is left foot,} \end{cases}$$

where n_{sp} is the total number of stance phase of both feet, and x_s is the horizontal position of the innermost point of the landing foot's silhouette in the stance phase s.

In the assessment of duck-footed walking (Fig. 2.5), we first extracted the foot silhouette's region by setting bounding boxes for the left/right foot that were proportional to the full-body bounding box. The proportion was pre-defined experimentally so that the bounding boxes can contain the left/right foot. We then computed the principal axes of the foot silhouette's region by principal component analysis (PCA) during the stance phase, and assessed the length of first and second principal axes. The ratio of the length of the first and second principal axes is close to 1 when the toe is straight forward, and is larger as the toe is oriented outwardly.

We therefore defined the sum of the ratios of the two feet as the score of duck-footed walking. The score is defined by

$$q_{DW} = \sum_{f \in \{left,right\}} \frac{1}{n_{sp}^f} \sum_{s=1}^{n_{sp}^f} \frac{r_s^{f,1}}{r_s^{f,2}},$$

where $r_s^{f,k}$ is the length of the k-th principal component of the foot f's region in stance phase s, and n_{SP}^f is the number of stance phases of the foot f.

Unlike the first two features, wide-base gait and duck-footed walking are assessed from only two walking cycles where the silhouettes are extracted.

### 2.2.3   Judgment of the CSF tap test

We used the four gait features (i.e., lateral sway, petit-pas gait, wide-base gait, and duck-footed walking) extracted from a gait image sequence and the walking time to improve the judgment for the CSF tap test. Because it was previously reported that the speed of turning around is important for the CSF tap tests [115, 116], we also used the time for turning around at the halfway point of the TUG and the TMW, in addition to the total walking time. Specifically, we manually set the frame where the patient moved his/her foot to take the first step as the start time, and set the frame where the patient crossed the finish line as the end time, to assess the total time spending for walking. Similarly, we assessed the turning time by selecting the frame where the patient changed direction to turn as the start time, and set the frame where the patient completed the turn and took the first step to walk straight as the end time.

We used a statistical method to calculate a score for the judgment. We employed a t-test for each feature, and calculated the p-values of each assessment before and after the tap test. We then selected the minimum p-value of all the features as the judgment scores (i.e., we selected the feature with the best improvement).

2.3   Results

### 2.3.1   Feature assessment

Assessment of the four features (i.e., lateral sway, petit-pas gait, wide-base gait, and duck-footed walking) is shown in Fig. 2.6. Since there are no other studies that have quantitatively measured these gait features by video analysis, we compared the assessed values of the proposed method with the GSSR scores determined by the physicians. We grouped patients by the GSSR scores and sorted them by the assessed values within each group. The assessed
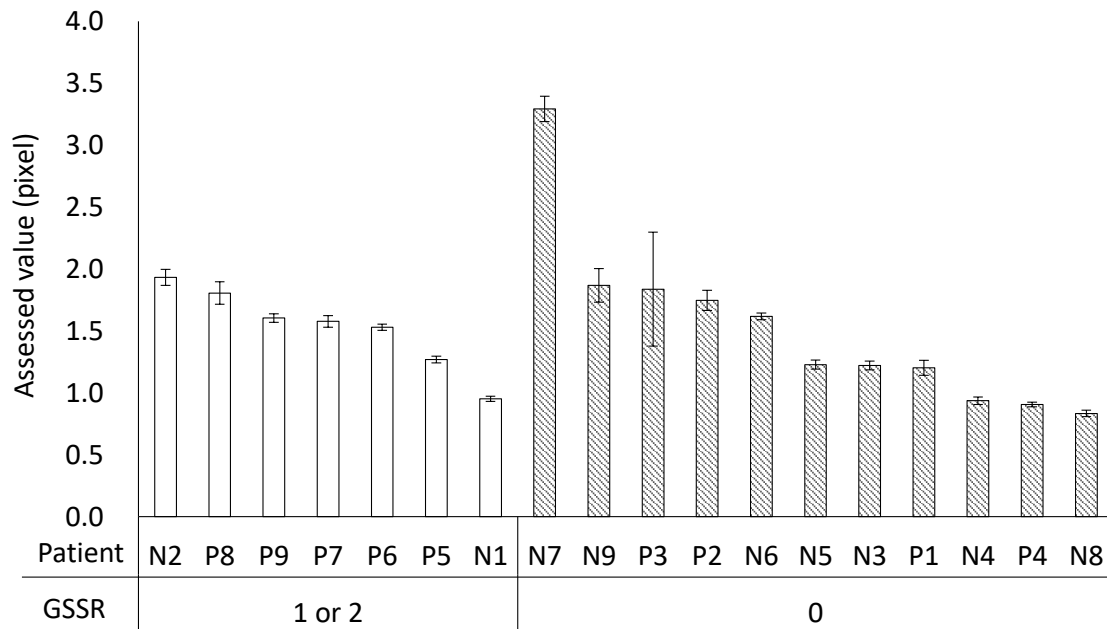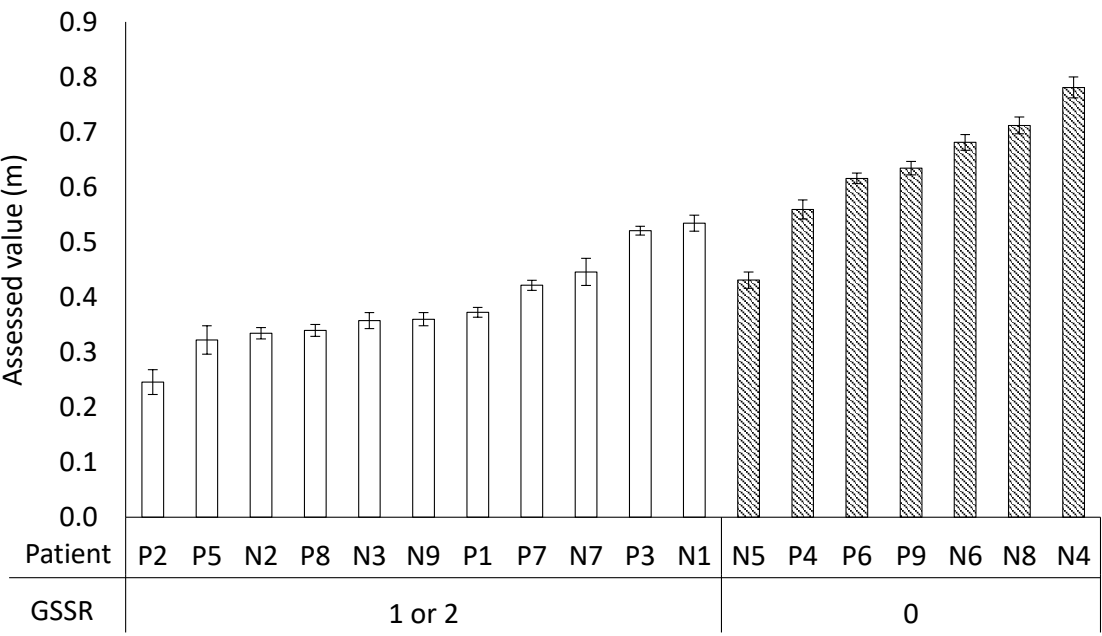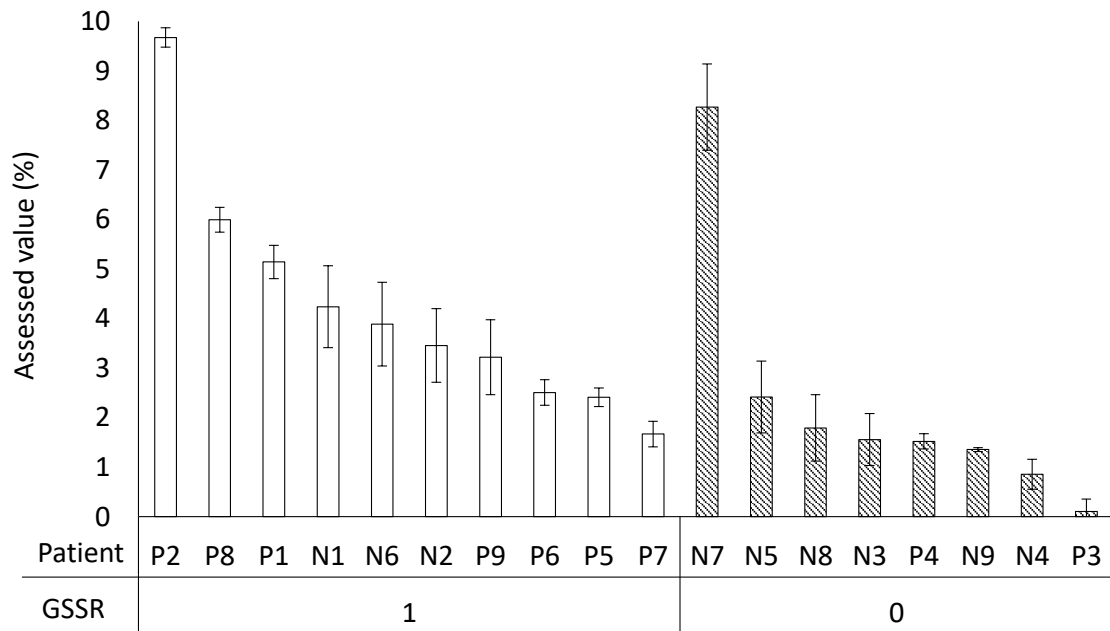
(a) Lateral sway, pre-tap.
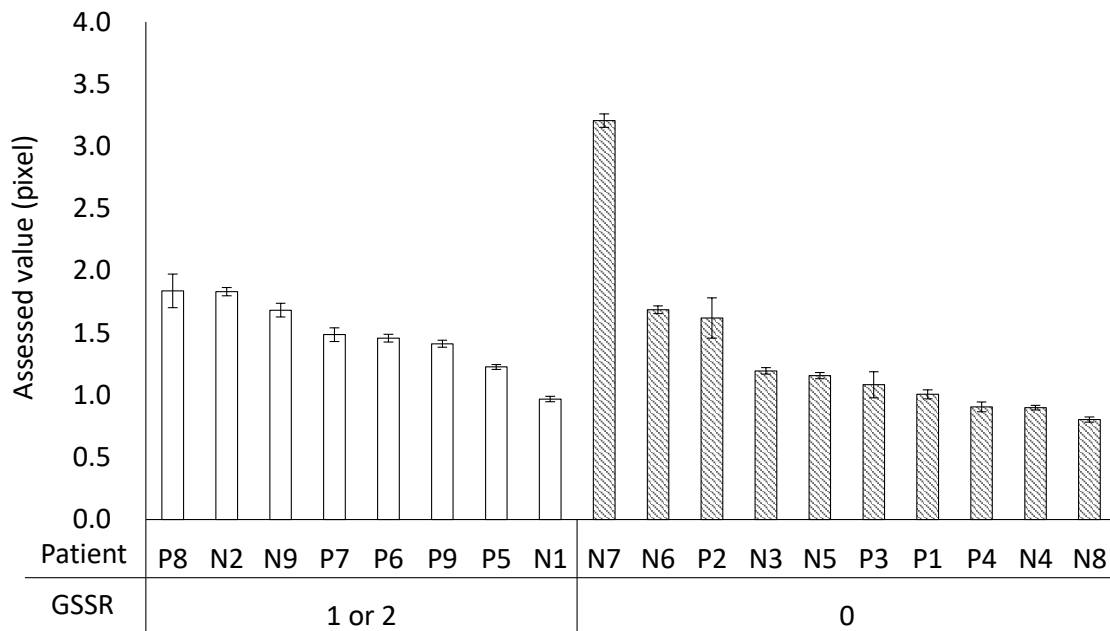
(b) Lateral sway, post-tap.
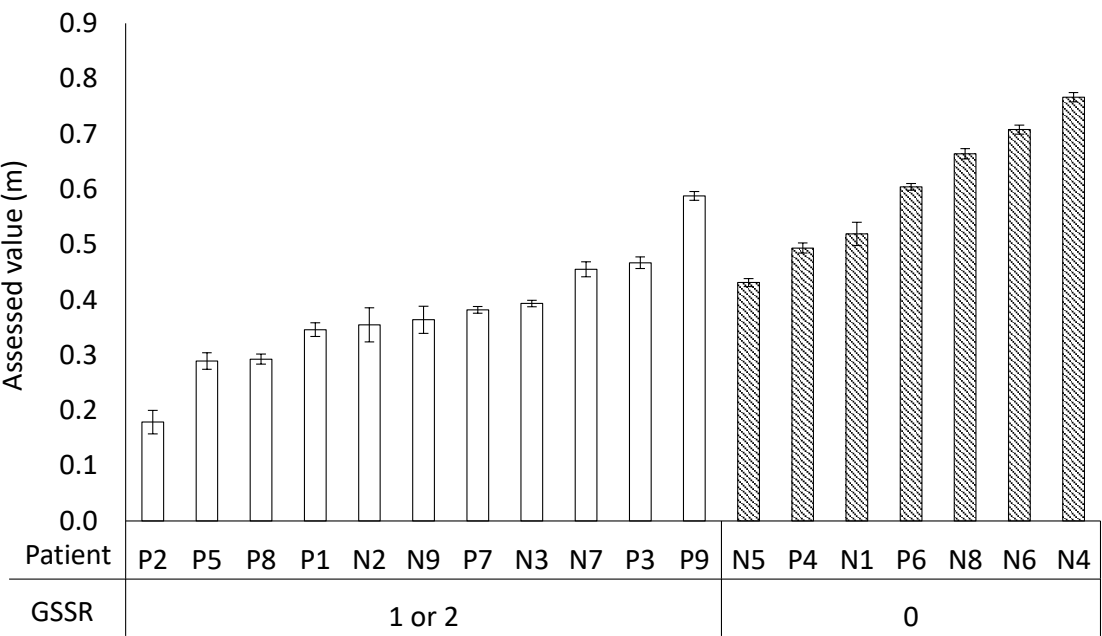
(c) Petit-pas gait, pre-tap.

(d) Petit-pas gait, post-tap.

(e) Wide-base gait, pre-tap.
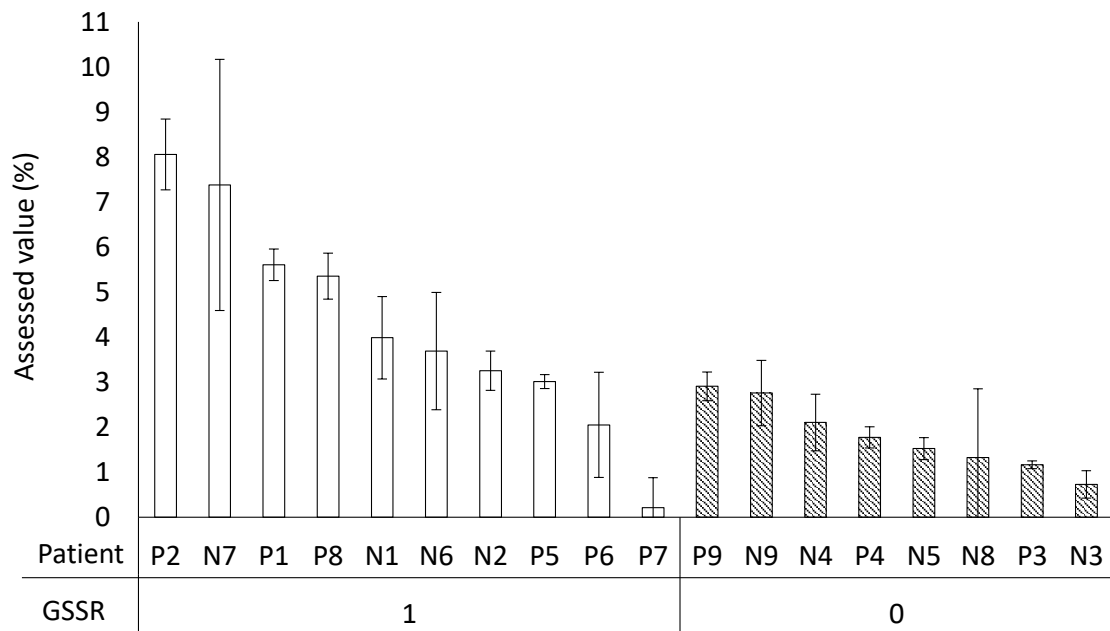
(f) Wide-base gait, post-tap.
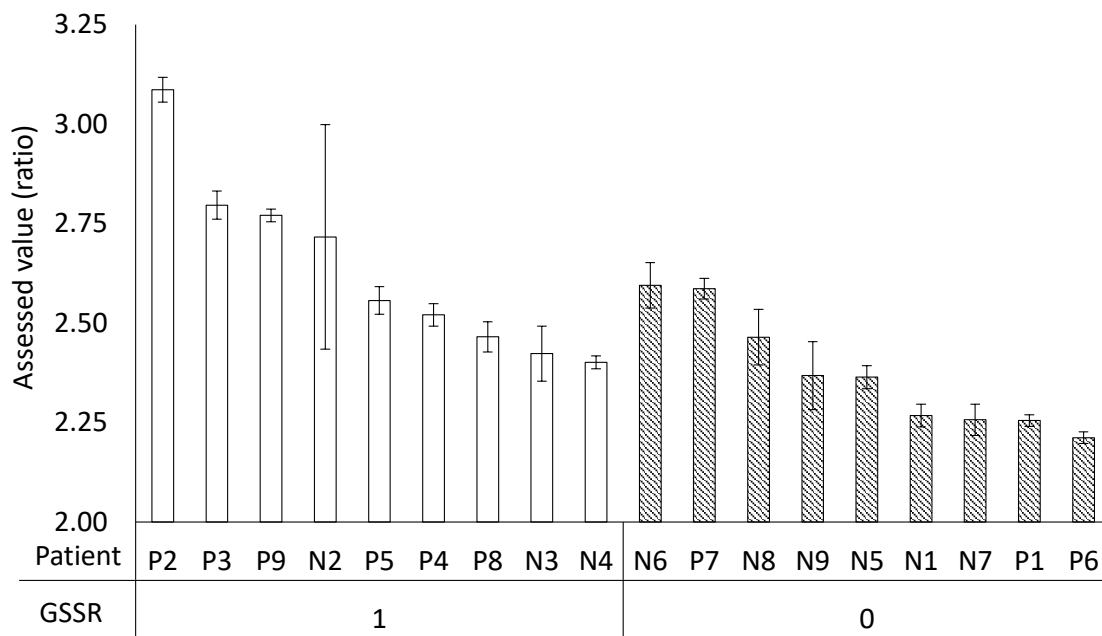
(g) Duck-footed walking, pre-tap.
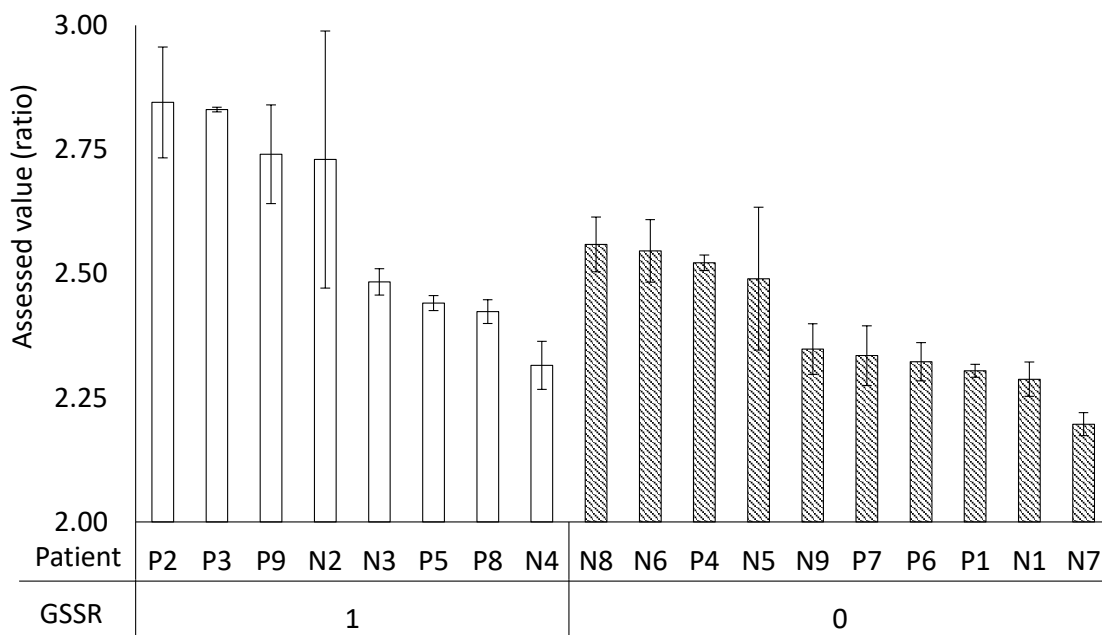
(h) Duck-footed walking, post-tap.

Fig. 2.6 The mean ($\pm$ standard deviation for sample mean) of the assessed values compared with GSSR scores that are labeled by the treatment physician. For lateral sway, wide-base gait, and duck-footed walking, patients with the symptoms should have a higher value than those without symptoms, while for petit-pas gait the symptomatic patients should have a lower value.
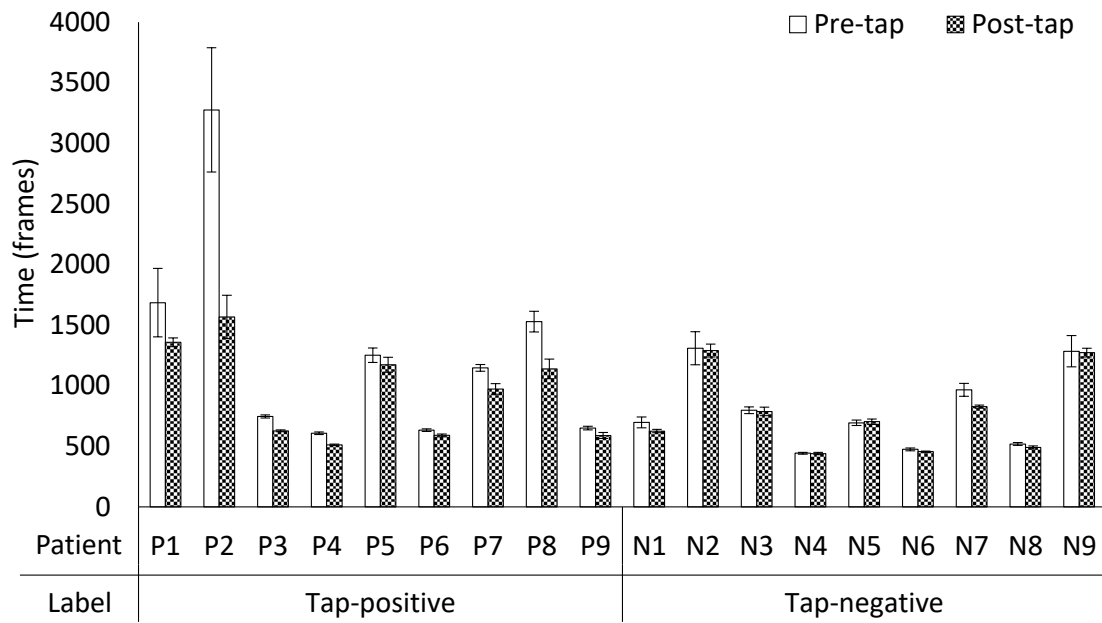
(a) Total time of TMW.
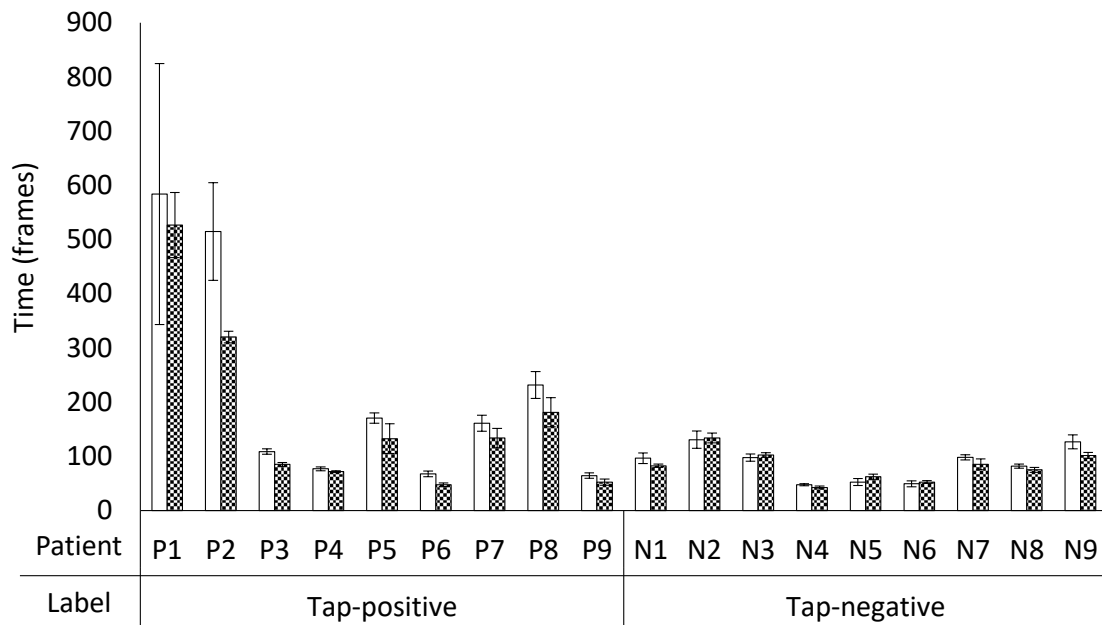
(b) Turning time in TMW.
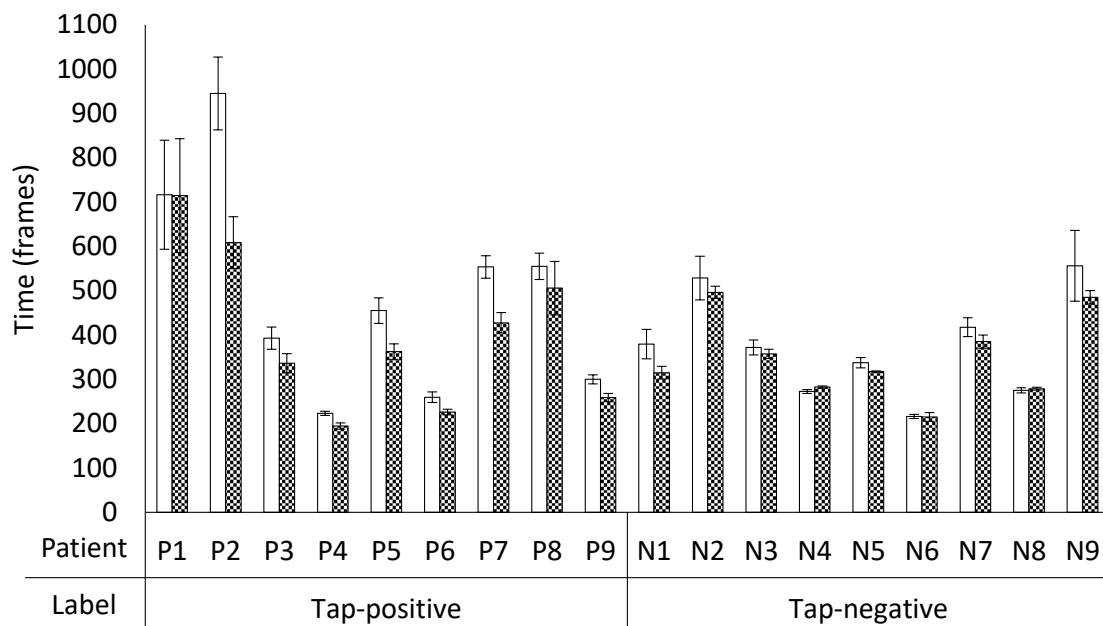
(c) Total time of TUG.
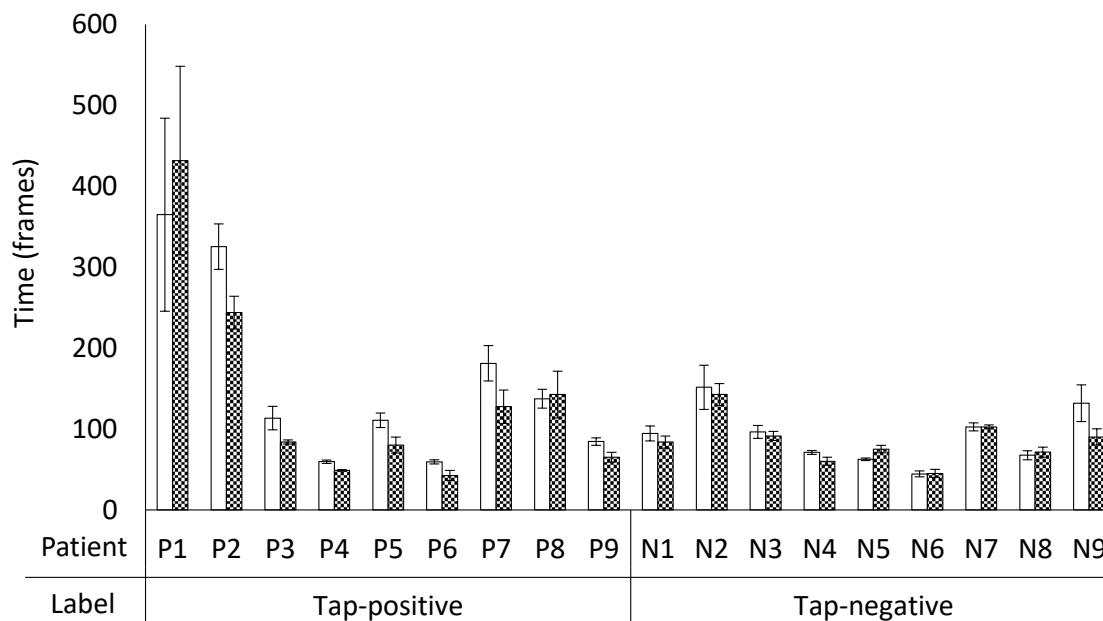
(d) Turning time in TUG.

Fig. 2.7 Time measurements, as shown by the average ± standard deviation for the sample mean. The improvement in the total walking time is slightly different from that in Table 2.2, e.g., P1, P3, P8, and P9. This is because the physician uses the best score to judge improvement, while we use the mean and standard deviation for the sample mean.

values of the time measurements are also shown in Fig. 2.7, which are used to help judging the results of the tap-test in current clinical practice.

As shown in Fig. 2.7, there was a significant improvement in the walking time of some tap-positive patients (e.g., patients P2, P3, and P6), while the walking time is a quantitative feature used in current clinical practice. The tap-positive patients did not always exhibit significant time improvements (e.g., patient P1), implying that the walking time alone is insufficient for accurate judgment of the tap-test. However, we noticed that patient P1 had a significant improvement in lateral sway ($p<0.01$), while the speed-negative patients P8 and P9 had a significant improvement in lateral sway and petit-pas gait.

We also compared the assessed values obtained by our quantitative assessment method with the GSSR scores determined by the subjective physician assessments. As shown in Fig. 2.6, the assessed values and the GSSR scores showed a similar trend for petit-pas gait, wide-base gait, and duck-footed walking, except for a few cases. However, there were no differences in the assessed values of lateral sway between the patients with a score of 0 and those with a score of 1 or 2.

2.3.2 Judgment of the CSF tap test

For each patient, we calculated the p-values before and after the tap-test for each gait feature and for walking time. The results are shown in Table 2.2, and the p-value was compared among all patients for each feature in Fig. 2.8 The selected minimum p-value for each patient is shown in Fig. 2.9, which is used to judge the results of the tap test. The assessed values of the corresponding features before and after the CSF tap test are listed in Table 2.3. The minimum p-values for the two speed-negative patients (i.e., P8 and P9) were <0.005, which supports the requirement for multi-faceted assessments to improve clinical judgment by the physician.

We also found that the minimum p-values of all tap-positive patients were always less than those of all tap-negative patients. Thus, we can successfully confirm tap-positive and tap-negative patients based on the minimum p-value derived from the assessed gait disturbance features with an appropriate threshold (e.g., $p=0.01$).

Table 2.2 The p-values of all gait features and walking time measurements. The minimum p-value of each patient is marked in bold. LS: Lateral Sway, PG: Petit-pas Gait, WG: Wide-base Gait, DW: Duck-footed Walking, TMW/TUG total: Total time for TMW/TUG test, TMW/TUG turn: Turning time in TMW/TUG test.

| Patient	LS	PG	WG	DW	TMW total	TWM turn	TUG total	TUG turn
P1	**0.006**	0.059	0.824	0.990	0.144	0.412	0.496	0.649
P2	0.244	0.032	0.044	0.038	0.006	0.034	**0.004**	0.020
P3	0.061	0.001	0.999	0.825	**7.E-06**	0.002	0.060	0.040
P4	0.485	0.021	0.809	0.519	**2.E-05**	0.101	0.013	4.E-04
P5	0.105	0.198	0.988	**0.003**	0.213	0.190	0.016	0.054
P6	0.036	0.174	0.381	0.912	0.016	**0.004**	0.017	0.034
P7	0.105	0.006	0.120	0.052	0.009	0.137	**0.002**	0.054
P8	0.576	0.005	0.158	0.184	**0.003**	0.102	0.251	0.562
P9	**3.E-05**	0.010	0.362	0.407	0.043	0.072	0.008	0.021
N1	0.697	0.278	0.426	0.657	0.078	0.100	**0.053**	0.194
N2	**0.090**	0.724	0.414	0.513	0.451	0.564	0.274	0.388
N3	0.283	0.957	**0.122**	0.768	0.420	0.731	0.241	0.308
N4	0.145	0.291	0.872	0.153	0.429	0.116	0.956	**0.114**
N5	**0.059**	0.503	0.158	0.730	0.610	0.874	0.076	0.968
N6	0.951	0.865	0.455	0.304	**0.088**	0.691	0.462	0.536
N7	0.230	0.389	0.402	0.130	**0.018**	0.195	0.131	0.500
N8	0.185	**0.020**	0.408	0.827	0.075	0.139	0.704	0.666
N9	0.109	0.555	0.848	0.427	0.467	**0.062**	0.211	0.072
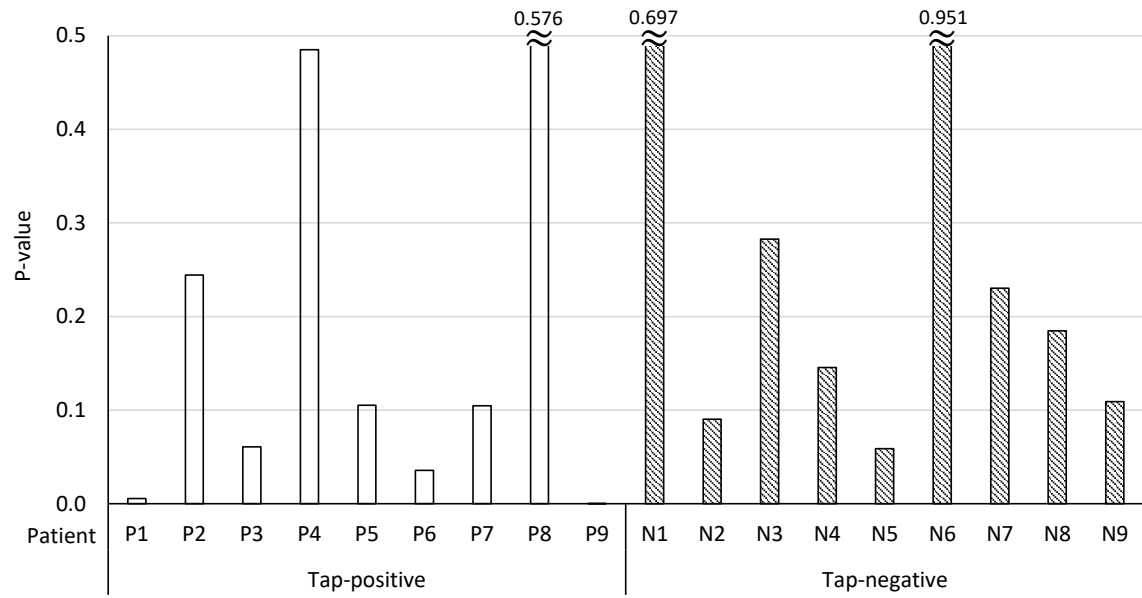
(a) Lateral sway.
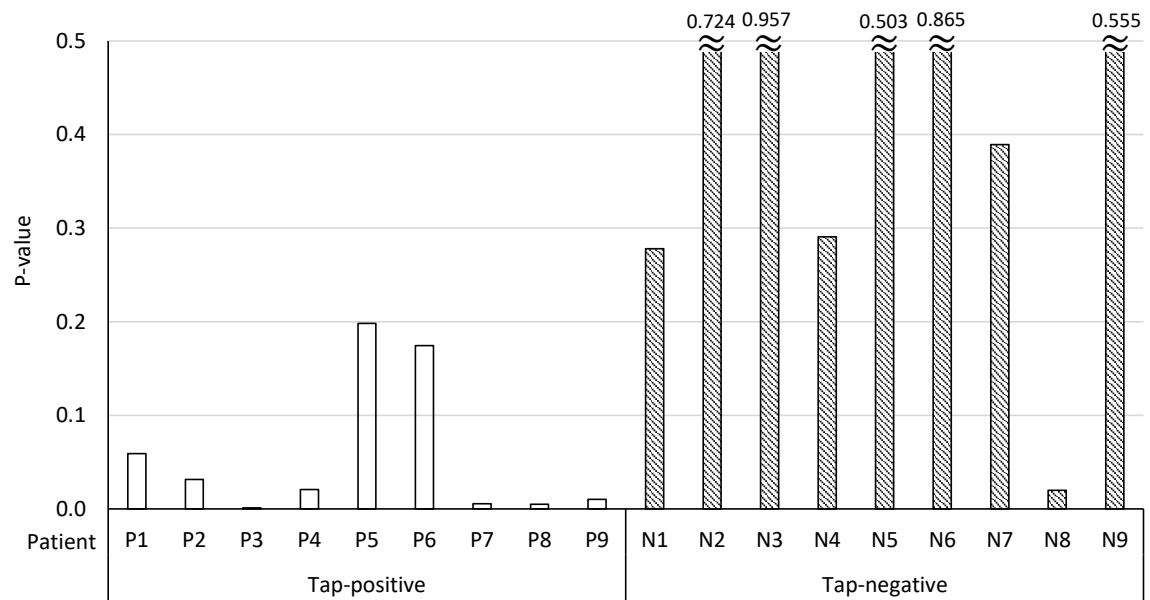
(b) Petit-pas gait.
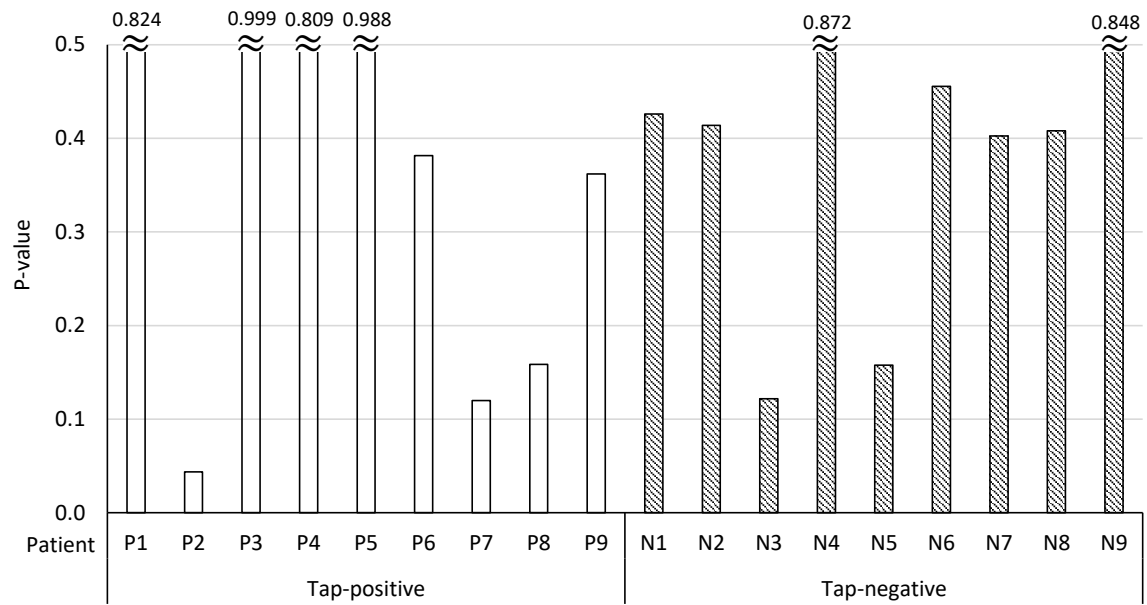
(c) Wide-base gait.
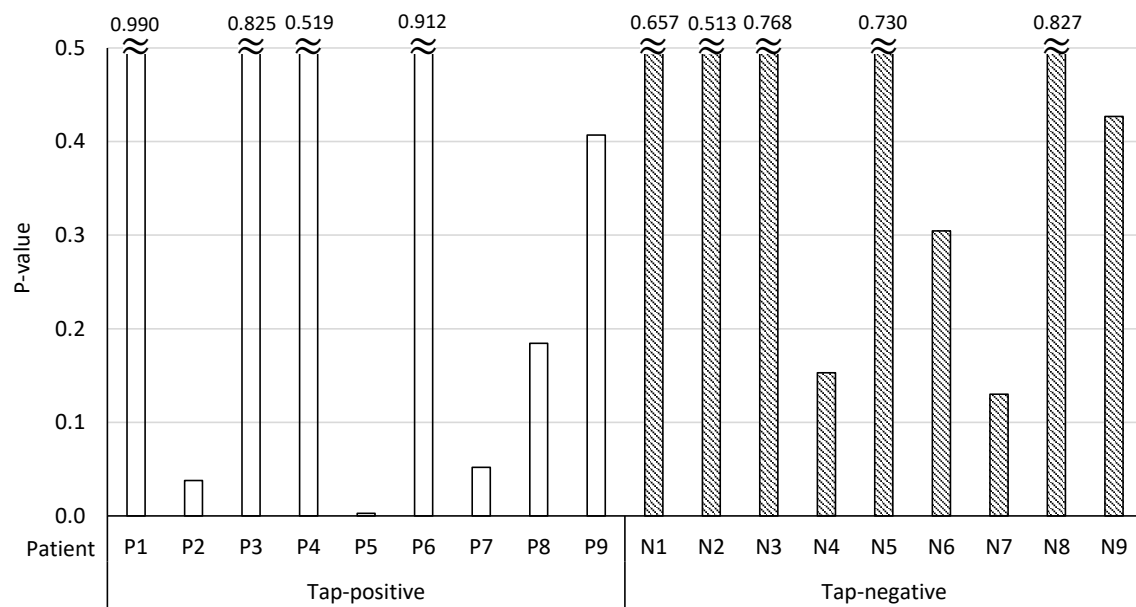
(d) Duck-footed walking.
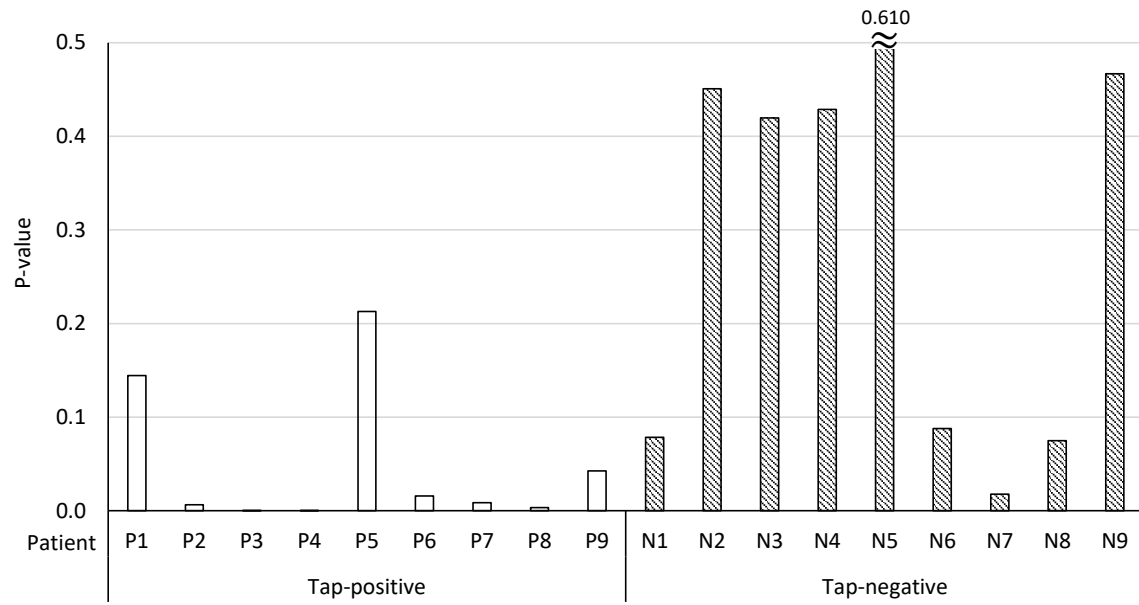
(e) Total time of TMW.
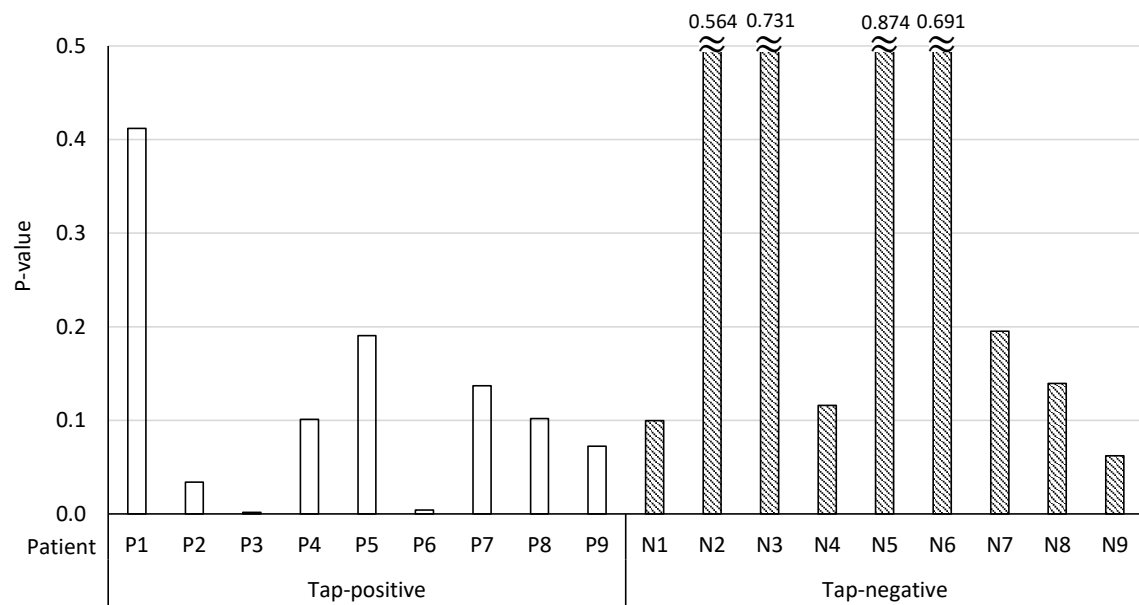
(f) Turning time in TMW.
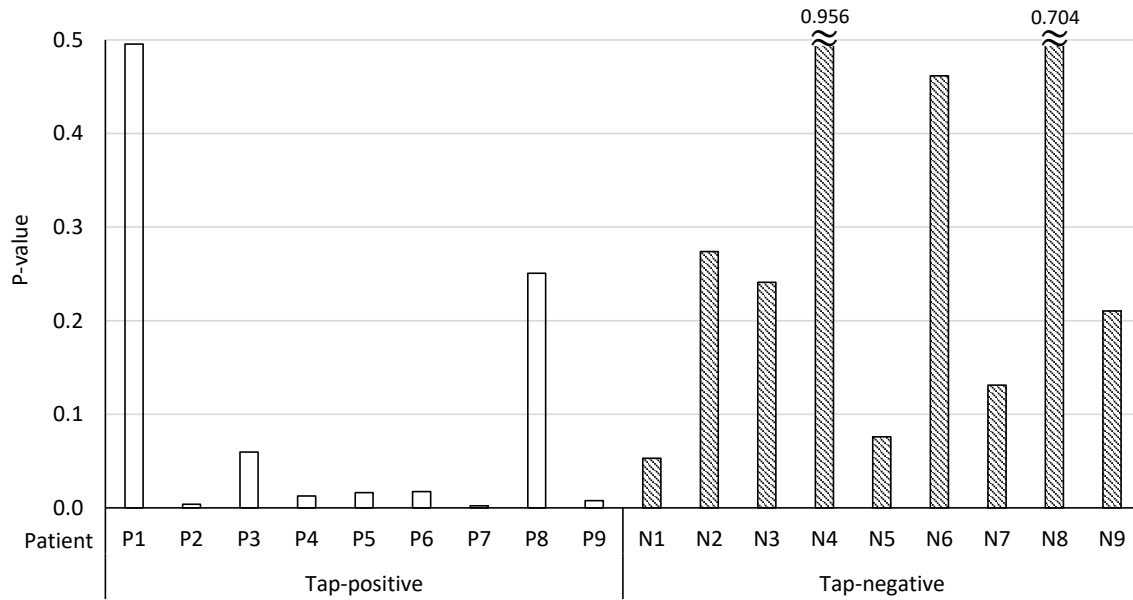
(g) Total time of TUG.
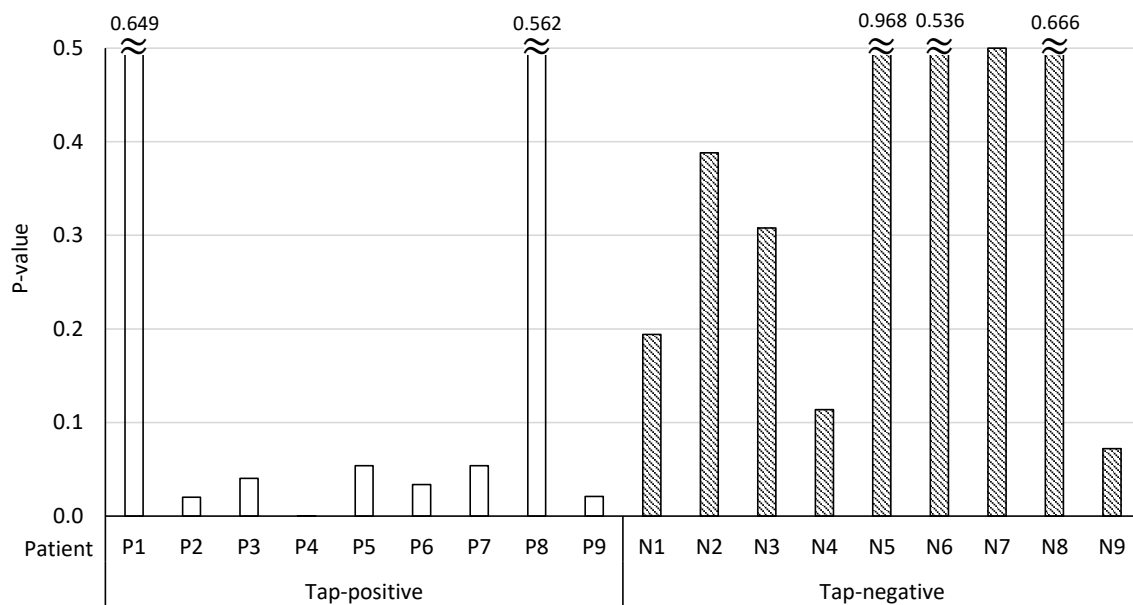
(h) Turning time in TUG.

Fig. 2.8 Comparison of p-values in each feature. Histograms for p-values above 0.5 are truncated and labeled with their exact values.

Table 2.3 The corresponding features of each patient's minimum p-value and the assessment values. LS: Lateral Sway, PG: Petit-pas Gait, WG: Wide-base Gait, DW: Duck-footed Walking, TMW/TUG total: Total time for TMW/TUG test, TMW/TUG turn: Turning time in TMW/TUG test.

Patient		P1	P2	P3	P4	P5	P6	P7	P8	P9
Feature (Unit)		LS (%)	TUG total (sec)	TMW total (sec)	TMW total (sec)	DW (ratio)	TMW turn (sec)	TUG total (sec)	TMW total (sec)	LS (%)
Pre-tap	Mean	1.20	944.8	745.6	607.5	2.56	67.5	533.5	1528.6	1.60
	SD	0.06	82.3	12.8	11.6	0.03	5.2	25.3	86.0	0.03
Post-tap	Mean	1.01	608.5	627.5	511.7	2.44	47.5	427.5	1138.8	1.41
	SD	0.04	58.3	7.8	6.0	0.02	3.2	23.1	80.7	0.03

Patient		N1	N2	N3	N4	N5	N6	N7	N8	N9
Feature (Unit)		TUG total (sec)	LS (%)	WG (%)	TUG turn (sec)	LS (%)	TMW total (sec)	TMW total (sec)	PG (m)	TMW turn (sec)
Pre-tap	Mean	379.4	1.93	1.56	71	1.23	474.4	965.6	0.66	126.5
	SD	33.4	0.06	0.52	2.3	0.04	12	52.6	0.01	12.9
Post-tap	Mean	315	1.83	0.73	60	1.16	455	827.5	0.71	101.5
	SD	14.3	0.03	0.3	5	0.02	5	12.5	0.02	5.5
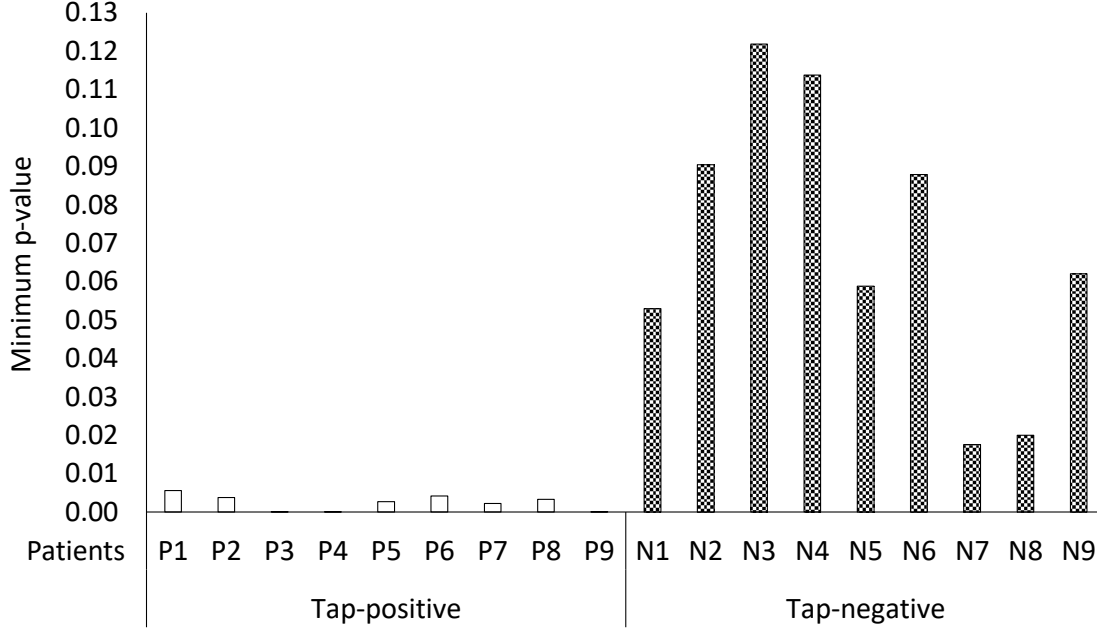
Fig. 2.9 The minimum p-values calculated for each patient. Although the minimum p-value for two tap-negative patients were <0.05, the minimum p-value for tap-positive patients were generally smaller than for tap-negative patients.

2.3.3 Noise tolerance of the judgement

To prove the reliability of the above judgment method, we add noise to the measured feature values and try to use them to judge the results of CSF tap test.

First, we calculated the mean of all measured values $\bar{v}_{ft}$ in all the 8 features (including gait features and timing features) , and randomly generated a Gaussian noise with $\mu = 0$, $\sigma = \alpha \bar{v}_{ft}$ where $\alpha \in (0, 0.3]$, added it to the measured values. Subsequently, we calculated the p-values of these noise-added measured values according to the proposed method. We then judged tap-positive and tap-negative labels by taking the median of the minimum p-value of all patients as the threshold, which makes the false-negative and false-positive rates equal.

The accuracy of the judgment is shown in Fig. 2.10. When the standard deviation of the added noise is within 10% of the $\bar{v}_{ft}$, the proposed method throws can obtain a high accuracy. However, as the noise gradually increases, the accuracy decreases.
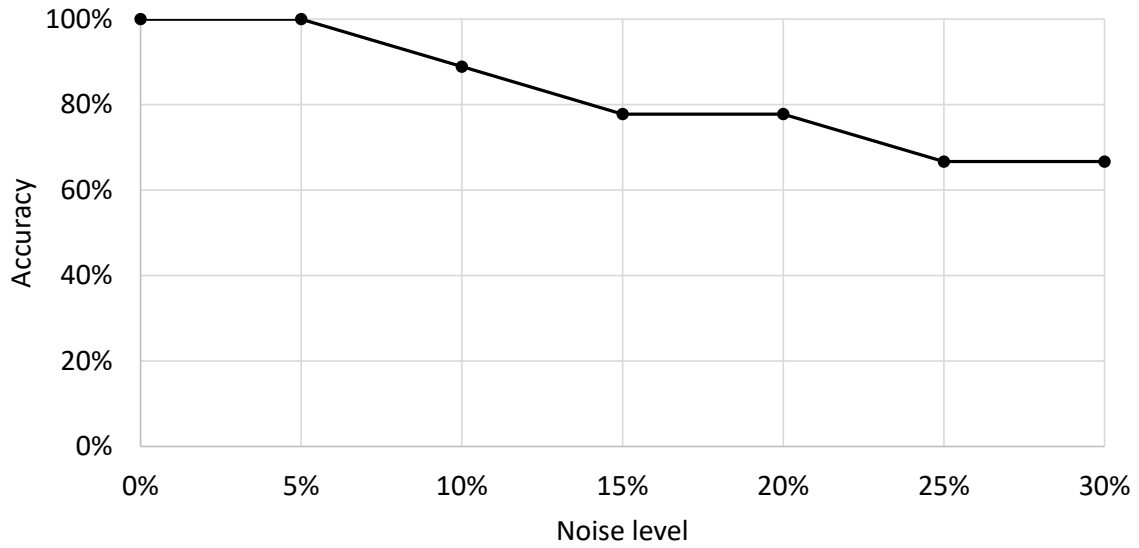
Fig. 2.10 Noise tolerance of the judgement. The x-axis is the value of α, i.e., the coefficient of the mean of feature values used as standard deviation of the noise.

2.4 Discussion

We assessed the four gait disturbance features of iNPH quantitatively, and confirmed that most of the assessed values of petit-pas gait, wide-base gait, and duck-footed walking were consistent with the physicians' judgment, while the assessment of lateral sway has some discrepancies. We also used the assessed value to improve the judgment for the CSF tap test, and concluded that judgment using silhouette-based gait assessment is more effective than existing quantitative judgment methods, which only use the recorded walking time.

Next, we will further discuss the challenges encountered in the development of our methodology, including some inconsistencies in lateral sway, and compare our method to that used by the physicians.

2.4.1 Lateral sway

According to the definition of lateral sway in the GSSR, a score of 2 indicates a fluctuation in the patient's heel strike position, a score of 1 indicates trunk sway, and a score of 0 indicates no sway or heel strike abnormality. Because our proposed method assesses trunk sway, it was expected to appropriately reflect the 0 and 1 scores. However, during our interviews with the physicians, the physicians stated that they assigned a score of 1 to patients who exhibited a trunk sway or those with a small degree of heel strike fluctuation, while they only assigned a score of 2 when frequent fluctuations in heel strike position were observed. Thus,

we asked the physicians to re-assess lateral sway using the strict definition of GSSR, which provided more consistency in the assessed values between our video-based analysis and the physicians' judgment. This is because physicians often associated lateral sway with a sense of balance during walking, and therefore, associated the frequency of fluctuations in heel strike with lateral sway scores. Thus, we suggest that the inconsistency of findings using our method compared with physicians' judgment is derived, at least in part, from differences in understanding of the definitions.

The ambiguity of the subjective judgment also influences the judgment of the physicians. From the interviews, we found that physicians often assign a better lateral sway score to patients exhibiting a lateral sway than that derived from iNPH. For example, physicians assigned a score of 0 to a patient walking fast during the test despite the presence of lateral sway because typical iNPH patients walk slowly. A lateral sway derived from fast walking was considered unimportant when rating the lateral sway for iNPH. Another patient was assigned a score of 0 despite the presence of a body sway because she had no petit-pas gait, wide-base gait, duck-footed walking, or other typical features of iNPH gait disturbance. Thus, the physicians felt that the sway was from another cause. As such, the judgment of the physicians often deviated from the definition of lateral sway.

2.4.2 Wide-base gait and duck-footed walking

As previously described, we assessed the wide-base gait and duck-footed walking features of all patients using a selected subsequence of two gait cycles in which the patients walked stably. We asked the physicians to review the selected gait segment, and they confirmed that the assigned values for these sequences were reasonable. Thus, we concluded that the proposed silhouette analysis-based method for wide-base gait and duck-footed walking was appropriate.

However, physicians assessed the entire gait sequence, which includes the stable gait sequence as well as the unstable gait sequence that often occurs when the patient turns or begins to walk. As such, our proposed method may miss gait disturbances if they are observed outside of our selected sequence. The assessed values of several patients with walking labels that were inconsistent with physician observations may relate to this subsequence selection problem. In future studies, we will extend our analysis to the entire gait sequence.

2.5   Conclusion

We propose a video-based method to improve gait disturbance assessment for iNPH. The proposed method can assess gait features independently of a physician's experience, and we validated that the method using the assessed value is more accurate than current qualitative methods for judging the CSF tap test. We assessed lateral sway, petit-pas gait, wide-base gait, and duck-footed walking, as well as the walking times, of the tests using gait silhouette analysis. We confirmed the effectiveness of our proposed method with experiments on CSF tap test judgment with nine tap-positive and nine tap-negative patients.

Some of the assessed values, however, remained partially inconsistent with the GSSR scores assigned by the physicians. One reason for this failure is that the proposed gait feature extraction and assessment method is different from the physicians' own customized criteria for GSSR. Future studies are required to improve our proposed method by considering the frequency of fluctuations in each patient's heel strike position and the multi-factorial dependency of the GSSR.

Another potential issue for clinical application of our method is that the manual work (e.g., silhouette modification, bounding box assigning, and step counting) is time consuming. To improve this, we are currently examining automatic extraction of the silhouette using a deep learning-based method (e.g., the RefineNet [117]). In future studies, we will also capture kinematic information of patients using depth sensors for more intuitive and accurate assessment of the features.

Chapter 3

Health indicator estimation by video-based gait analysis

3.1 Introduction

Body composition refers to the proportions of major components of the human body such as water, fat, and muscle. Common health indicators such as muscle mass and basal metabolic rate can be calculated from the body composition information. It is also known that the body composition changes and muscle mass decreases with aging [119]. The body composition information is therefore used for health management including prevention and mitigation of lifestyle-related diseases and for management of physical training [120–123].

Most existing commercially available body composition meters employ bioelectrical impedance [124, 125]. These methods involve sending a weak alternating current from the subject's hands and feet into the body, and analyzing the impedance to measure the body composition and estimate related health indicators. Although these methods can accurately measure the indicators, it requires relatively time-consuming steps: removing socks, cleaning hands and feet, and standing on the body composition meter, which may take several minutes in total. Furthermore, the same machine cannot measure multiple people simultaneously.

Because body composition is closely related to human motor function [126–128] as well as body shape [129–132], there is a possibility that we can estimate body composition and the related health indicators by analyzing the human motor function and body shape, which could be observed from gait videos. We therefore raise a research question on the possibility of health indicator estimation from gait videos as an alternative for the body composition

meter. One of the most promising approaches is visually observing the subject walking, i.e., assessing the gait. A gait video contains information pertaining to the subject's motor function and body shape. Thus, may be possible to estimate the health indicators using video-based gait analysis techniques. Moreover, because gait is the most fundamental mode of locomotion, we have more opportunities to observe it than other actions (e.g., jumping or doing bending exercises).

Video-based gait analysis is the subject of a large body of literature, including but not limited to gait-based estimation of gender, age, and health status [108, 133, 47, 134, 135]. Generally, the related techniques require just one gait period of a gait video (i.e., approximately one second) for analysis, thus suggesting that the health indicators can also be obtained much faster than with body composition meters.

The above-mentioned studies on video-based gait analysis typically employ machine learning techniques (e.g., classifiers for gender classification, regression techniques for age estimation) and the amount of training data is crucial to the success of these algorithms (e.g., pairs consisting of the gait video and ground-truth age label for age estimation), in particular, in this deep learning era.

It is, however, relatively difficult to collect a sufficiently large amount of training data, particularly if medical data is required. For example, studying the relationship between gait and a certain disease requires establishing collaboration with a hospital and obtaining informed consent from patients. Furthermore, it requires the effort of collecting gait videos of the patients and asking a medical professional to annotate ground-truth labels for each gait video. Moreover, the number of instances related to a specific disease is considerably limited compared to healthy subject data. As a result, the researcher must work with small-scale data, and hence tend to rely on handcrafted gait features and classical machine learning or even rule-based methods [136, 137, 134, 138]. This is in contrast to general computer vision tasks, such as object recognition and human pose estimation, which rely on deep learning frameworks [139, 140] with the help of publicly available and well-organized large-scale databases [141, 78].

Some large-scale publicly available databases exist for video-based gait analysis [69, 142, 68] containing over 10,000 subjects. Although these databases are not aimed at medical/healthcare-oriented applications, but at individual recognition [43, 143, 144]. We may be able to use their data to pre-train a backbone network for video-based gait analysis, and subsequently fine-tune the network for our task, i.e., health indicator estimation with a limited training data. This is a similar approach to computer vision research that fine-tunes networks pre-trained by standard datasets [78, 141] for specific tasks.

In this chapter, we therefore propose a deep learning framework to estimate health indicators via video-based gait analysis even with limited ground-truth training data. The contributions of this work are summarized as follows.

1. Health indicator estimation from a walking video.

To our knowledge, this is the first work to estimate health indicators based entirely on a video of the subject walking. This potentially increases the efficiency and ease of health indicator measurement compared to the conventional body composition meter.

2. Pre-training and fine-tuning strategies with gait primitives relevant to health indicators.

To enable the use of a deep learning framework even with the limited number of training data for gait-based health indicators, we use pre-training and fine-tuning strategies. Presently, there has not been extensive research on a pre-training strategy before fine-tuning in video-based gait analysis. Thus, we demonstrate that pre-training with gait primitives relevant to health indicators (e.g., arm swing, stride, the degree of stoop, and body width) is beneficial for the subsequent fine-tuning for the health indicator estimation task. This achieves an improvement in accuracy compared with training from scratch or pre-training using a conventional reconstruction task with an auto-encoder.

3.2 Related work

3.2.1 Video-based gait analysis

There is a rich body of literature on video-based gait analysis which we address in this subsection. Several comprehensive surveys on this topic provide further details for the interested reader [145–148].

The majority of studies on this topic pertain to gait recognition, i.e., person identification from walking videos [149–151, 61, 152–154]. Other popular related topics include gender classification [108] and age estimation [155].

Early studies in video-based gait analysis mainly focused on developing handcrafted gait features using both model-based approaches [150] and appearance-based approaches [156, 152, 157–159], and applied classical machine learning techniques (e.g., linear discriminant analysis and support vector machines) to the gait features for tasks such as person identification, gender classification, or age estimation. Publicly available gait video databases [61, 59, 160, 161] played an important role in training the machine learning models and providing a basis for performance evaluation, although database sizes were limited (i.e., at most a few hundred subjects).

41

Subsequently, deep learning frameworks were applied to video-based gait analysis [162, 163, 42, 164–168, 47, 169, 170, 49] in a similar fashion to other computer vision topics such as face recognition, object recognition, semantic segmentation, and human pose estimation. Because deep learning frameworks generally require substantial training data, the aforementioned studies typically used larger-scale gait databases [66, 68, 69, 142] (e.g., over 60,000 subjects in [68, 142]) for training and performance evaluation.

In contrast, it is difficult to collect large-scale gait data in the field of medical/healthcare-oriented gait analysis, both because the number of eligible subjects is limited, and annotation by a medical professional is required. Thus, studies in this field typically rely on small-scale gait databases. As a result, handcrafted features and classical machine learning or rule-based techniques are often employed because they can achieve suitable performance without substantial training data. For example, Liao et al. [136] and Ajay et al. [138] conducted gait analysis studies for specific diseases by using a small number of gait data including only 20 to 30 subjects. Aoki et al. [137] and Matsuura et al. [134] tackled the more general topic of cognitive function estimation from gait for hundreds of subjects, which is a larger dataset than those used in [136, 138], but still too small to apply a deep learning framework.

There are some studies that have applied deep learning frameworks in the field of medical gait analysis. For example, Zhang et al. [171] and Camps et al. [172] applied deep learning frameworks with only 18 and 21 subjects, respectively, with the help of data augmentation to cover the shortage of training data. The data augmentation, however, does not necessarily reflect subject diversity.

3.2.2   Fine-tuning in deep learning

Because deep learning frameworks usually require substantial training data that is not necessarily easy to collect, a researcher may decide to use a network pre-trained for general tasks on large-scale annotated databases (e.g., ImageNet [78], MS-COCO [141]), and fine-tune the network for a specific task such as object recognition [173, 174], action recognition [175], image retrieval [176], age estimation [46], person re-identification [177], and computer-aided diagnosis [178–181].

The networks pre-trained for general tasks with general image databases (e.g., for object recognition with ImageNet [78]), however, this data is not necessarily effective for fine-tuning for a specific task, particularly, when there are different domains between the general tasks and the specific tasks. To address this issue, some researchers prepare another training set for pre-training, whose domain/task is more relevant to the target task. For example, Yang et al. [46] pre-trained another network using facial images in addition to the network trained using a large-scale generic image database, and fine-tuned the two pre-trained networks

Table 3.1 Health indicators.

Indicator		Abbreviation	Unit	Measured value	
				Mean	SD
Weight		Weight	kg	60.5	12.3
Total Body Water		TBW	L	33.2	6.90
Protein		Protein	kg	8.93	1.91
Minerals		Minerals	kg	3.15	0.64
Soft Lean Mass		SLM	kg	42.7	8.92
Fat Free Mass		FFM	kg	45.3	9.44
Skeletal Muscle Mass		SMM	kg	24.9	5.76
Segmental	Mean of Arms	LBM_MA	kg	2.15	0.67
Lean Body	Mean of Legs	LBM_ML	kg	7.30	1.72
Mass	Trunk	LBM_T	kg	19.2	4.26
Limbs' Lean Body Mass		L_LBM	kg	18.9	4.71
Body Mass Index		BMI	kg/m^2	22.2	3.53
InBody Score		IBS	N/A	72.3	5.40
Waist-Hip Ratio		WHR	N/A	0.812	0.0545
Obesity Degree		OD	%	103.3	16.0
Body Fat Mass		BFM	kg	15.2	6.95
Percent Body Fat		PBF	%	24.7	8.46
Segmental	Mean of Arms	BFM_MA	kg	1.00	0.59
Body Fat	Mean of Legs	BFM_ML	kg	2.54	1.06
Mass	Trunk	BFM_T	kg	7.13	3.62
Visceral Fat Level		VFL	Level	5.78	3.41

for face-based age estimation. Wang et al. [182] pre-trained a network using pairs of real magnetic resonance images and manual segmentation annotations by radiologists, and fine-tuned the network for medical image segmentation. Gong et al. [183] used simulated positron emission tomography images to pre-train a network, and fine-tuned the network to improve the quality of the images.

Similarly, we pre-train our network using large-scale databases for gait analysis, add some layers to the network, and then fine-tune the network for the target task. Specifically, we explore suitable tasks for pre-training in this work to achieve suitable fine-tuning for the subsequent health indicator estimation.

3.3   Health indicators

We briefly introduce the health indicators used in our work. The indicators are summarized in Table 3.1. The health indicators are measured using a body composition meter InBody270 (InBody Japan Inc.), and the descriptions of the indicators are as follows:

- Weight, total body water (TBW), protein, minerals, body fat mass (BFM), soft lean mass (SLM), and skeletal muscle mass (SMM) signify the body weight and the mass of each body composition factor measured directly. The sum of TBW, protein, minerals, and BFM is equal to the weight. SLM is the mass of all the muscles, and SMM is the mass of muscles to move the skeleton, which is more directly related to the motor function. The percent body fat (PBF) is the ratio [%] of BFM to the body weight.

- Fat free mass (FFM) is calculated as the weight minus BFM. The basal metabolic rate (BMR), which is the energy a person consumes daily to sustain vital activities, can be calculated from the FFM using the Cunningham equation [184]:

$$v_{\mathrm{BMR}} = 370 + 21.6 \times v_{\mathrm{FFM}}.$$

  Because the right-hand side of this equation is calculated using only FFMs and constants, we omit the BMR estimation.

- Body mass index (BMI) is calculated using weight w and manually inputted height h by the following equation:

$$v_{\mathrm{BMI}} = w/h^2.$$

- InBody score (IBS) is an indicator originally defined by the manufacturer of the body composition meter, and is calculated by comparing the measured value and standard value of BFM and FFM. Obesity degree (OD) is the ratio of measured weight to standard weight. The standard weight is dependent on the height.

- Waist-hip ratio (WHR) is the estimated ratio of waist circumference and hip circumference. Visceral fat level (VFL) is a grading of visceral fat in the horizontal section of the abdomen around the navel. Both indicators are estimated by the body composition meter using the measured data.

- Segmental lean body mass (LBM)/body fat mass (BFM) refers to the muscle/body fat mass of each arm/leg and the trunk. The output of the body composition meter includes the value for each limb. Because we observe the subject walking from a side view in

this work, it is difficult to distinguish motions derived from left and right arms/legs. We therefore use an average of the LBM/BFM over the left and right arms/legs.

- Skeletal muscle index (SMI) evaluates the motor function and is calculated using the total muscle mass of the limbs $v_{\text{L\_LBM}}$ and the height h by the following equation:

$$v_{\text{SMI}} = v_{\text{L\_LBM}}/h^2.$$

This indicator is output by the body composition meter, but data acquisition errors for some subjects preclude us from including it in our work. We therefore substitute SMI with the summation of segmental LBM of the arms and legs to form the limb lean body mass (L_LBM).

3.4 Health indicator estimation using gait-primitive networks

3.4.1 Overview

Fig. 3.1 provides an overview of the proposed method to estimate the health indicators from a walking video. First, we design a deep neural network whose input is a gait template image (e.g., [156, 152, 157–159]) and the output is a gait primitive (e.g., arm swing, body width). We pre-train this network using publicly available large-scale gait databases (e.g., [68, 69, 142]) and refer to it as a gait primitive network. Secondly, we design a deep neural network where the input is the gait template image, and the output is the indicators. We achieve this by adding some layers to the gait primitive network and fine-tuning it with a limited number of training data. We describe the details in the following subsections.

3.4.2 Gait template image

Gait recognition researchers have proposed a variety of gait template images such as gait energy image (GEI) [156] a.k.a. averaged silhouette [185], frequency-domain features [152], gait flow image [158], chrono-gait image [151], and masked GEI [159]. Among them, GEI is the most frequently used method in video-based gait analysis because it is simple yet effective.

The GEI is obtained by averaging cropped silhouette images, whose size is 88 by 128 pixels, over one gait period, as shown in Fig. 3.1 (see inputs for the network). It encodes both static body shape (e.g., body part width with white pixels) and dynamic motion (e.g., arm
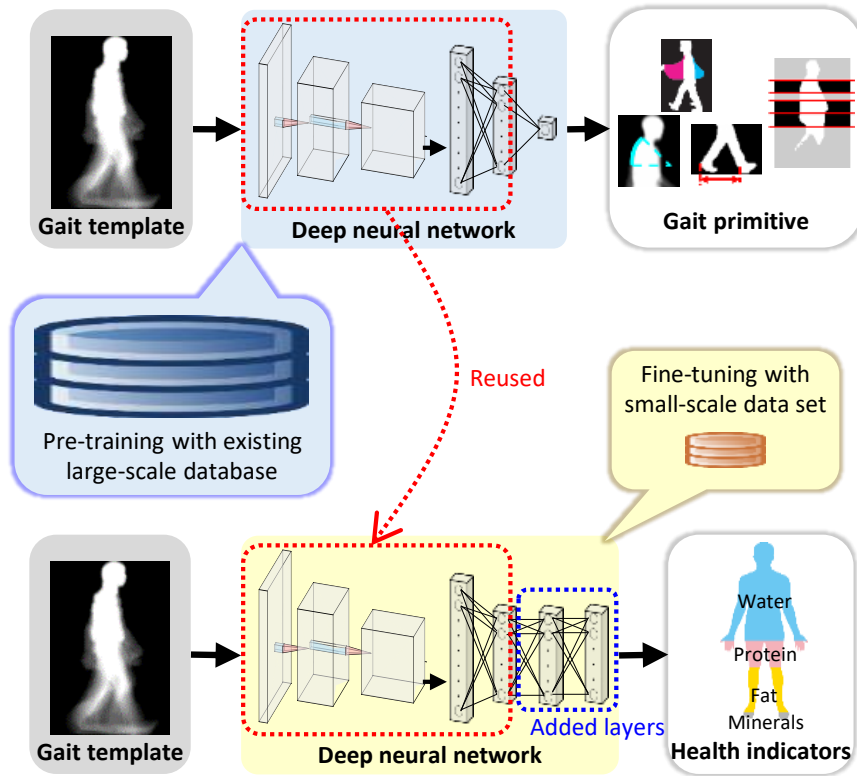
Fig. 3.1 Overview of the proposed method.

swing and leg movement with grey pixels). There are some literature reported correlations between muscle strength and walking ability, e.g., walking speed [41, 186, 187], which can be considered as correlations between muscle mass (i.e., SLM, SMM, etc.) and motion characteristics of gait; while others reported correlations between body fat mass (BFM) or percent body fat (PBF) and body shape parameters such as waist circumference [129–132]. In addition, the waist-hip ratio (WHR) itself is an indicator that describes body shape. This suggests that to estimate all of the indicators, the input data is required to contain both motion and shape information. The GEI is therefore suitable as the input for the network to estimate the health indicators as well as the gait primitives. The GEI is therefore suitable as the input for the network to estimate the health indicators as well as the gait primitives.

3.4.3 Gait primitives

The gait primitives refer to the fundamental components describing a subject's gait characteristics including motion characteristics, like arm swing, stride, and the degree of motion

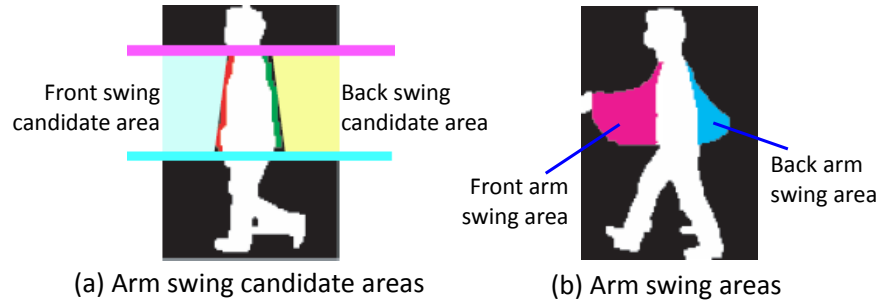(a) Arm swing candidate areas (b) Arm swing areas

Fig. 3.2 Measurement of forward and backward arm swing. Reprinted with permission from Springer [107], COPYRIGHT (2009).

symmetry; pose characteristics, like the degree of stoop and pitch; and shape characteristics, like height and body width, etc. To effectively pre-train the network, the gait primitives must be relevant to the motor function and the body shape, which are subsequently relevant to the health indicators. Furthermore, it is preferable to automatically extract the gait primitives without laborious manual annotation for a large-scale gait database.

Considering these criteria, we select four gait primitives: forward arm swing, backward arm swing, back straightness, and stride length, as proposed in [107]. These gait primitives are relevant to motion characteristics of walking, and are measurable automatically from a cropped silhouette sequence. Although relatively simple handcrafted methods are used to extract these gait primitives [107], their reliability has been demonstrated through an experience-based long-run exhibition of video-based gait analysis conducted in a science museum, where over 70,000 visitors joined the demonstration over approximately one year [188, 189].

We briefly explain each of the gait primitives in the following paragraphs and refer interested readers to [107] for further details.

- **Forward arm swing (FAS) and backward arm swing (BAS)**

  The front-end and back-end lines of the torso are extracted from a cropped silhouette at a single support phase and then forward/backward arm swing candidate regions are set. Areas swept by silhouettes in the forward/backward arm swing candidate region are counted as the degree of forward/backward arm swing, as shown in Fig. 3.2

- **Back straightness (BS)**

  The back-end line of the torso is extracted, and the slope of the back-end line is regarded as the back straightness. Specifically, if the line is more vertical, the back is straighter, and a less vertical result signifies stoop.

GEI   Static constituent   Regions for body
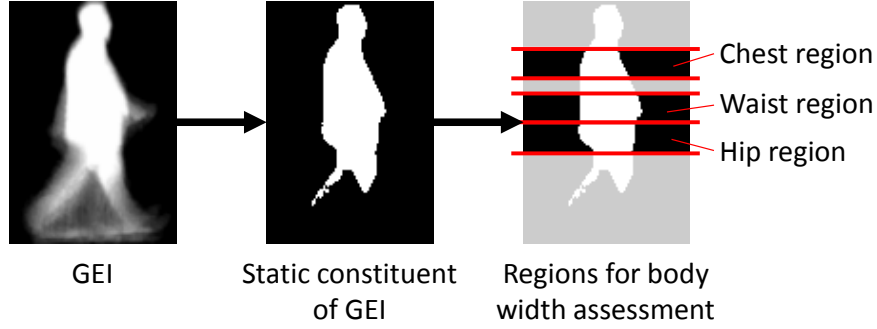of GEI   width assessment

Fig. 3.3 Measurement of body width. The average row width of the silhouette in each region is calculated as the body width.

- **Stride length (SL)**

  The gait period is obtained as the time shift that maximizes the auto-correlation of the cropped silhouette sequence along the temporal axis [152]. The total traveling distance during a walking sequence is computed by subtracting the start position from the end position, and the walking speed is then computed by dividing the total traveling distance by the elapsed time during the walking sequence. Finally, the stride length is regarded as the traveling distance over half of a gait period, and hence it is calculated as the product of the walking speed and half of the gait period.

Furthermore, since the above-mentioned four gait primitives only contain motion and pose characteristics and do not contain shape characteristics, we designed an automatically extracted shape-oriented primitive, i.e., body width shown in Fig. 3. We first extract an almost static foreground part of the GEI by thresholding with 75% of the maximum intensity and compute the average width of the static part for different height ranges: a chest region ($24 \leq y \leq 40$), a waist region ($48 \leq y \leq 64$), and a hip region ($64 \leq y \leq 80$). This measurement lacks support in the literature, but we have demonstrated its correlation with some health indicators by analyzing the data measured by the subjects (Fig. 3.4).

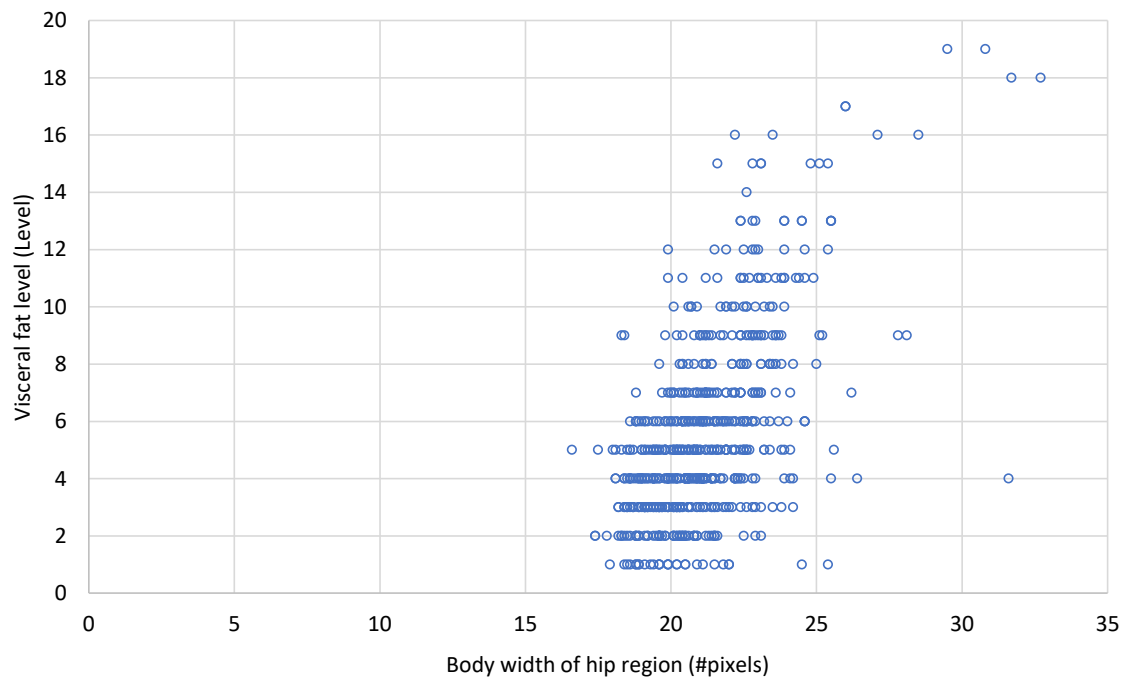We showed the distribution of the gait primitives in Fig. 3.5.

3.4.4 Pre-training the gait primitive network

We used GEINet [42] as the backbone for our gait primitive network. GEINet is a standard convolutional neural network (CNN) and is utilized in tasks such as person identification [42], gender classification, and age estimation [47].
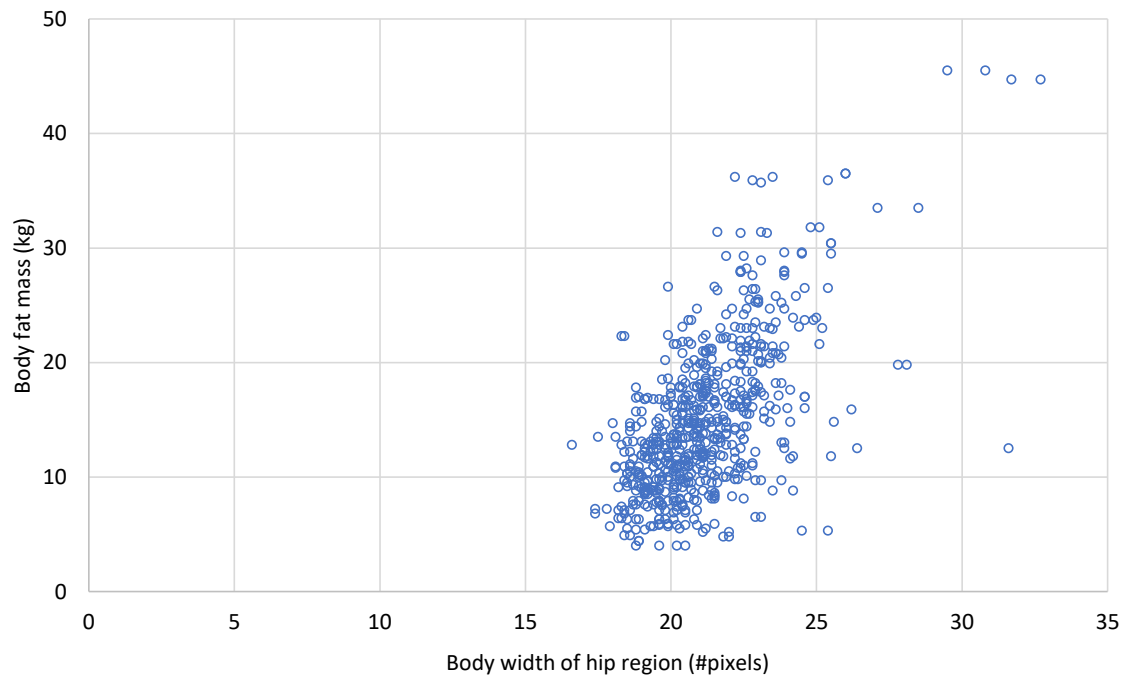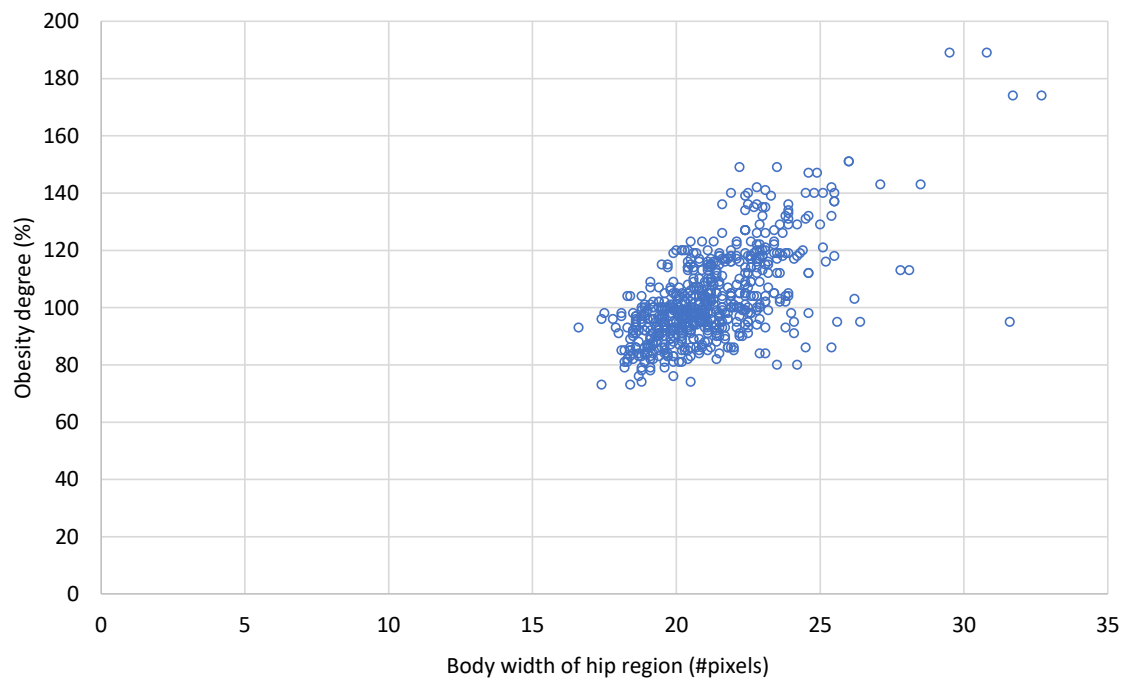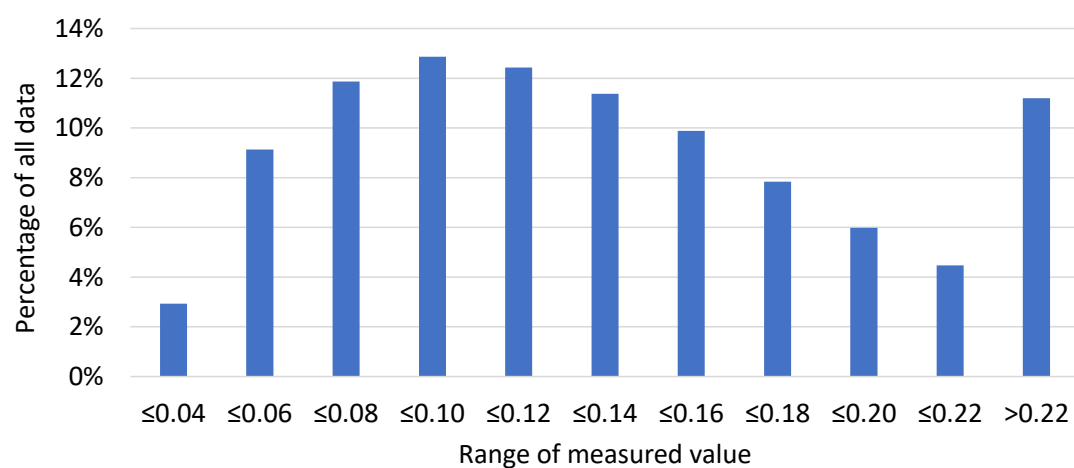
(a) vs. BFM.

(b) vs. VFL.

(c) vs. OD.

Fig. 3.4 Relationship between body width and some health indicators. It can be seen that these three health indicators and hip body width are positively correlated.

(a) Forward arm swing.
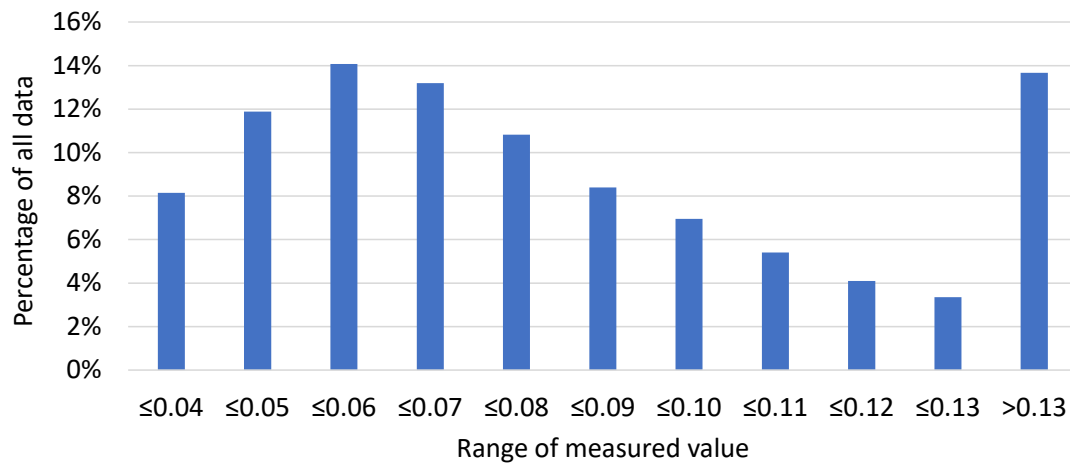
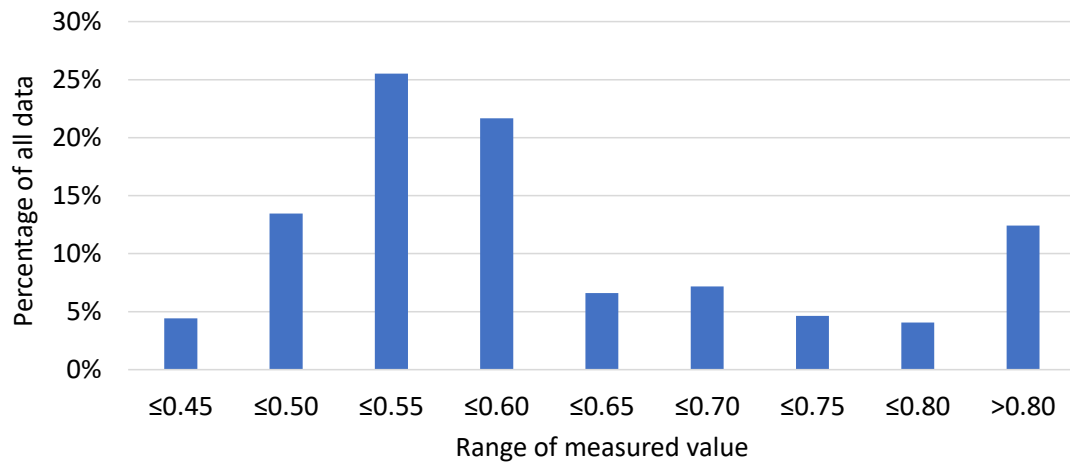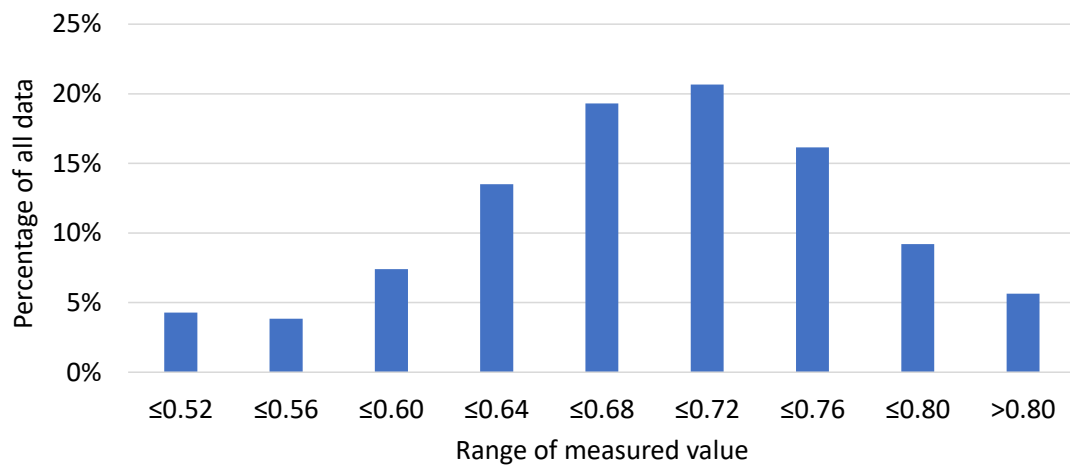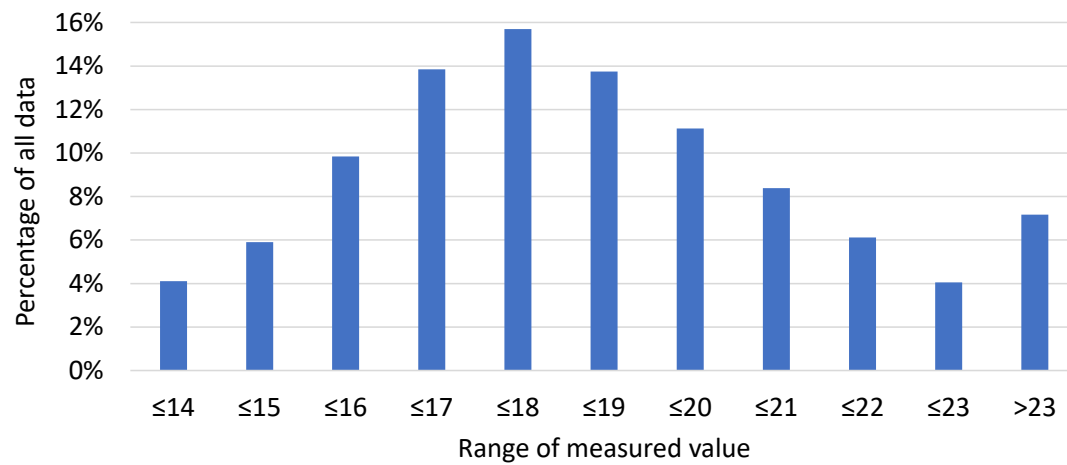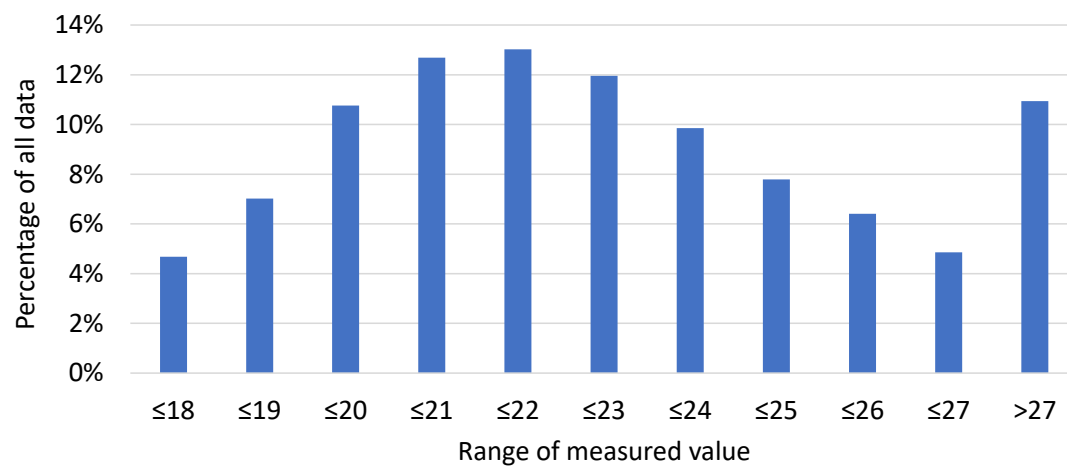(b) Backward arm swing.

(c) Back straightness.

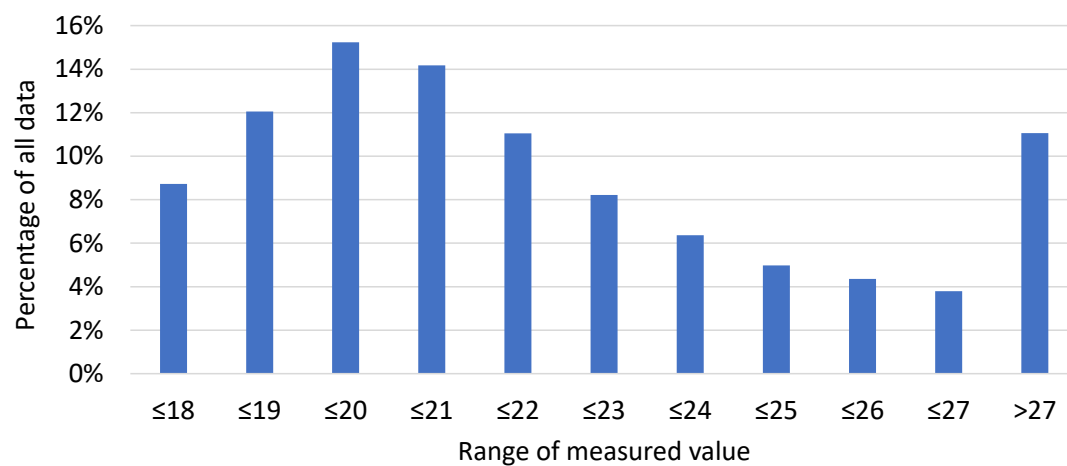(d) Stride length.

(e) Body width of chest region.

(f) Body width of waist region.

(g) Body width of hip region.

Fig. 3.5 Samples of gait primitive values for 1,000 subjects randomly selected.

The structure of the gait primitive network is almost the same as the original GEINet, as shown in Fig. 3.6, where the leading six layers are two sequential triplets of convolution, pooling, and normalization layers, followed by two fully connected layers. While the last layer outputs 1,024 units in the original GEINet for person identification, the last layer (fc2) in our gait primitive network outputs only one unit, because each gait primitive is defined as a scalar value (i.e., one dimension). The penultimate layer (fc1) outputs 1024 units. The other detailed configurations for the convolution and pooling layers are shown in Table 3.2, where the number and size of the kernels are slightly different from the original GEINet to suit our health indicator estimation task.
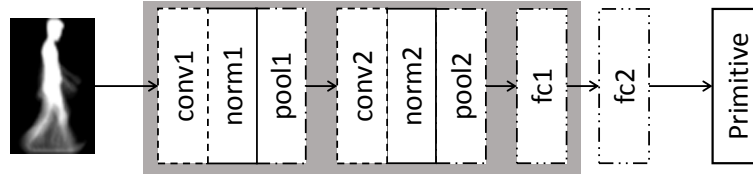
Fig. 3.6 Structure of the gait primitive network that takes a GEI as input and outputs the gait primitive. The abbreviations conv, norm, pool, and fc refer to convolutional layer, normalization layer, pooling layer, and fully connected layer, respectively. This convention is consistent throughout this paper. The layers with gray background rectangles are used in fine-tuning.

Table 3.2 Layer configurations for GEINet. Act. denotes the activation function.

Layer	#Kernels	Size/stride	Act.	Pooling
conv1	81	$5 \times 5 \times 1/1$	ReLU	
pool1		$3 \times 3/2$		Max pooling
conv2	45	$7 \times 7 \times 81/1$	ReLU	
pool2		$2 \times 2/2$		Max pooling

Once the gait primitive network structure is designed, we pre-train each gait primitive network using a training set composed of pairs of GEIs and each corresponding gait primitive so as to minimize a loss function, i.e., mean absolute error (MAE) between estimated gait primitives through the gait primitive network and gait primitives measured by the handcrafted method in [107] (i.e., a sort of the ground-truth gait primitives).

Fig. 3.7 Network structure for health indicator estimation in conjunction with two gait primitive networks. The outputs from the penultimate layers of the two gait primitive networks are concatenated and fed into a fully connected layer to estimate health indicators. The two sets of layers with gray background rectangles are from the pre-trained primitive networks.
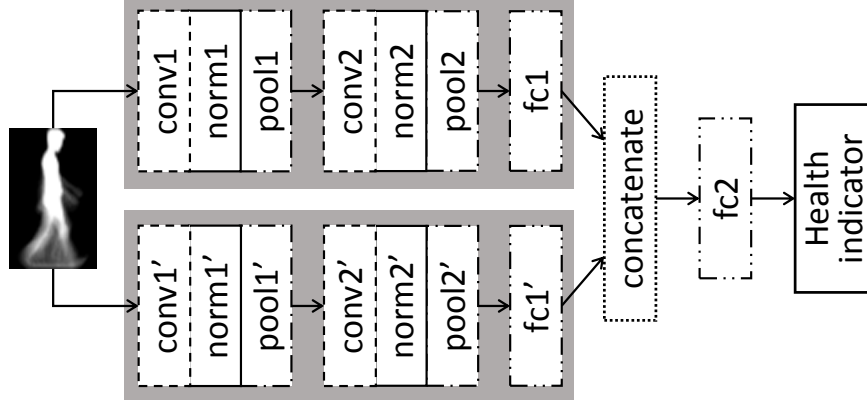
3.4.5 Fine-tuning the gait primitive network for health indicator estimation

We fine-tune the gait primitive networks for health indicator estimation. Because the output from the last layer (fc2) of the gait primitive network is just a single unit (i.e., one dimension), the output itself does not contain sufficient information for health indicator estimation. Instead, we utilize the output of the penultimate layer (fc1), which comprises 1024-dimensions, for health indicator estimation. The output from the penultimate layer is fed into a fully connected layer (fc2) to estimate one health indicator. Specifically, we fine-tune the layers from the gait primitive network (i.e., conv1, conv2, fc1) while we train the newly added fully connected layer fc2 from scratch. For this purpose, we first define the MAE for the health indicator as $L_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|$, where N is the number of training data, $\hat{y}_i$ and y_i are the estimated health indicator through the network (i.e., an output of fc2) and the ground-truth of the health indicator for the i-th sample. We then use the L_{MAE} as the loss function to minimize during network training.

We also use multiple gait primitive networks for further improvement, as shown in Fig. 3.7. We concatenate the outputs from the penultimate layers of the multiple primitive networks, and the concatenated feature is fed into a fully connected layer fc2. Thereafter, we train the network similarly to the aforementioned case of the single gait primitive network.

Some of the health indicators, such as the body weight and the mass of various body components, are closely related to the subject's height. However, the input to the network

(i.e., the GEI) lacks height information because we use height-normalized cropped silhouettes. The health indicators that are represented by mass in kilograms (e.g., weight and SLM) and volume in liters (e.g., TBW) are also normalized by the height, specifically, divided by the cubic height, and height-normalized health indicators are set as estimation targets. We require the subject's height information for this normalization, and hence we assume that the subject provides the height information or we automatically measure the subject's height from a captured image with a calibrated camera and a ground plane constraint [190]. Once we estimate the height-normalized health indicators, we obtain the original health indicators by un-normalizing them (i.e., multiplying by the cubic height).

3.5 Experiments

3.5.1 Dataset

We used OU-ISIR Gait Database, Large Population Dataset with Age (OULP-Age) [68] for pre-training the gait primitive networks. More specifically, we extracted a subset of 40,000 subjects out of 63,846 subjects.

We conducted experiments to collect the data used for training and evaluating health indicator estimation because there is no such publicly available database. We conducted the experiments twice in March 2019 and September 2019, and recruited a total of 332 subjects (167 females and 165 males). The ages of the subjects range from 8 to 71 years old, with a mean age of 34.6 years and a standard deviation of 15.6 years. We obtained informed consent from the subjects to use the data for research purposes.

We asked each subject to walk on a pre-defined course of approximately 8 meters at his/her natural walking speed in two sessions: in the morning and in the afternoon. We thus obtained 653 valid gait videos after removing those with unnatural walking actions (e.g., touching the head or hands on the waist while walking). A camera was set up approximately five meters away from the walking course to capture the walking subject from a side view. The captured walking video's frame size was 640 by 480 pixels and it contained 30 frames per second (fps) for 2.5 seconds (75 frames in total). We extracted silhouettes from the original walking video, cropped the height-normalized silhouettes, and then extracted the GEIs.

We also measured the health indicators of each subject using the InBody270 body composition meter (Fig. 3.8). Because the body composition meter requires the subject's height to measure the health indicators, we asked each subject his/her height in advance. Because of time limitations, each subject's health indicators were measured only in one

session, we thus associated the single health indicator measurement with both GEIs extracted in the two walking sessions for each subject.

3.5.2 Setup

We employed 20-fold cross-validation to fine-tune and evaluate performance. We randomly divide the entire dataset into 20 groups with approximately equal number of subjects. We use one group as the validation set, and the rest as the training set in the fine-tuning. That is, 315 or 316 subjects were used as the training set and 16 or 17 subjects were used as the validation set.

We evaluated the accuracy using the relative errors of the health indicators, defined as the ratio of the MAE of the estimated health indicator to the mean of the measured (ground-truth) values. This scheme enables us to evaluate the accuracy despite the varying scales of the different health indicators.

In the training stage, the size of mini-batches was 128 and the initial learning rate was set to 0.001. The number of epochs was set to 200 and 250 for pre-training and fine-tuning, respectively.

3.5.3 Benchmarks

Because this is the first work to address health indicator estimation from gait, there are no benchmarks for this specific task. We therefore prepare a suite of benchmarks for this purpose.

Firstly, to verify the effectiveness of our pre-training and fine-tuning strategy with the gait primitive networks, we trained the entire network for health indicator estimation from scratch, denoted as DL (scratch). Note that this is equivalent to the proposed method without pre-training.

Secondly, to verify the effectiveness of the gait primitives for pre-training, we pre-trained the network using an auto-encoder (see Fig. 3.9) to minimize the reconstruction errors of the GEIs. This network contains knowledge related only to the GEI and does not have any relevance to motor function. The pre-trained encoder part of the network, whose structure is identical to that of the gait primitive network except for the absence of the last layer (fc2), was used for fine-tuning in the same way as the proposed method, and is denoted as DL (auto-encoder + fine-tuning).

In addition, we fine-tune the ResNet50 model pre-trained using ImageNet to compare the primitive network with the general trained model commonly used for fine-tuning methods. The trained ResNet50 model can also be considered as a network trained with primitives

(a) The body composition meter.

(b) Measuring health indicators with InBody270.

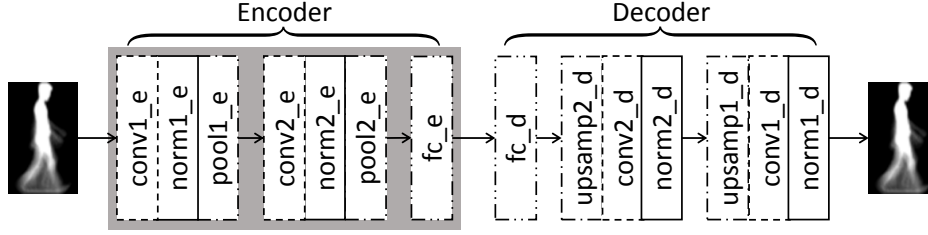Fig. 3.8 InBody270 and the measurement.

Fig. 3.9 Network structure of the auto-encoder. The abbreviation upsamp indicates an upsampling layer. We used the encoder part (layers with gray background rectangles) for fine-tuning.

that are larger in scale yet further removed in domain from the task, hence it is also possible to compare the impact on pre-training performance between sample size and primitive-task relevance. For regression of the health indicators, we replaced the top layer of the trained ResNet50 model with two serial full-connected layers with the same configuration as fc1 and fc2 in the proposed approach (see Section 3.4.4). Moreover, due to the insufficient scale of the dataset, we froze the early layers and trained only the last convolutional block and the newly added fully-connected layers.

Finally, we employed support vector regression (SVR) as a family of classical machine learning. The classical machine learning usually works better even with a smaller number of training data than deep learning-based approaches, while its capability is usually inferior to the deep learning-based approaches. We therefore investigated the trade-off between the number of training data and the capability by comparing the proposed method with SVR.

3.5.4 Ablation studies on gait primitives

We conducted ablation studies to determine the best use schema of the primitive networks.

For the gait primitives with motion and pose characteristics, we report the relative errors for particular health indicators when each single gait primitive, each combination of two gait primitives, and all four gait primitives are used in Fig. 3.10. We see that the estimation error increases significantly when we pre-train the networks with the step length gait primitive, and did not show significant variation when using the other gait primitives. Furthermore, pre-training using all the gait primitives does not outperform dual network combinations. Therefore, considering cost-effectiveness, we selected the dual network combination yielding the highest accuracy — forward/backward arm swing — as the best schema.

Fig. 3.10 Comparison of the relative errors among the gait primitives with motion/pose characteristics. Hatched colored bars indicate a single gait primitive, while solid-colored bars indicate combinations of two gait primitives. The solid black bar indicates all four gait primitives.
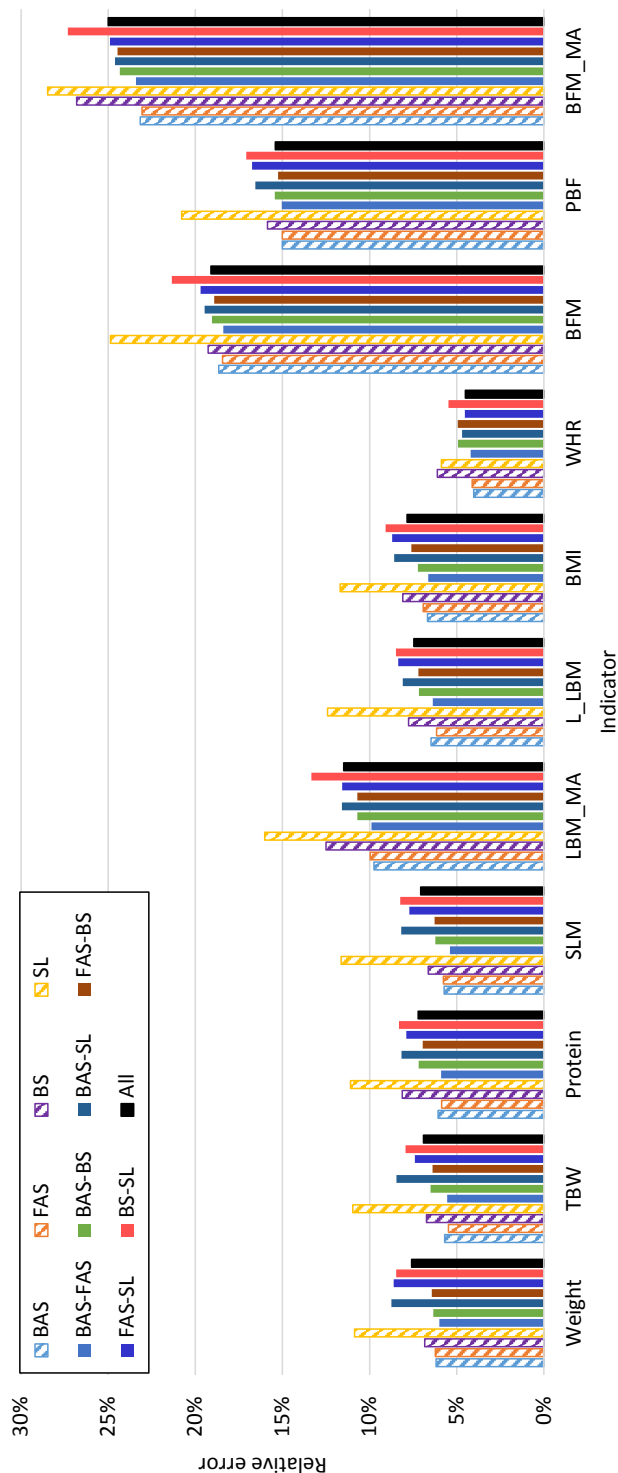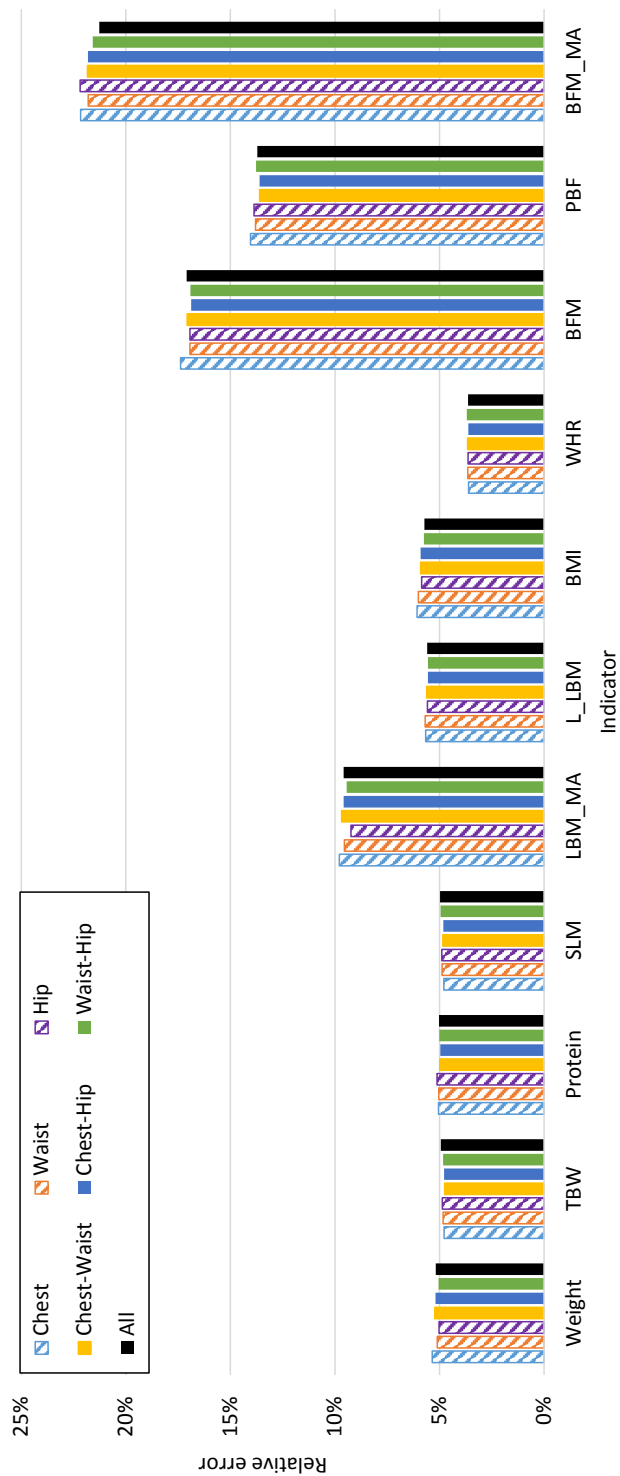
Fig. 3.11 Comparison of the relative errors among the gait primitives with shape characteristics. Same as Fig. 3.10, hatched colored bars indicate a single gait primitive, while solid-colored bars indicate combinations of two gait primitives. The solid black bar indicates all three gait primitives.

(a) [Proposed] BAS-FAS.

(b) SL.

(c) BAS-SL.

(d) All motion/shape primitives.

(e) vs. SL.

(f) vs. BAS-SL.
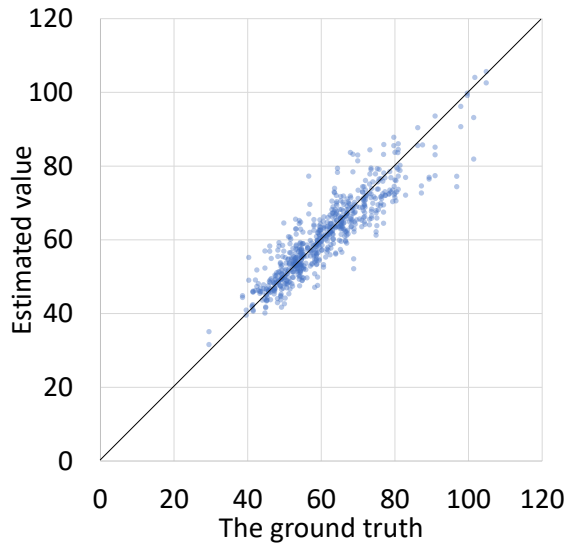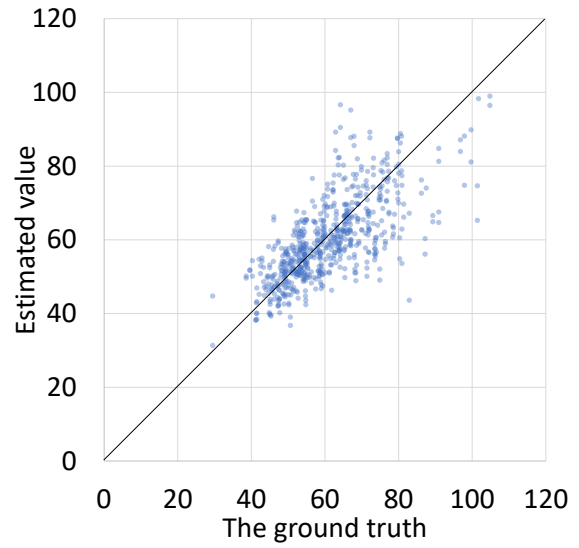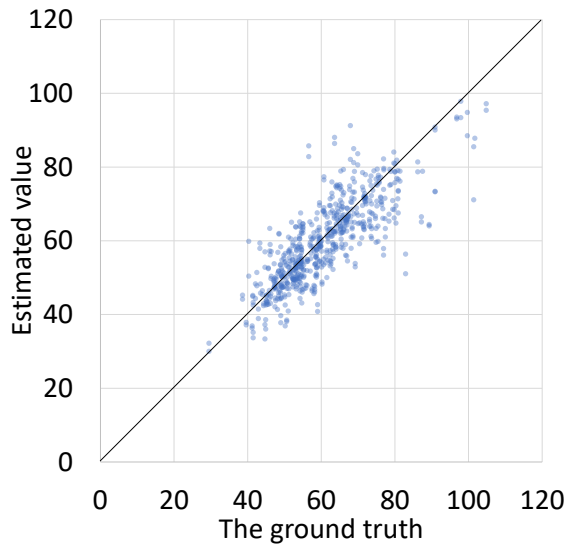
(g) vs. all motion/pose primitives.

Fig. 3.12 Scatter plots of estimated values vs. the ground truth (i.e., (a)–(d)) and errors of proposed method vs. other primitive networks (i.e., (e)–(g)), in weight estimation using motion/pose primitives. In the former plots, the solid black line indicates the equality between the estimated value and the ground truth, and in the latter ones, the smaller the horizontal distribution of the points than their vertical distribution, the better the accuracy of the proposed method than the compared method, and the same for the following.

(a) [Proposed] BAS-FAS.

(b) SL.

(c) BAS-SL.

(d) All motion/shape primitives.

(e) vs. SL.

(f) vs. BAS-SL.
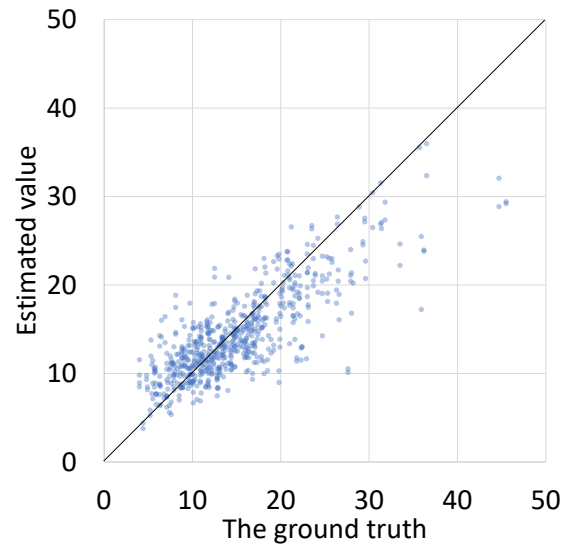
(g) vs. all motion/pose primitives.
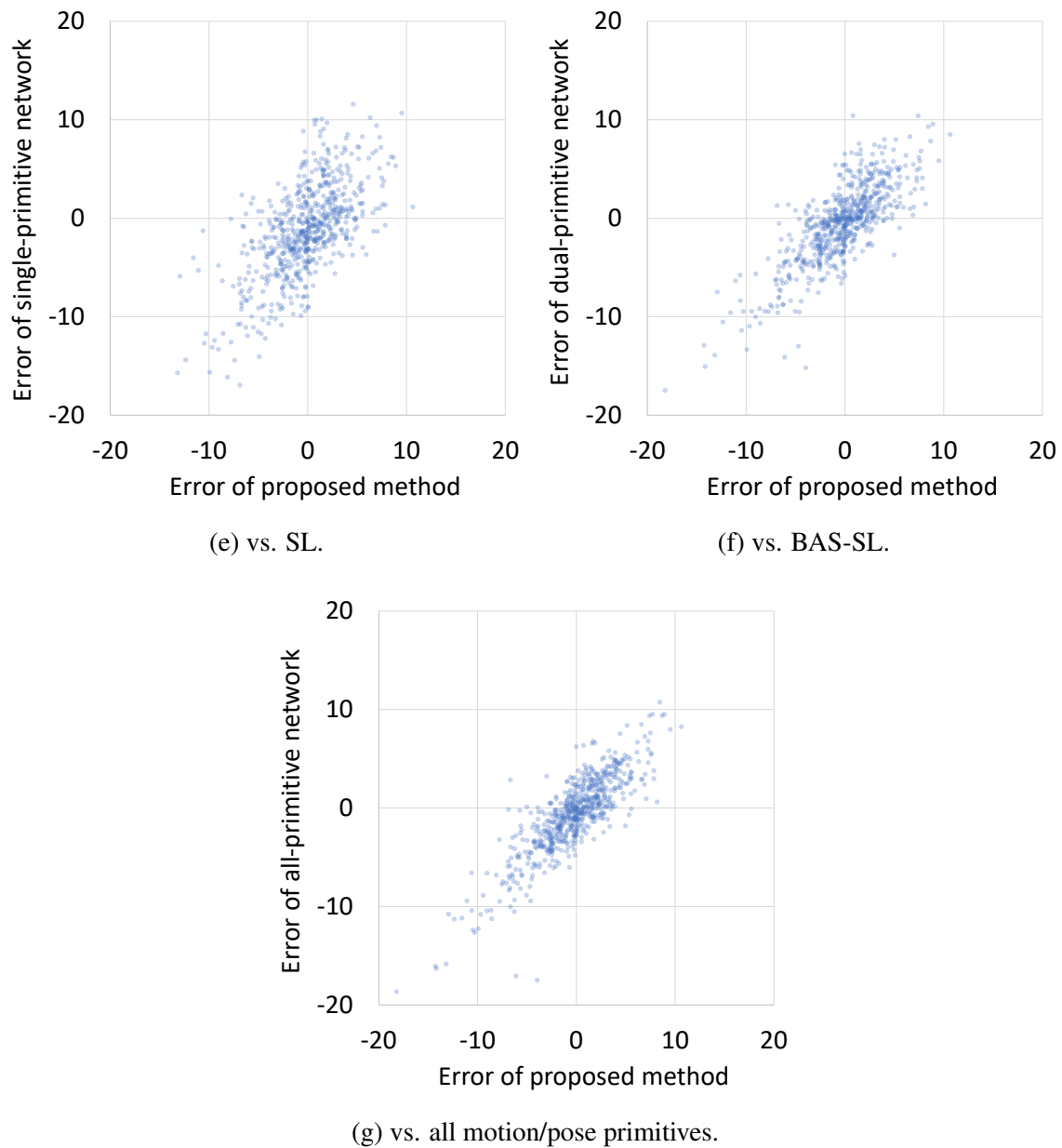
Fig. 3.13 Scatter plots of estimated values vs. the ground truth and errors of proposed method vs. other primitive networks, in BFM estimation using motion/pose primitives.

(a) [Proposed] BAS-FAS.

(b) SL.
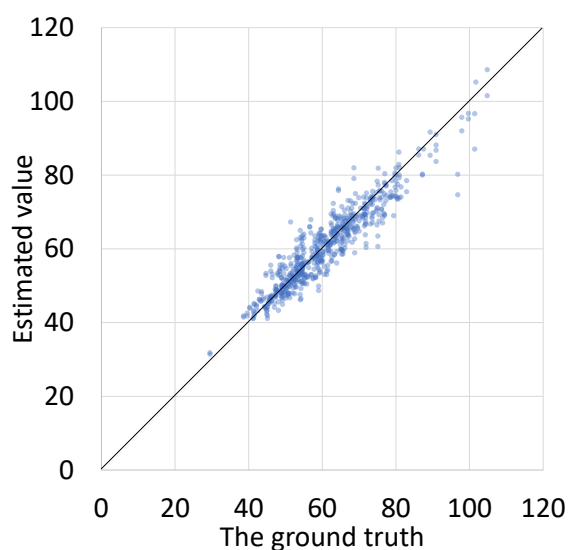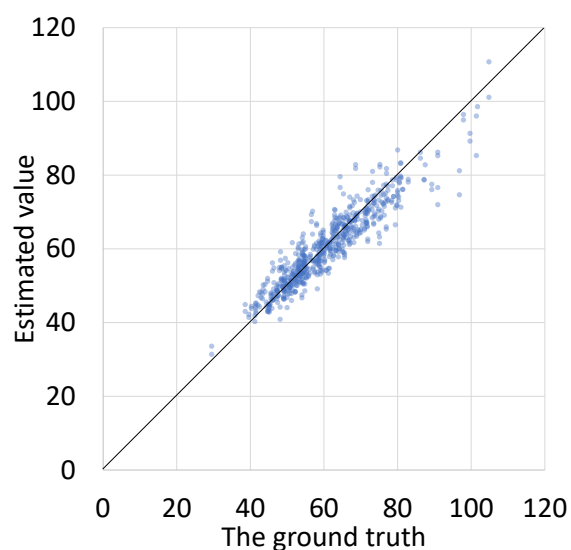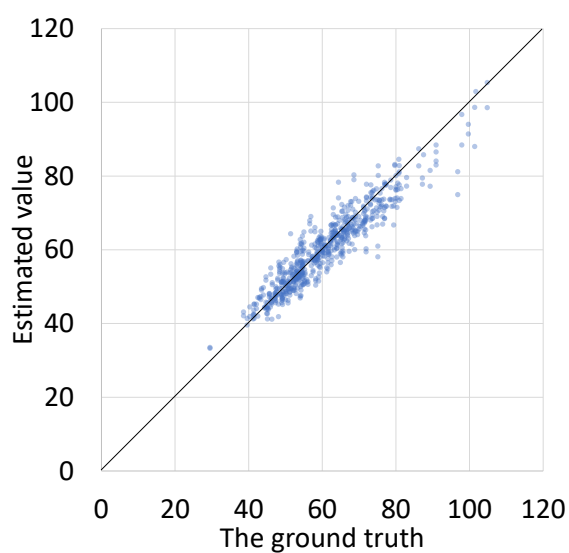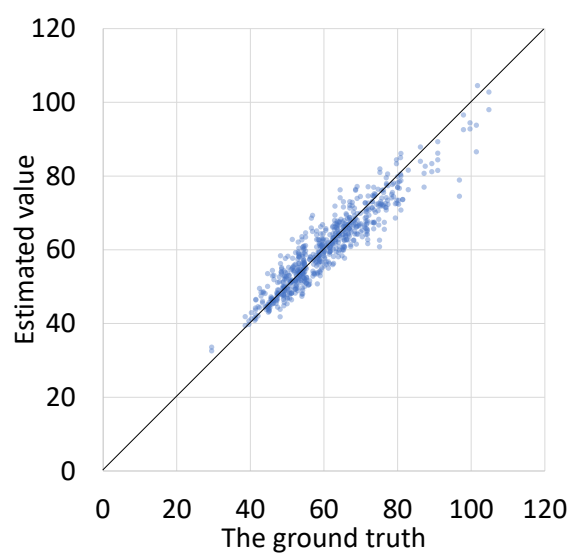
(c) BAS-SL.

(d) All shape primitives.
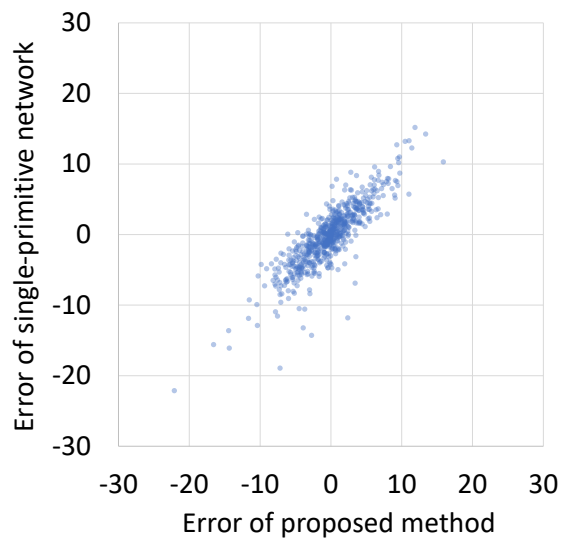
(e) vs. SL.
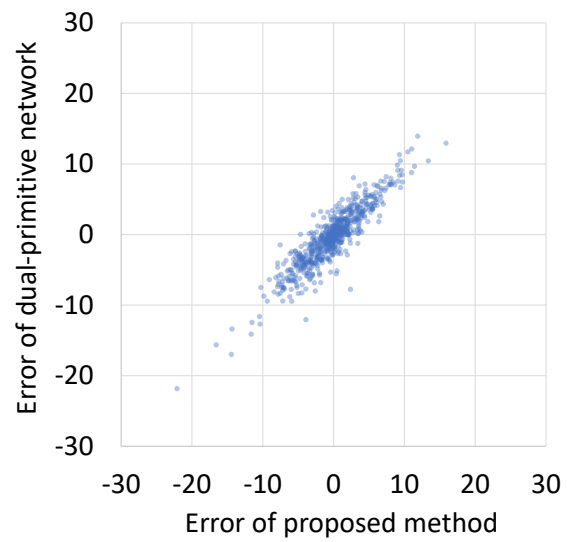
(f) vs. BAS-SL.
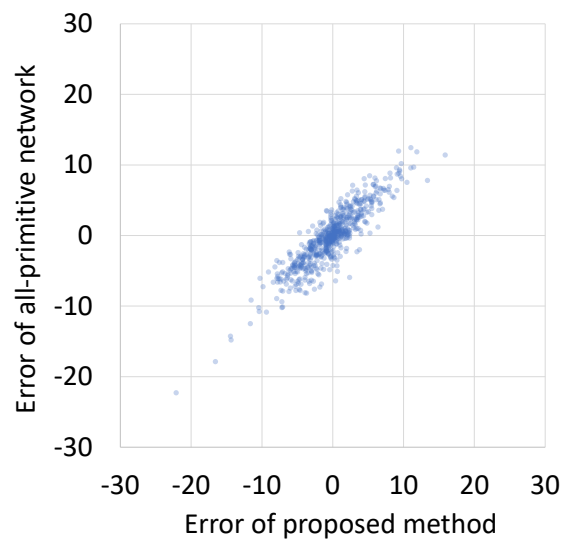
(g) vs. all shape primitives.

Fig. 3.14 Scatter plots of estimated values vs. the ground truth and errors of proposed method vs. other primitive networks, in weight estimation using shape primitives.
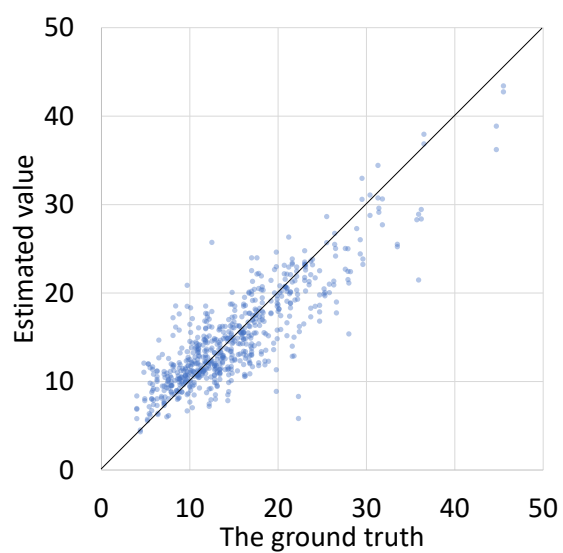
(a) [Proposed] BAS-FAS.

(b) SL.
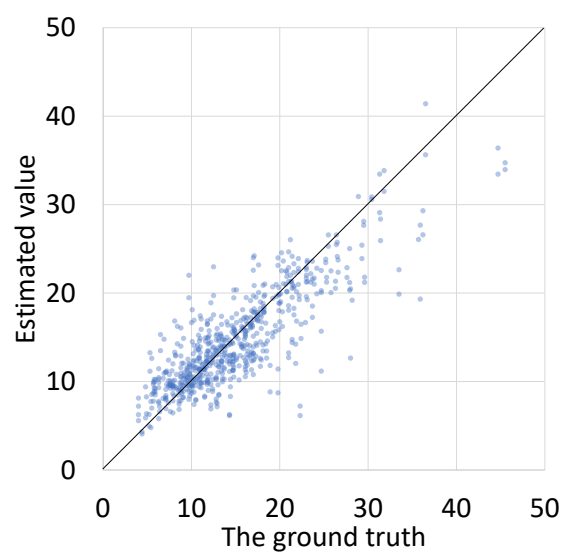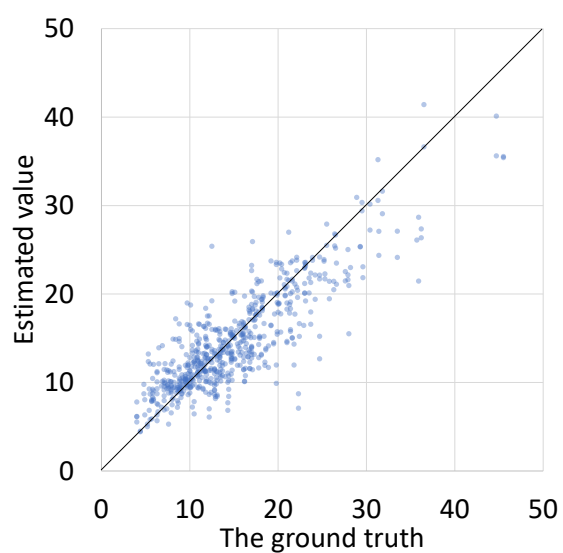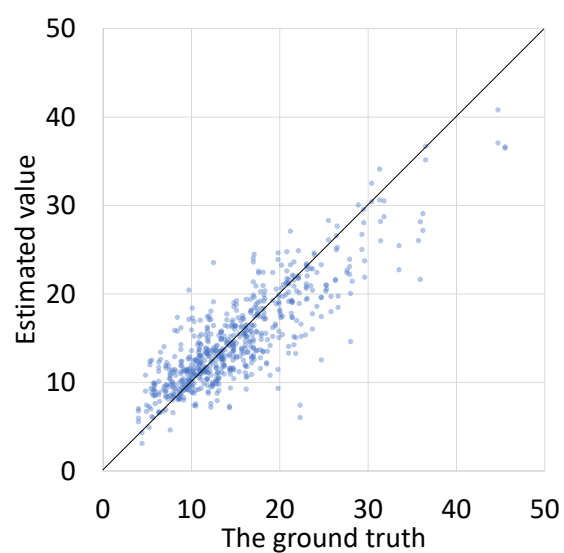
(c) BAS-SL.

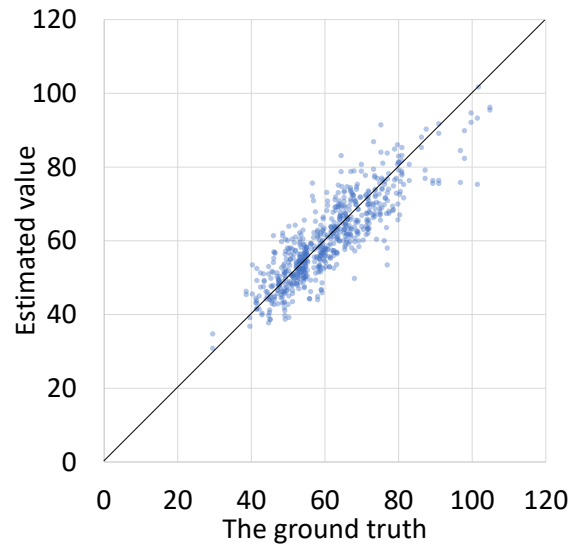(d) All shape primitives.
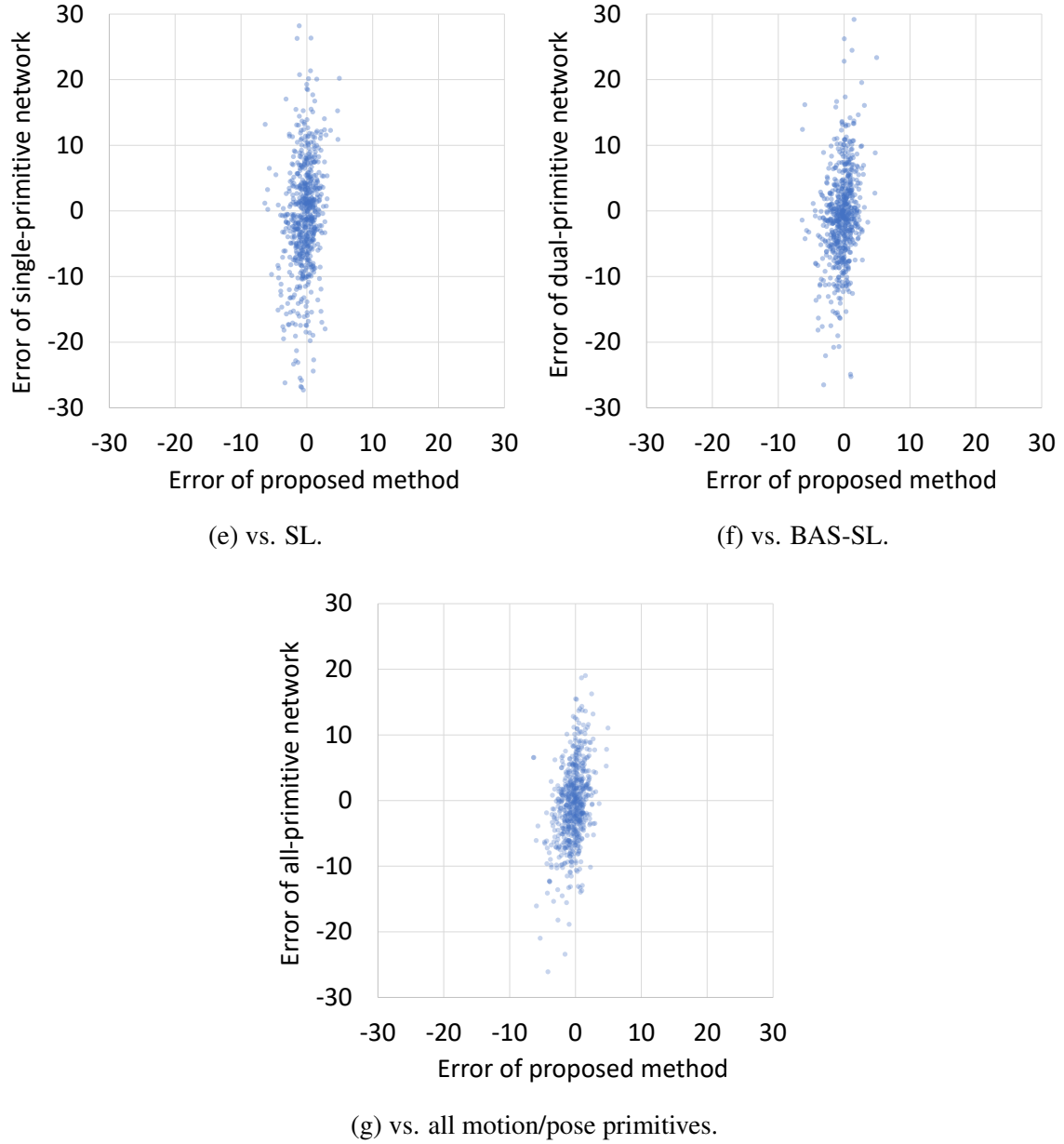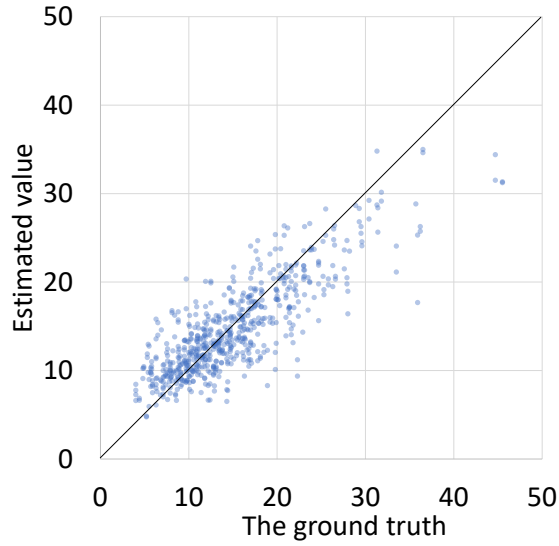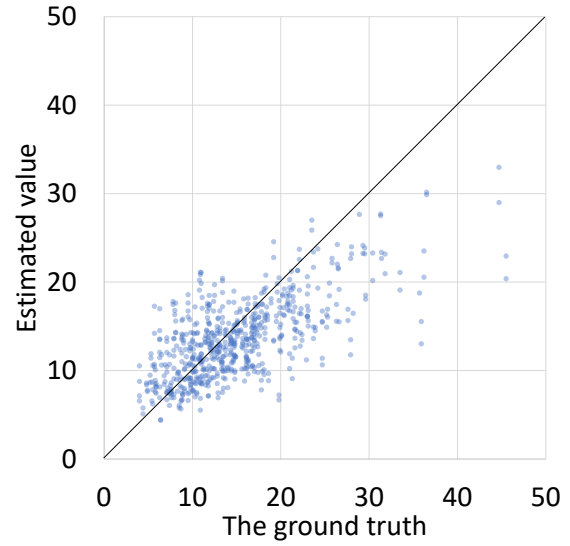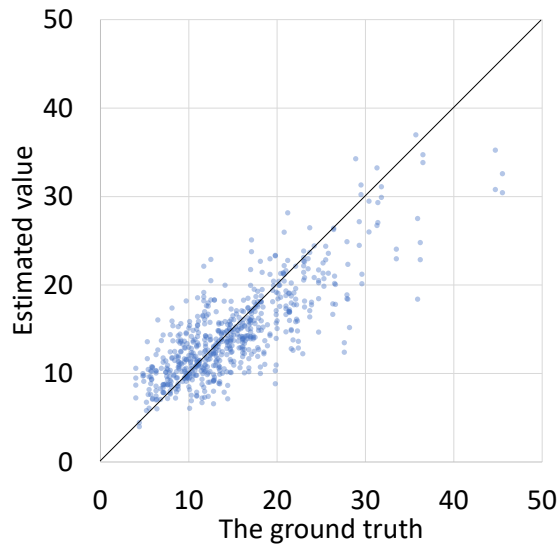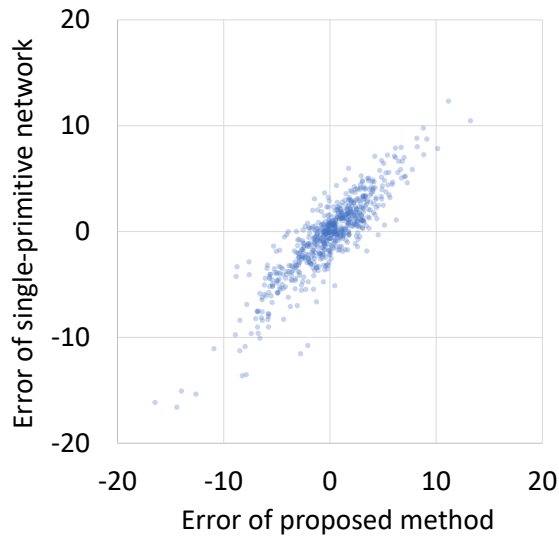
(e) vs. SL.
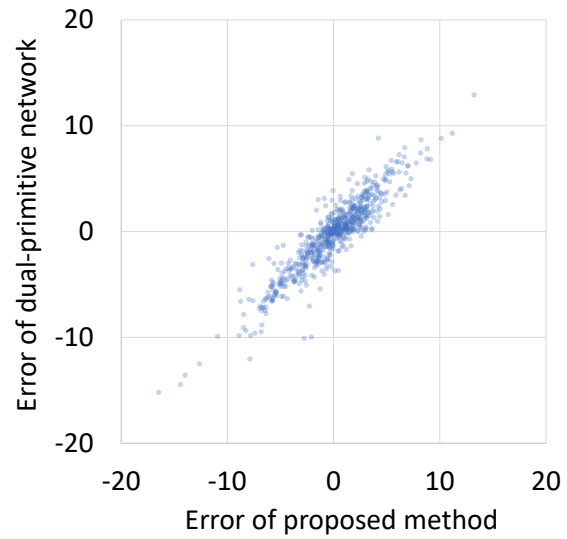
(f) vs. BAS-SL.
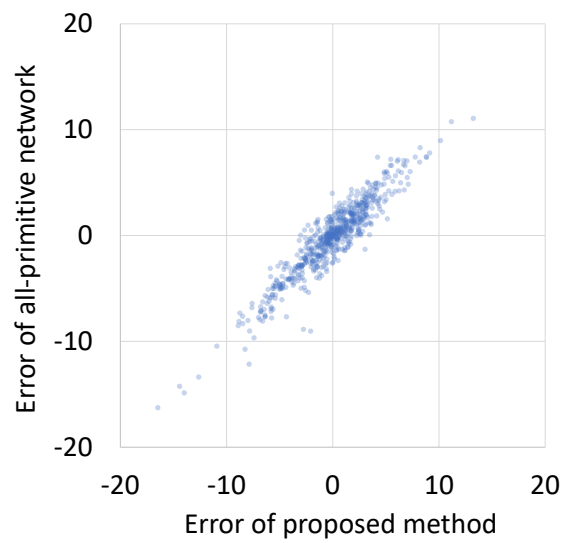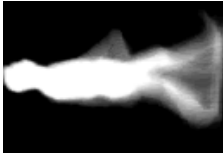
(g) vs. all shape primitives.

Fig. 3.15 Scatter plots of estimated values vs. the ground truth and errors of proposed method vs. other primitive networks, in BFM estimation using shape primitives.

Table 3.3 MAEs and relative errors of the proposed method of using gait primitive with motion/pose characteristics and shape characteristics, respectively.

Indicator	Mean MAE		Relative error	
	Motion/pose	Shape	Motion/pose	Shape
Weight	3.63	3.04	6.00%	5.02%
TBW	1.84	1.62	5.54%	4.86%
Protein	0.53	0.46	5.90%	5.12%
Minerals	0.18	0.16	5.69%	4.95%
SLM	2.30	2.09	5.39%	4.89%
FFM	2.45	2.20	5.41%	4.86%
SMM	1.53	1.41	6.15%	5.63%
LBM_MA	0.21	0.20	9.89%	9.24%
LBM_ML	0.43	0.37	5.94%	5.02%
LBM_T	1.15	1.06	6.00%	5.50%
L_LBM	1.20	1.05	6.37%	5.58%
BMI	1.47	1.30	6.64%	5.85%
IBS	3.89	3.34	5.38%	4.62%
WHR	0.03	0.03	4.21%	3.63%
OD	6.51	5.84	6.30%	5.65%
BFM	2.80	2.57	18.40%	16.93%
PBF	3.72	3.43	15.03%	13.87%
BFM_MA	0.23	0.22	23.41%	22.18%
BFM_ML	0.43	0.41	16.95%	16.12%
BFM_T	1.51	1.37	21.14%	19.14%
VFL	1.51	1.40	26.14%	24.27%

Table 3.4 Examples of GEI and corresponding estimated values and ground truth. M/P: Estimated values using motion/pose primitive network, S: Estimated values using shape primitive network, GT: Ground truth.

Subject	Subject A			Subject B			Subject C			Subject D			Subject E		
GEI															
Values	M/P	S	GT	M/P	S	GT	M/P	S	GT	M/P	S	GT	M/P	S	GT
Weight	62.6	56.4	59.8	50.8	50.5	49.3	57.1	56.0	54.7	48.1	48.1	52	71.4	68.3	73.5
TBW	38.4	38.1	36.4	26.0	25.9	27.5	31.7	31.8	32.4	23.1	22.0	22.4	36.6	33.3	33.3
Protein	10.4	10.3	9.7	6.83	6.88	7.3	8.47	8.61	8.8	6.12	5.80	6	9.25	8.93	8.9
Minerals	3.65	3.68	3.39	2.56	2.56	2.65	2.87	2.89	2.93	2.21	2.14	2.17	3.23	3.21	3.44
SLM	50.1	49.4	46.7	33.5	33.8	35.3	42.3	41.9	41.7	30.4	27.7	28.8	45.5	43.2	42.8
FFM	53.3	52.2	49.5	36.1	35.6	37.5	43.8	44.1	44.1	33.0	30.7	30.6	47.8	44.9	45.6
SMM	28.1	28.4	27.3	18.8	19.1	20.2	24.0	24.3	24.3	16.7	16.2	16.1	24.8	24.8	24.8
LBM_MA	2.50	2.56	2.31	1.46	1.44	1.5	2.00	2.10	2.2	1.29	1.24	1.3	2.01	1.98	2.13
LBM_ML	8.94	9.15	8.99	5.82	5.77	5.94	7.10	7.14	7.05	5.02	4.62	4.56	6.69	6.95	7.04
LBM_T	23.6	21.7	20.9	14.6	14.8	15	18.8	19.2	19.5	13.4	13.0	13.4	19.2	18.3	19.2
L_LBM	23.1	22.5	22.6	14.3	14.2	14.9	19.1	18.3	18.5	11.7	11.9	11.7	18.2	18.0	18.3
BMI	17.9	17.9	18.1	21.5	20.4	20	21.7	21.1	20.1	24.3	24.2	25.1	25.2	23.0	27.3
WHR	0.787	0.803	0.8	0.775	0.78	0.76	0.802	0.800	0.83	0.820	0.838	0.79	0.784	0.814	0.87
OD	80.7	79.3	82	101	97.6	95	98.5	94.7	91	122	124	119	122	109	127
BFM	7.62	7.04	10.3	14.0	16.0	11.8	11.2	12.8	10.6	17.1	18.8	21.4	19.8	21.5	27.9
VFL	2.11	2.26	4	7.00	7.13	4	4.07	5.04	4	8.81	10.4	10	6.34	7.22	13
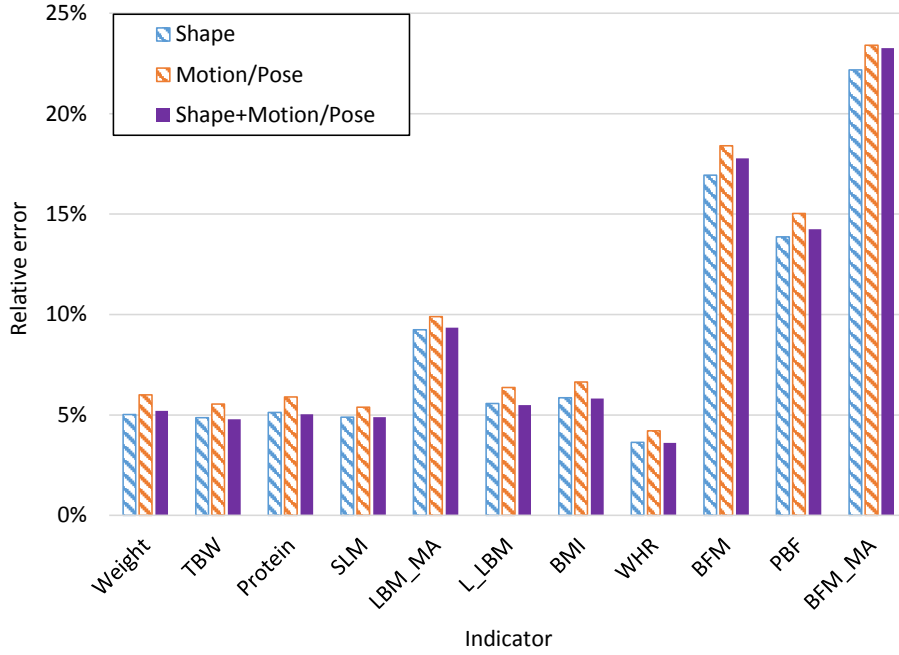
Fig. 3.16 Comparison of the respective best schemes for gait primitives with motion/pose characteristics and shape characteristics, and their combinations. The "shape" one is body wide in the hip region and the "motion/pose" one is the combination of forward and backward arm swing.

On the other hand, for the gait primitives with shape characteristics, we similarly report the relative errors when pre-training using single, dual or all the three regional body widths as gait primitives in Fig. 3.11. We found that combining multiple primitive networks resulted in marginal improvement in accuracy when using gait primitives with shape characteristics. Therefore, we chose the body width of the hip region, with the lowest error among the single primitives, as the best schema.

We compare the original MAEs and relative errors of the two best schemas in Table 3.3, and showed some examples of GEI with the estimated values and ground truth in Table 3.4. The comparison shows that pretraining using gait primitives with shape characteristics results in smaller errors in all indicators than using those with motion characteristics. We also tried to combine the primitive networks of the two best schemas (see Fig. 3.16), but no significant improvement in accuracy was obtained. Therefore, we decided to directly use the two best schemas, i.e., forward/backward arm swing and hip region body width, as the proposed method.

As examples of results, we plotted the estimated values against the ground truth, and compared the estimation errors of the proposed method with those of other primitive networks,

for two of the health indicators, i.e., weight and BFM, in Fig. 3.12, Fig. 3.13, Fig. 3.14, and Fig. 3.15.

On the other hand, we see that the proposed method with the body shape-oriented gait primitive can relatively accurately estimate the health indicators other than body fat-relevant ones (i.e., the relative errors are approximately 5%). However, the indicators related to body fat show higher errors (e.g., from 15 to 30% error) than the other indicators. We hypothesize that this may be related to the subjects' clothing — participants in the experiment conducted in March generally wore thicker clothing, and their body shape would have been obscured and not properly represented by the silhouette. Furthermore, since the body composition meter needs to ensure that the sum of TBW, protein, minerals and BFM output is exactly equal to the weight, it is possible that the measured weight minus the first three values will be used as the BFM measurement; and the deviation of the measured weight of the subjects will also vary due to the different clothing. Further validation is therefore still needed in future.

3.5.5   Comparison with benchmarks

We evaluated the relative errors among the benchmarks as shown in Fig. 3.17. SVR (i.e., a classical machine learning technique) underperforms compared to deep learning from scratch for body fat-related indicators, but these two methods are comparable for the other indicators. Among the deep learning methods, deep learning frameworks trained from scratch show higher estimation errors for most of the indicators as compared to the method that is fine-tuned with auto-encoder or ResNet50. Furthermore, the proposed deep learning framework fine-tuned with the gait primitive network yields the highest accuracy. This indicates the effectiveness of the proposed pre-training and fine-tuning strategy compared to training from scratch or fine-tuning a topic-independent general trained model when using a limited number of training samples for the target task. Also, we observe that it is essential to pre-train the network with features relevant to the target task (i.e., gait primitives for health indicator estimation) rather than pre-training in a general way (i.e., using an auto-encoder). Without proper pre-training the deep learning methods lose their effectiveness and may become inferior to classical machine learning methods (e.g., while SVR yields less than 10% error for TBW, pre-training with auto-encoder yields over 10% error).
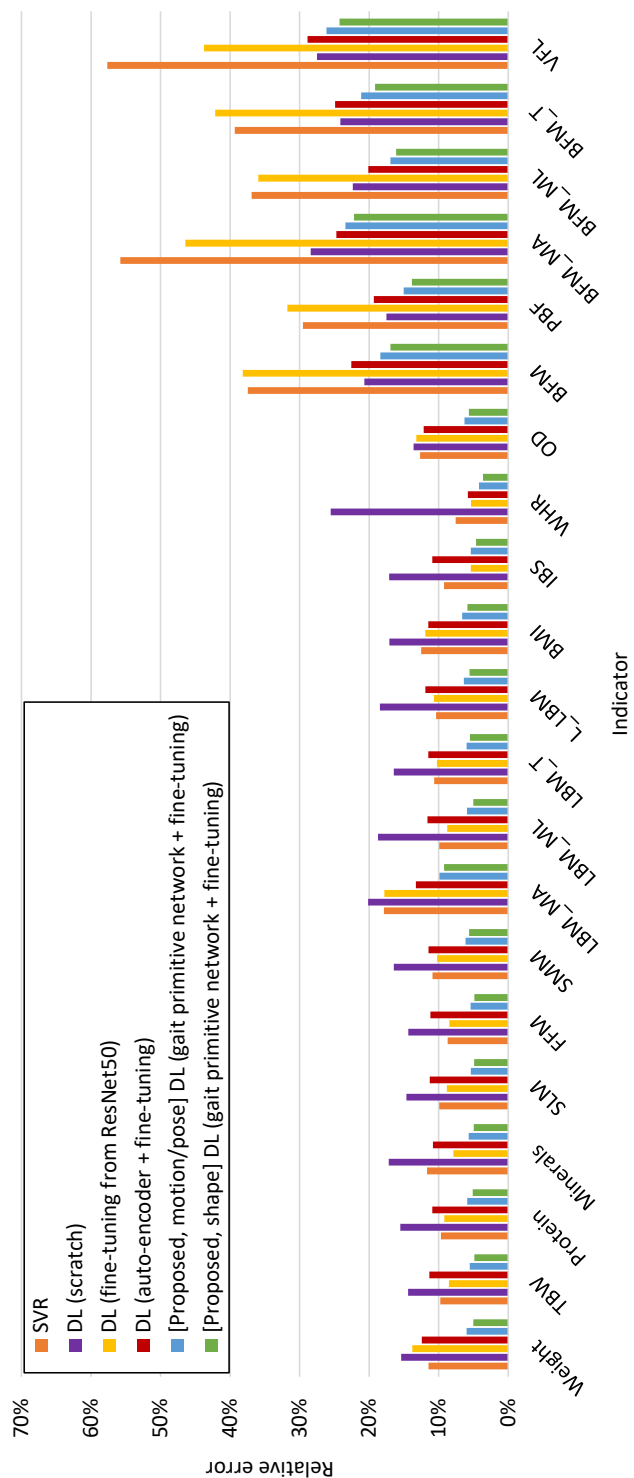
Fig. 3.17 Comparison of the relative errors with the benchmarks.
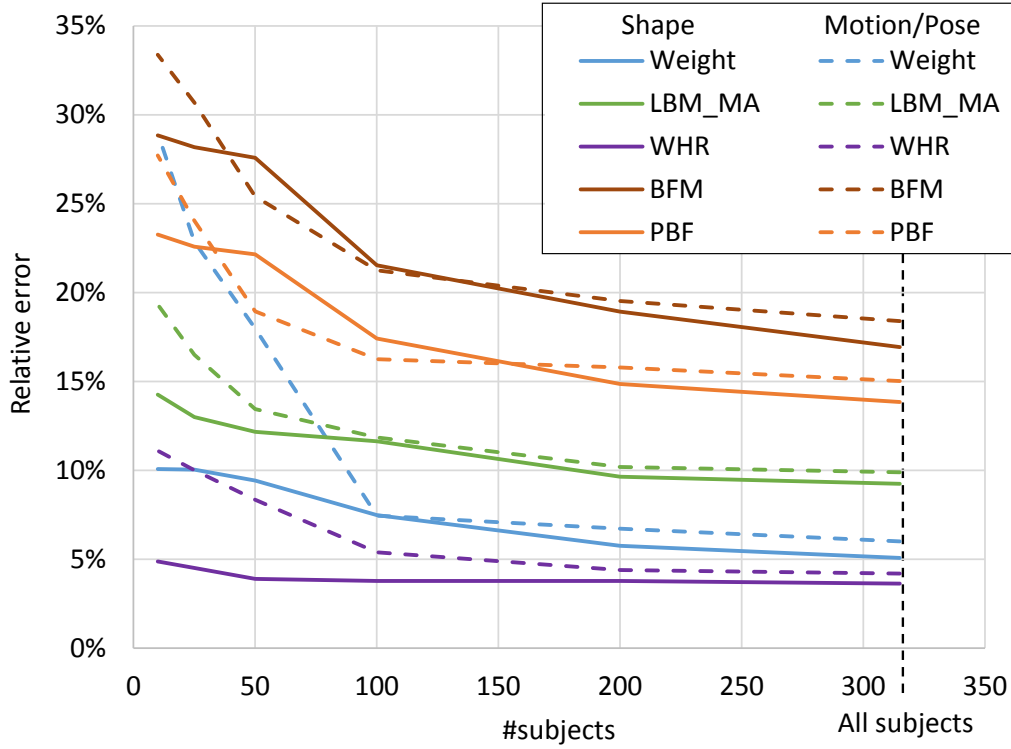
Fig. 3.18 Sensitivity analysis of the number of training samples on the relative errors.

3.5.6   Sensitivity analysis of the number of training samples

Because we claim the effectiveness of the proposed method for a relatively small number of training samples for the target task, we conducted a sensitivity analysis of the number of training samples on the accuracy of health indicator estimation. For this purpose, we evaluated the relative errors of the proposed method for selected health indicators as we decreased the number of training samples, as shown in Fig. 3.18. This experiment indicates that the proposed method can adequately accomplish the target task when the number of training samples exceeds 100.

3.5.7   Discussion

Although the proposed method still has some limitations, such as a large error in estimation of body fat-related indicators, its advantages over existing body composition meters still exist. In our data collection experiments, it took about 2 minutes per subject to measure health indicators using a body composition meter, including preparation time (e.g., taking off and putting on shoes and socks). On the other hand, it took only took only 6 to 8 seconds per

subject to capture a gait video, i.e., much faster than the body composition meter. Moreover, while the body composition meter requires subject's contact to the device, the gait video capturing does not require the subject's contact, which is preferable under the situation of the current COVID-19 pandemic.

3.6   Conclusion

We proposed a method to estimate health indicators related to body composition using video-based gait analysis by a deep learning framework. To address the challenge of the small sample size, we pre-trained the gait primitive networks with a large-scale gait database, and fine-tuned them with a limited number of health indicator training samples. We confirmed that the proposed method achieved fairly low relative errors (approximately 5%) for the health indicators other than body fat-relevant ones, and outperformed the benchmarks.

Future research directions include incorporating age and gender information for health indicator estimation because the correlation between these factors is well established. In addition, the temporal information of gait is not exploited by the estimation due to using GEI as input. Therefore, we will try to use gait features that contain temporal information, e.g., silhouette sequences. Furthermore, although we used only side-view gait images in this study, we believe that estimation using data from other view angles is also possible. Since the multi-view large-scale gait database [68] has been already available, we will try to extend the proposed method to allow the estimation of health indicators using an arbitrary-view gait video.

Chapter 4

Discussion

In the above study, the author tried the proposed strategies and obtained certain results.

In the diagnostic study of iNPH, the dataset is extremely small. Therefore, the authors adopted strategy I, i.e., manually designed gait features using the existing medical domain knowledge for analysis, and successfully judged the results of the tap test.

In the study of health indicator estimation, the scale of the dataset is larger but still insufficient for a from-scratch deep learning method. The author then adopted strategy II, i.e., pre-trained the CNN using general gait features as primitives and fine-tuned the primitive network using the dataset. The obtained estimator yields an acceptable accuracy.

Both the studies share the characteristic of a small dataset scale, and the different strategies used for the analyses are due to the different dataset scales and the relevance of the available domain-specific knowledge to the topic. In the study of iNPH diagnosis, the number of subjects for each label (tap-positive and tap-negative) was less than 10, which is not even amenable for classical machine learning methods, but the knowledge of the medical field was very exhaustive. On the other hand, in the study of health indicator estimation there were more than 300 subjects, but domain-specific knowledge was only marginally relevant and could not be directly designed manually for features. The analysis methods used in both studies were designed separately based on the above characteristics. If these characteristics are not considered when designing methods, it is hardly possible to obtain good results.

In this chapter, the author conducted some additional experiments as benchmarks to justify the previous discussion. In addition, the scope of the two strategies is also discussed in the last section.

4.1   iNPH diagnosis using deep learning approach with the pre-training-fine-tuning strategy

First, the author attempted to use deep learning approach to judge the results of CSF tap test in the study of iNPH diagnosis. As in studies of health indicator estimation, the author attempted to pre-train the network using gait primitives extracted from a large-scale gait database, and then fine-tuned it using the dataset of iNPH patients.

For the gait primitives, based on the principle of using the information with the strongest relevance to the study topic, the author chose three gait symptoms of iNPH, i.e., lateral sway, wide-base gait and petit-pas gait, which could be extracted automatically from a large-scale database without any manual work. The extraction methods are described below.

Lateral sway: The author defined the region of 16 pixels at the top of the silhouette (which with a whole vertical size of 128 pixels) as the head region, and calculated the center of gravity of its foreground pixels as the head position. The score for lateral sway was calculated from the trajectory of the head position using the method described in Section 2.2.2.

Wide-base gait: The author defined the region of 28 pixels at the bottom of the silhouette as the foot region, and extracted the lateral coordinates of the leftmost and rightmost foreground pixels of the foot region in the whole image sequence, and then calculated the distance between them as the score of the wide-base gait.

Petit-pas gait: Since all image sequences in the database have the same walking distance, the author used an existing method [103] to calculate the gait period of the subjects, which can be considered as a reciprocal of the step length.

Since the dataset of iNPH patients has only front-view images, the author chose 16,000 image sequences from the OU-ISIR MVLP database [67], which also has front-view gait images, as the pre-training dataset. The GEI and iNPH-related gait primitives extracted from the dataset were used to pre-train the primitive network. The structure and layer configuration for the pre-training network are exactly the same as those described in Section 3.4.4.

Subsequently, the author fine-tuned each of the three primitive networks into classifiers for judging the result labels of CSF tap test. Since judging the results of the CSF tap test requires comparing the gait of pre- and post-tap, the author calculated differential images for all combinations of GEI pre- and post-tap for each patient (as shown in Fig. 4.1), and used them with the ground truth (i.e., labels judged by the physicians) as the training set for fine-tuning. The structure of the fine-tuning network is almost the same as shown in Fig. 3.6, except that the output of the last layer (fc2) was changed to two dimensions, i.e., the predicted probabilities of the two labels (i.e., tap-positive and tap-negative).

Post-tap GEI   Pre-tap GEI   differential image
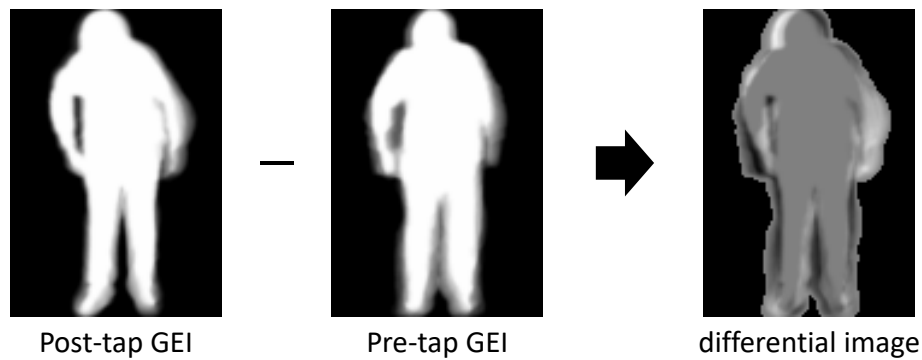
Fig. 4.1 Example of the differential image for pre- and post-tap GEI. It is calculated by subtracting the pre-tap GEI from the post-tap GEI. Where the difference in pixel values is 0, the value of the corresponding pixel on the differential image is set to 128 (gray as in the center of the differential image); when the pixel value of the post-tap GEI is greater than the pre-tap GEI, the corresponding pixel is closer to white, and vice versa.

Table 4.1 Accuracy for CSF tap test judgement using deep learning method.

Primitive	Lateral sway	Petit-pas gait	Wide-base gait
Accuracy	33.3%	51.2%	55.6%

The author used rank-1 accuracy, i.e., the output is considered correct when the label with the highest probability is the same as the ground truth label, to assess the accuracy of the prediction, and used leave-one-out cross-validation to test the performance of the classifier. The data from one patient was retained as the test set, while one tap-positive and one tap-negative each were randomly selected from the other 17 patients as the validation set, and the rest were used for training.

As shown in Table 4.1, the accuracy rates of the test sets in classification are quite low. Moreover, as shown in Fig. 4.2, the classifier is almost always consistent across sequences for each patient, resulting in either a 100% or 0% accuracy rate. This is rather an identification of the patient than a determination of the patient's symptoms, suggesting that the available data are inadequate for this task.
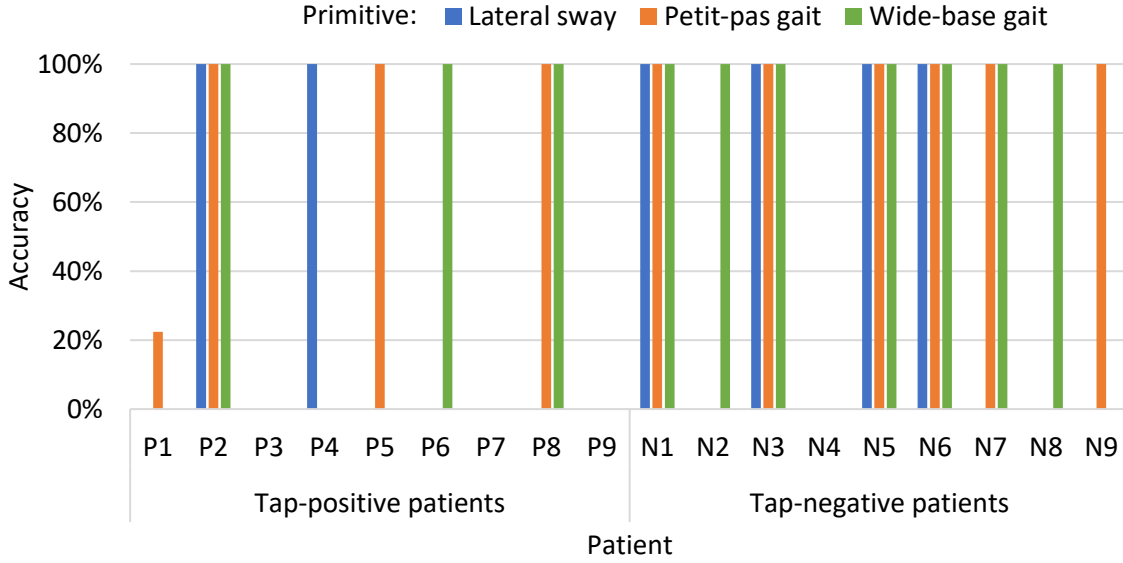
Fig. 4.2 Result of CSF tap test which judged by deep learning classification method.

4.2 Health indicator estimation via manually designed feature

Second, the author attempted to estimate health indicators using manually designed features. However, without any relevant knowledge about the estimation of health indicators from gait, the author can only design features by general understanding, that the amount of weight and body composition is related to body width and area of the silhouette. Therefore, the author used the three body width data (chest, waist, and hip) described in Section 3.4.3, with the static component area of the GEI and the whole area of the GEI as manually designed features, and calculated health indicators using support vector regression.

The author performed a 20-fold cross-validation same as the proposed method to assess the performance of this method. As shown in Fig. 4.3 the estimation accuracy of the method using manually designed features is lower than that of the proposed method, which uses domain-specific knowledge to pre-train the primitive network and fine-tune it with the dataset, in almost all indicators. Meanwhile, the SVR estimation results using manually designed features as input have higher accuracy than when GEI is used directly as input. This indicates that even if the analysis method is not entirely appropriate, it has a role in improving the accuracy as long as domain-specific knowledge is used.

Fig. 4.3 Results of health indicator estimation using manually designed features, compared with proposed methods and SVR estimation directly from GEI, which used as a benchmark in Section 3.5.5.

4.3 iNPH diagnosis via classical machine learning method with classical handcrafted feature

Classical machine learning methods are believed to work well on small datasets as well. However, when the dataset scale is as small as the one in the study of iNPH diagnosis, these methods are also incapable. To demonstrate this, the author tried to use classical machine learning with classical handcrafted features to judge the results of CSF tap tests for iNPH patients.

The author extracted local binary pattern (LBP) features using the differential GEIs illustrated in Section 4.1 and Fig. 4.1, and trained a classifier for tap-positive and tap-negative labels using support vector machine (SVM).

Table 4.2 Classification accuracy for CSF tap test judgement using SVM with LBP feature.

Label	Tap-positive	Tap-negative
Accuracy	24.9%	32.6%
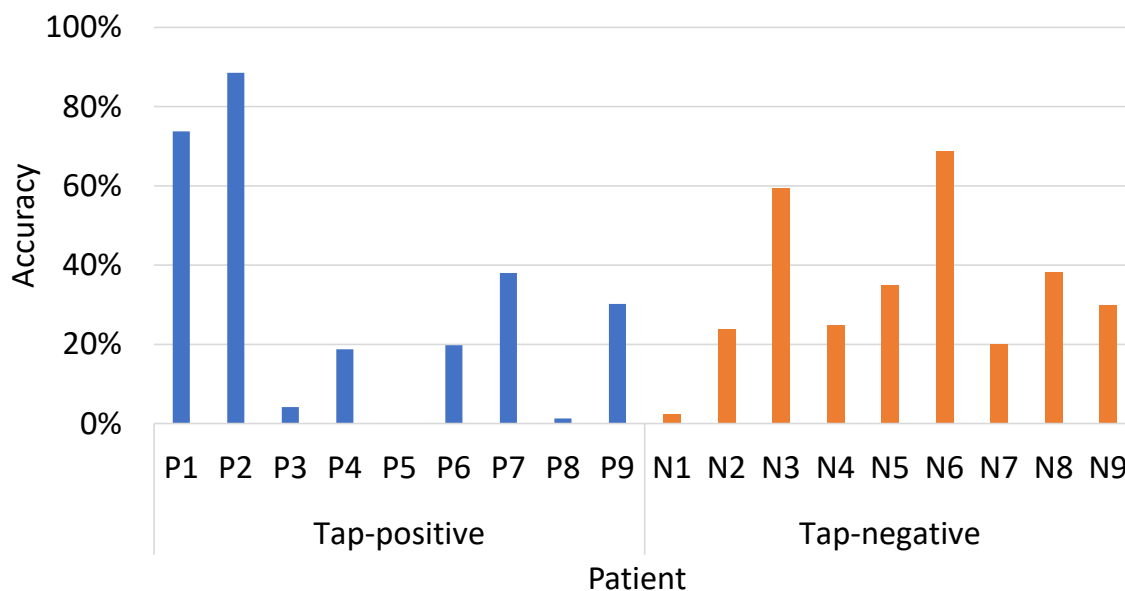
Fig. 4.4 Result of CSF tap test which judged by SVM with LBP feature.

As in Section 4.1, the authors evaluate rank-1 accuracy and use the leave-one-out cross-validation to test the performance of the classifier. Table 4.2 showed the accuracy of the tap-positive and tap-negative labels respectively, and both of them are very low. Fig. 4.4 showed the accuracy for each patient separately, and it can be seen that despite the results are not as extreme as those using the deep learning method in Section 4.1, the variation between the accuracies of patients within the same label is still significant.

4.4 Scope of the two strategies

As shown in the results of the sensitivity analysis of the study on health indicator estimation (Fig. 3.18), when the dataset scale is less than 100 subjects, it becomes difficult for the deep learning approach to obtain good results. Therefore, a threshold of 100 subjects can be used in selecting the applied strategy, i.e., a dataset with a size less than 100 subjects should be considered as insufficient to competently fine-tune the primitive network (i.e., the scale is very small in Fig. 1.2), while a dataset larger than 100 subjects but still insufficient for direct training of the deep neural network can be tried using strategy II.

This threshold is also should be adjusted according to the task to be achieved and the sample distribution in the actual dataset, etc.

In classification tasks, the interclass distribution of sample size can significantly affect the classification accuracy. For example, in the study of iNPH diagnosis, if the number of patients with tap-positive increases many times while the number of tap-negative remains unchanged, then despite the overall scale of the dataset will become larger, it will still not be beneficial for the classification task. Therefore, in classification tasks, strategies should be selected according to the scale of the class with the smallest number of samples, rather than the scale of the entire dataset.

As for the regression tasks, the distribution of the dataset also affects the accuracy. For example, if the dataset for the study of health indicator estimation has only subjects with standard physique without any over- or underweight ones, the trained estimator will have difficulty in yielding accurate results for the latter. Therefore, the actual scale of the dataset can be used for selection only if its distribution obeys the prior probability, otherwise the scale needs to be treated as a smaller one accordingly in the selection.

Chapter 5

Conclusion

This thesis describes strategies for solving the problem of undersized datasets which is often experienced when using computer vision methods in the medical topics. Deep learning methods are frequently used and very effective in current computer vision studies, however, a dataset with sufficiently large scale is a prerequisite for the availability of this approach. In the case of topics in the medical field, it is difficult to obtain adequately data, which are usually only tens to hundreds in size, due to the double limitation of the number of subjects and professionals. Datasets of this scale are difficult to train deep learning networks with adequate accuracy, and classical machine learning methods are not competent for complex analysis of these topics. Therefore, strategies are needed to compensate for the lack of information due to insufficient datasets for the study of this kind of topics.

The key to this issue is the lack of information, and therefore more information needs to be introduced into the study as compensation. Thus, what kind of information should be introduced and how it should be used becomes critical. In this regard, two strategies are proposed, i.e., using domain-specific knowledge that is as closely related to the topic as possible, and deciding on the analysis method for the data with sufficient consideration of the scale of the dataset and the relevance of the domain-specific knowledge for the topic. The author conducted two application studies as a practice of these strategies, namely the study of video-based gait analysis for iNPH diagnosis and the study of health indicator estimation by video-based gait analysis.

In the study on iNPH diagnosis, the dataset obtained by the author contained only 18 subjects. Also, due to site limitations, only front- and back-view walking images can be captured, which contain less gait information than side-view images. On the other hand, gait symptoms of iNPH are well described in the medical field, and physicians can make judgments about a patient's gait symptoms from front-view videos. Therefore, the author manually designed gait features that can be extracted from front view images as described in

medical field works, and developed a method for judging CSF tap test results by analyzing the values of these features extracted from the subject's gait videos.

In the study on health indicator estimation, the author obtained a dataset of more than 300 individuals due to fewer restrictions on data collection. However, although it is sufficiently likely that gait will indicate health indicators, there is still no existing literature that describes how health indicators can be calculated from gait. The author also tried to train the deep learning network from scratch using the gait-health indicator dataset, but its results were not satisfactory. Since there is no direct knowledge of both gait and health indicators, the author chose to use the sub-optimal, i.e., features that can be measured by gait, which are gait primitives such as arm swing amplitude and body width. It is possible that these gait primitives also indicate health indicators and therefore have a certain relevance to this topic. And since only gait information is required for their measurements, a normal gait database can be used as a training dataset. The author therefore used a large-scale gait database OULP-Age, extracted gait primitives using existing methods and trained the primitive networks. Subsequently, the author fine-tuned the primitive networks using the gait-health indicator dataset and obtained estimators with acceptable accuracy.

Through the two studies mentioned above, the author demonstrated the effectiveness of the two strategies proposed in this work. In studies of health indicator estimation, the results of benchmark experiments demonstrate that using domain-specific knowledge with strong relevance is better than using it with weaker relevance or not using it at all. On the other hand, in order to demonstrate the necessity of deciding the data analysis methods according to the scale of the dataset and the relevance of domain-specific knowledge, the author tried to exchange the analysis methods in the two studies as additional experiments. The results showed that the analysis methods were not interchangeable when the scale of the datasets and the relevance of domain-specific knowledge were different.

The above practice proves the effectiveness of the proposed strategies, but there are still issues remaining. For topics like iNPH diagnosis where existing domain-specific knowledge is exhaustive, all the researcher needs to do is to make full use of that knowledge. However, for topics like health indicator estimation, a specific solution on how to select and utilize domain-specific knowledge is still lacking. Gait characteristics such as arm swing and stride length were used as primitives in this study because there are existing, well-established measurement methods; while body width was added in consideration of the difficulty of expressing body shape information by the characteristics measured by existing methods, and its measurement method was developed originally and immature. In the end, however, body width works better as a gait primitive instead. Therefore, such an approach may not be effective in other topics, and the gait primitives and their usage now chosen in this topic

may not be the best option. In the absence of strongly associated domain-specific knowledge, strategies for the selection and use of existing knowledge still need further research.

References

[1] Peter Savadjiev, Jaron Chong, Anthony Dohan, Maria Vakalopoulou, Caroline Reinhold, Nikos Paragios, and Benoit Gallix. Demystification of AI-driven medical image interpretation: past, present and future. *European Radiology*, 29(3):1616–1624, March 2019.

[2] A.D. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, March 2000. Conference Name: IEEE Transactions on Medical Imaging.

[3] A. Hoover and M. Goldbaum. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transactions on Medical Imaging*, 22(8):951–958, August 2003. Conference Name: IEEE Transactions on Medical Imaging.

[4] Takashi Matsuyama. Expert systems for image processing: Knowledge-based composition of image analysis processes. *Computer Vision, Graphics, and Image Processing*, 48(1):22–49, October 1989.

[5] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on Medical Imaging*, 8(3):263–269, September 1989. Conference Name: IEEE Transactions on Medical Imaging.

[6] Sharon A. Stansfield. ANGY: A Rule-Based Expert System for Automatic Segmentation of Coronary Vessels From Digital Subtracted Angiograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(2):188–199, March 1986. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[7] Toshiyuki Okada, Marius George Linguraru, Masatoshi Hori, Ronald M. Summers, Noriyuki Tomiyama, and Yoshinobu Sato. Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. *Medical Image Analysis*, 26(1):1–18, December 2015.

[8] S.K. Warfield, K.H. Zou, and W.M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, July 2004. Conference Name: IEEE Transactions on Medical Imaging.

[9] Juan Eugenio Iglesias and Mert R. Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205–219, August 2015.

References

[10] M. Goldbaum, S. Moezzi, A. Taylor, S. Chatterjee, J. Boyd, E. Hunter, and R. Jain. Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 695–698 vol.3, September 1996.

[11] Hyunjin Park, P.H. Bland, and C.R. Meyer. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on Medical Imaging*, 22(4):483–492, April 2003. Conference Name: IEEE Transactions on Medical Imaging.

[12] Terence D. Valenzuela, Denise J. Roe, Shan Cretin, Daniel W. Spaite, and Mary P. Larsen. Estimating Effectiveness of Cardiac Arrest Interventions. *Circulation*, 96(10):3308–3313, November 1997. Publisher: American Heart Association.

[13] Stephen S. Raab, Julia C. Lenel, and Michael B. Cohen. Low grade transitional cell carcinoma of the bladder. Cytologic Diagnosis by Key Features as Identified by Logistic Regression Analysis. *Cancer*, 74(5):1621–1626, 1994. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-0142%2819940901%2974%3A5%3C1621%3A%3AAID-CNCR2820740521%3E3.0.CO%3B2-E.

[14] R. L. Kennedy, H. S. Fraser, L. N. McStay, and R. F. Harrison. Early diagnosis of acute myocardial infarction using clinical and electrocardiographic data at presentation: derivation and evaluation of logistic regression models. *European Heart Journal*, 17(8):1181–1191, August 1996.

[15] Lisa Ekselius, Eva Lindström, Lars von Knorring, Ove Bodlund, and Gunnar Kullgren. A Principal Component Analysis of the DSM-III-R Axis II Personality Disorders. *Journal of Personality Disorders*, 8(2):140–148, June 1994. Publisher: Guilford Publications Inc.

[16] J.S. Taur and C.W. Tao. Medical image compression using principal component analysis. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 2, pages 903–906 vol.2, September 1996.

[17] Jean-Claude Rymer, Robert Sabatier, Alain Daver, Jacques Bourleaud, Marcel Assicot, Jacqueline Bremond, Jacqueline Rapin, Sharon Lynn Salhi, Bruno Thirion, Anne Vassault, and others. A new approach for clinical biological assay comparison and standardization: application of principal component analysis to a multicenter study of twenty-one carcinoembryonic antigen immunoassay kits. *Clinical chemistry*, 45(6):869–881, 1999. Publisher: Oxford University Press.

[18] K Veropoulos, N Cristianini, and C Campbell. The application of support vector machines to medical decision support: a case study. *Advanced Course in Artificial Intelligence*, pages 1–6, 1999.

[19] Martin Bonneville, Jean Meunier, Yoshua Bengio, and Jean-Paul Soucy. Support vector machines for improving the classification of brain PET images. In *Medical Imaging 1998: Image Processing*, volume 3338, pages 264–273. SPIE, June 1998.

[20] Kai Yang, Bi Zhou, Fei Yi, Yan Chen, and Yingsheng Chen. Colorectal Cancer Diagnostic Algorithm Based on Sub-Patch Weight Color Histogram in Combination of Improved Least Squares Support Vector Machine for Pathological Image. *Journal of Medical Systems*, 43(9):306, August 2019.

[21] M. Kamber, R. Shinghal, D.L. Collins, G.S. Francis, and A.C. Evans. Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images. *IEEE Transactions on Medical Imaging*, 14(3):442–453, September 1995. Conference Name: IEEE Transactions on Medical Imaging.

[22] Johannes Mair, Bernd Puschendorf, Jörn Smidt, Peter Lechleitner, and Franz Dienstl. A Decision Tree for the Early Diagnosis of Acute Myocardial Infarction in Nontraumatic Chest Pain Patients at Hospital Admission. *Chest*, 108(6):1502–1509, December 1995.

[23] E. Viikki, M. Kentala, I. Juhola, and K. Pyykko. Decision tree induction in the diagnosis of otoneurological diseases. *Medical Informatics and the Internet in Medicine*, 24(4):277–289, January 1999. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/146392399298302.

[24] Yoshua Bengio. *Learning Deep Architectures for AI*. Now Publishers Inc, 2009. Google-Books-ID: cq5ewg7FniMC.

[25] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7553 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Computer science;Mathematics and computing Subject_term_id: computer-science;mathematics-and-computing.

[26] Wenhua Shao, Haiyong Luo, Fang Zhao, Cong Wang, Antonino Crivello, and Muhammad Zahid Tunio. DePedo: Anti Periodic Negative-Step Movement Pedometer with Deep Convolutional Neural Networks. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6, May 2018. ISSN: 1938-1883.

[27] Debleena Paul, Gaurav Sanap, Snehal Shenoy, Dnyaneshwar Kalyane, Kiran Kalia, and Rakesh K. Tekade. Artificial intelligence in drug discovery and development. *Drug Discovery Today*, 26(1):80–93, January 2021.

[28] Raquel Dias and Ali Torkamani. Artificial intelligence in clinical and genomic diagnostics. *Genome Medicine*, 11(1):70, November 2019.

[29] Daniel S. W. Ting, Yong Liu, Philippe Burlina, Xinxing Xu, Neil M. Bressler, and Tien Y. Wong. AI for medical imaging goes deep. *Nature Medicine*, 24(5):539–540, May 2018. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 5 Primary_atype: News & Views Publisher: Nature Publishing Group Subject_term: Diagnosis;Machine learning;Medical imaging Subject_term_id: diagnosis;machine-learning;medical-imaging.

[30] Paras Lakhani and Baskaran Sundaram. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 284(2):574–582, August 2017. Publisher: Radiological Society of North America.

References

[31] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, February 2017. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7639 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Diagnosis;Machine learning;Skin cancer Subject_term_id: diagnosis;machine-learning;skin-cancer.

[32] Ryan Poplin, Avinash V. Varadarajan, Katy Blumer, Yun Liu, Michael V. McConnell, Greg S. Corrado, Lily Peng, and Dale R. Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158–164, March 2018. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 3 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Cardiology;Computational science;Engineering;Risk factors Subject_term_id: cardiology;computational-science;engineering;risk-factors.

[33] Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C. S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K. Prasadha, Jacqueline Pei, Magdalene Y. L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A. N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A. Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, and Kang Zhang. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, 172(5):1122–1131.e9, February 2018.

[34] Eliza Strickland. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 56(4):24–31, April 2019. Conference Name: IEEE Spectrum.

[35] Shuo Jin, Bo Wang, Haibo Xu, Chuan Luo, Lai Wei, Wei Zhao, Xuexue Hou, Wenshuo Ma, Zhengqing Xu, Zhuozhao Zheng, Wenbo Sun, Lan Lan, Wei Zhang, Xiangdong Mu, Chenxin Shi, Zhongxiao Wang, Jihae Lee, Zijian Jin, Minggui Lin, Hongbo Jin, Liang Zhang, Jun Guo, Benqi Zhao, Zhizhong Ren, Shuhao Wang, Zheng You, Jiahong Dong, Xinghuan Wang, Jianming Wang, and Wei Xu. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. Technical report, Institute for Precision Medicine, Tsinghua University, March 2020. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Type: article.

[36] Salih A. Salih, Richard Wootton, Elaine Beller, and Len Gray. The validity of video clips in the diagnosis of gait disorder. *Journal of Telemedicine and Telecare*, 13(7):333–336, 2007. _eprint: https://doi.org/10.1258/135763307782215406.

[37] Michael W. Hayes, Shanti Graham, Peter Heldorf, Gregory de Moore, and John G. L. Morris. A video review of the diagnosis of psychogenic gait: Appendix and commentary. *Movement Disorders*, 14(6):914–921, 1999. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1531-8257%28199911%2914%3A6%3C914%3A%3AAID-MDS1002%3E3.0.CO%3B2-B.

[38] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. View-invariant Estimation of Height and Stride for Gait Recognition. In Massimo Tistarelli, Josef Bigun, and Anil K. Jain, editors, *Biometric Authentication*, Lecture Notes in Computer Science, pages 155–167, Berlin, Heidelberg, 2002. Springer.

[39] DAVID M. BUCHNER, ERIC B. LARSON, EDWARD H. WAGNER, THOMAS D. KOEPSELL, and BARBARA J. DE LATEUR. Evidence for a Non-linear Relationship between Leg Strength and Gait Speed. *Age and Ageing*, 25(5):386–391, September 1996. _eprint: https://academic.oup.com/ageing/article-pdf/25/5/386/78499/25-5-386.pdf.

[40] Kristen M Beavers, Daniel P Beavers, Denise K Houston, Tamara B Harris, Trisha F Hue, Annemarie Koster, Anne B Newman, Eleanor M Simonsick, Stephanie A Studenski, Barbara J Nicklas, and Stephen B Kritchevsky. Associations between body composition and gait-speed decline: results from the Health, Aging, and Body Composition study. *The American Journal of Clinical Nutrition*, 97(3):552–560, January 2013. _eprint: https://academic.oup.com/ajcn/article-pdf/97/3/552/23818709/552.pdf.

[41] Stefano Volpato, Lara Bianchi, Fulvio Lauretani, Fabrizio Lauretani, Stefania Bandinelli, Jack M. Guralnik, Giovanni Zuliani, and Luigi Ferrucci. Role of Muscle Mass and Muscle Quality in the Association Between Diabetes and Gait Speed. *Diabetes Care*, 35(8):1672–1679, August 2012. Publisher: American Diabetes Association Section: Original Research.

[42] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. GEINet: View-invariant gait recognition using a convolutional neural network. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, June 2016.

[43] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. In *Proc. of the 33rd AAAI Conf. on Artificial Intelligence (AAAI 2019)*, 2019.

[44] Xiang Li, Yasushi Makihara, Chi Xu, and Yasushi Yagi. End-to-End Model-Based Gait Recognition Using Synchronized Multi-View Pose Constraint. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4106–4115, 2021.

[45] Chi Xu, Yasushi Makihara, Xiang Li, Yasushi Yagi, and Jianfeng Lu. Cross-View Gait Recognition Using Pairwise Spatial Transformer Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(1):260–274, January 2021. Conference Name: IEEE Transactions on Circuits and Systems for Video Technology.

[46] Xu Yang, Bin-Bin Gao, Chao Xing, Zeng-Wei Huo, Xiu-Shen Wei, Ying Zhou, Jianxin Wu, and Xin Geng. Deep Label Distribution Learning for Apparent Age Estimation. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 344–350, December 2015.

[47] Atsuya Sakata, Yasushi Makihara, Noriko Takemura, Daigo Muramatsu, and Yasushi Yagi. Gait-Based Age Estimation Using a DenseNet. In Gustavo Carneiro and Shaodi You, editors, *Computer Vision–ACCV 2018 Workshops*, Lecture Notes in Computer Science, pages 55–63. Springer International Publishing, 2019.

References

[48] Chi Xu, Yasushi Makihara, Ruochen Liao, Hirotaka Niitsuma, Xiang Li, Yasushi Yagi, and Jianfeng Lu. Real-Time Gait-Based Age Estimation and Gender Classification From a Single Image. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3460–3470, 2021.

[49] Taocheng Liu, Xiangbin Ye, and Bei Sun. Combining Convolutional Neural Network and Support Vector Machine for Gait-based Gender Recognition. In *2018 Chinese Automation Congress (CAC)*, pages 3477–3481, November 2018.

[50] Weijie Sheng and Xinde Li. Multi-task learning for gait-based identity recognition and emotion recognition using attention enhanced temporal graph convolutional network. *Pattern Recognition*, 114:107868, June 2021.

[51] Venkatraman Narayanan, Bala Murali Manoghar, Vishnu Sashank Dorbala, Dinesh Manocha, and Aniket Bera. ProxEmo: Gait-based Emotion Learning and Multi-view Proxemic Fusion for Socially-Aware Robot Navigation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8200–8207, October 2020. ISSN: 2153-0866.

[52] Uttaran Bhattacharya, Christian Roncal, Trisha Mittal, Rohan Chandra, Kyra Kapsaskis, Kurt Gray, Aniket Bera, and Dinesh Manocha. Take an Emotion Walk: Perceiving Emotions from Gaits Using Hierarchical Attention Pooling and Affective Mapping. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, pages 145–163, Cham, 2020. Springer International Publishing.

[53] Yasushi Makihara, Yuta Hayashi, Allam Shehata, Daigo Muramatsu, and Yasushi Yagi. Estimation of Gait Relative Attribute Distributions using a Differentiable Tradeoff Model of Optimal and Uniform Transports. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8, August 2021. ISSN: 2474-9699.

[54] Yasushi Makihara. Video-based Gait Analysis and Its Applications. In *Int. Conf. on Big Data, IoT and Machine Learning (BIM 2021)*, online, September 2021.

[55] Ralph Gross and Jianbo Shi. The CMU Motion of Body (MoBo) Database. Technical report, Carnegie Mellon University, 2001.

[56] M Nixon, J Carter, J Shutler, and M Grant. Experimental plan for automatic gait recognition. Technical report, Technical report, University of Southampton, 2001.

[57] Sina Samangooei, John Bustard, Mark S Nixon, and John N Carter. On acquisition and analysis of a dataset comprising of gait, ear and semantic data. *Multibiometrics for Human Identification*, pages 277–301, 2011. Publisher: Cambridge Univ. Press.

[58] Liang Wang, Huazhong Ning, Weiming Hu, and Tieniu Tan. Gait recognition based on Procrustes shape analysis. In *Proceedings. International Conference on Image Processing*, volume 3, pages III–III, September 2002. ISSN: 1522-4880.

[59] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 4, pages 441–444, August 2006. ISSN: 1051-4651.

[60] Daoliang Tan, Kaiqi Huang, Shiqi Yu, and Tieniu Tan. Efficient Night Gait Recognition Based on Template Matching. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 3, pages 1000–1003, August 2006. ISSN: 1051-4651.

[61] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanID gait challenge problem: data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):162–177, February 2005. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[62] David López-Fernández, Francisco José Madrid-Cuevas, Ángel Carmona-Poyato, Manuel Jesús Marín-Jiménez, and Rafael Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In Pier Luigi Mazzeo, Paolo Spagnolo, and Thomas B. Moeslund, editors, *Activity Monitoring by Multiple Distributed Sensing*, Lecture Notes in Computer Science, pages 26–39, Cham, 2014. Springer International Publishing.

[63] Brian DeCann, Arun Ross, and Jeremy Dawson. Investigating gait recognition in the short-wave infrared (SWIR) spectrum: dataset and challenges. In *Biometric and Surveillance Technology for Human and Activity Identification X*, volume 8712, pages 101–116. SPIE, May 2013.

[64] Yumi Iwashita, Ryosuke Baba, Koichi Ogawara, and Ryo Kurazume. Person Identification from Spatio-temporal 3D Gait. In *2010 International Conference on Emerging Security Technologies*, pages 30–35, September 2010.

[65] Yasushi Makihara, Atsushi Mori, and Yasushi Yagi. Temporal Super Resolution from a Single Quasi-periodic Image Sequence Based on Phase Registration. In Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, editors, *Computer Vision – ACCV 2010*, Lecture Notes in Computer Science, pages 107–120, Berlin, Heidelberg, 2011. Springer.

[66] Haruyuki Iwama, Mayu Okumura, Yasushi Makihara, and Yasushi Yagi. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, October 2012. Conference Name: IEEE Transactions on Information Forensics and Security.

[67] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, 10(1):4, February 2018.

[68] Chi Xu, Yasushi Makihara, Gakuto Ogi, Xiang Li, Yasushi Yagi, and Jianfeng Lu. The OU-ISIR Gait Database comprising the Large Population Dataset with Age and performance evaluation of age estimation. *IPSJ Transactions on Computer Vision and Applications*, 9(1):24, December 2017.

[69] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSJ Transactions on Computer Vision and Applications*, 10(1):4, Feb 2018.

References

[70] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face Recognition with Local Binary Patterns. In Tomás Pajdla and Jiří Matas, editors, *Computer Vision - ECCV 2004*, Lecture Notes in Computer Science, pages 469–481, Berlin, Heidelberg, 2004. Springer.

[71] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, June 2005. ISSN: 1063-6919.

[72] Srikanth Tammina. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10):p9420, January 2019.

[73] Long Wen, X. Li, Xinyu Li, and Liang Gao. A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 205–209, May 2019.

[74] Yong Li, Zegang Ding, Chi Zhang, Yan Wang, and Jing Chen. SAR Ship Detection Based on Resnet and Transfer Learning. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 1188–1191, July 2019. ISSN: 2153-7003.

[75] A. Sai Bharadwaj Reddy and D. Sujitha Juliet. Transfer Learning with ResNet-50 for Malaria Cell-Image Classification. In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pages 0945–0949, April 2019.

[76] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778, 2016.

[77] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, April 2015. arXiv: 1409.1556.

[78] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *Proc. of the 22nd IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.

[79] Muhammad Farooq and Abdul Hafeez. COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs. *arXiv:2003.14395 [cs, eess]*, March 2020. arXiv: 2003.14395.

[80] Adrian Bulat and Georgios Tzimiropoulos. Human Pose Estimation via Convolutional Part Heatmap Regression. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision - ECCV 2016*, Lecture Notes in Computer Science, pages 717–732, Cham, 2016. Springer International Publishing.

[81] Shuo Yuan, Xinguo Yu, and Abdul Majid. Robust Face Tracking Using Siamese-VGG with Pre-training and Fine-tuning. In *2019 4th International Conference on Control and Robotics Engineering (ICCRE)*, pages 170–174, April 2019.

[82] S. Hakim and R. D. Adams. The special clinical problem of symptomatic hydrocephalus with normal cerebrospinal fluid pressure: observations on cerebrospinal fluid hydrodynamics. *Journal of the neurological sciences*, 2(4):307–327, 1965.

[83] Chifumi Iseki, Toru Kawanami, Hikaru Nagasawa, Manabu Wada, Shingo Koyama, Kenji Kikuchi, Shigeki Arawaka, Keiji Kurita, Makoto Daimon, Etsuro Mori, and others. Asymptomatic ventriculomegaly with features of idiopathic normal pressure hydrocephalus on MRI (AVIM) in the elderly: a prospective study in a japanese population. *Journal of the neurological sciences*, 277(1):54–57, 2009.

[84] Naofumi Tanaka, Satoshi Yamaguchi, Hiroyasu Ishikawa, Hiroshi Ishii, and Kenichi Meguro. Prevalence of possible idiopathic normal-pressure hydrocephalus in japan: the osaki-tajiri project. *Neuroepidemiology*, 32(3):171–175, 2009.

[85] Etsuro Mori, Masatsune Ishikawa, Takeo Kato, Hiroaki Kazui, Hiroji Miyake, Masakazu Miyajima, Madoka Nakajima, Masaaki Hashimoto, Nagato Kuriyama, Takahiko Tokuda, and others. Guidelines for management of idiopathic normal pressure hydrocephalus. *Neurologia medico-chirurgica*, 52(11):775–809, 2012.

[86] Joachim K. Krauss, Jens P. Regel, Werner Vach, Dirk W. Droste, Jan J. Borremans, and Thomas Mergner. Vascular risk factors and arteriosclerotic disease in idiopathic normal-pressure hydrocephalus of the elderly. *Stroke*, 27(1):24–29, 1996.

[87] Koreaki Mori. Management of idiopathic normal-pressure hydrocephalus: a multi-institutional study conducted in japan. *Journal of neurosurgery*, 95(6):970–973, 2001.

[88] John Vassilouthis. The syndrome of normal-pressure hydrocephalus. *Journal of neurosurgery*, 61(3):501–509, 1984.

[89] Hajima Kitagaki, Etsuro Mori, Kazunari Ishii, Shigeru Yamaji, Nobutsugu Hirono, and Toru Imamura. CSF spaces in idiopathic normal pressure hydrocephalus: morphology and volumetry. *American Journal of Neuroradiology*, 19(7):1277–1284, 1998.

[90] C. Wikkelsø, H. Andersson, C. Blomstrand, and G. Lindqvist. The clinical effect of lumbar puncture in normal pressure hydrocephalus. *Journal of Neurology, Neurosurgery & Psychiatry*, 45(1):64–69, 1982.

[91] H. Stolze, J. P. Kuhtz-Buschbeck, H. Drücke, K. Jöhnk, C. Diercks, S. Palmie, H. M. Mehdorn, M. Illert, and G. Deuschl. Gait analysis in idiopathic normal pressure hydrocephalus–which parameters respond to the CSF tap test? *Clinical neurophysiology*, 111(9):1678–1686, 2000.

[92] Neill R. Graff-Radford and John C. Godersky. Normal-pressure hydrocephalus: onset of gait abnormality before dementia predicts good surgical outcome. *Archives of neurology*, 43(9):940–942, 1986.

[93] C. M. Fisher. Hydrocephalus as a cause of disturbances of gait in the elderly. *Neurology*, 32(12):1358–1358, 1982.

[94] Diane Podsiadlo and Sandra Richardson. The timed "up & go": a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society*, 39(2):142–148, 1991.

References

[95] Srinivas Chivukula, Zachary J. Tempel, Nathan T. Zwagerman, W. Christopher Newman, Samuel S. Shin, Ching-Jen Chen, Paul A. Gardner, Eric M. McDade, and Andrew F. Ducruet. The dynamic gait index in evaluating patients with normal pressure hydrocephalus for cerebrospinal fluid diversion. *World neurosurgery*, 84(6):1871–1876, 2015.

[96] E. W. Sankey, I. Jusue-Torres, B. D. Elder, C. R. Goodwin, S. Batra, J. Hoffberger, J. Lu, A. M. Blitz, and D. Rigamonti. Functional gait outcomes for idiopathic normal pressure hydrocephalus after primary endoscopic third ventriculostomy. *Journal of clinical neuroscience: official journal of the Neurosurgical Society of Australasia*, 22(8):1303–1308, 2015.

[97] Simon Agerskov, Maria Wallin, Per Hellström, Carsten Wikkelsö, and Mats Tullberg. The role of hypertension and diabetes mellitus in the severity of gait and balance disturbances in iNPH. *Fluids and Barriers of the CNS*, 12:O11, 2015.

[98] Per Hellström, P. Klinge, J. Tans, and Carsten Wikkelsø. A new scale for assessment of severity and outcome in iNPH. *Acta neurologica Scandinavica*, 126(4):229–237, 2012.

[99] Makiko Yogo, Shusaku Omoto, Masayo Morita, and Masahiko Suzuki. The quantitative motion analysis using portable gait rhythmogram after CSF drainage in the patients with idiopathic normal pressure hydrocephalus. *Fluids and Barriers of the CNS*, 12:P57, 2015.

[100] Valentina Agostini, Michele Lanotte, Marina Carlone, Marcello Campagnoli, Irene Azzolin, Roberto Scarafia, Giuseppe Massazza, and Marco Knaflitz. Instrumented gait analysis for an objective pre-/postassessment of tap test in normal pressure hydrocephalus. *Archives of physical medicine and rehabilitation*, 96(7):1235–1241, 2015.

[101] Stéphane Armand, Lara Allet, Theodor Landis, Olivier Beauchet, Frédéric Assal, and Gilles Allali. Interest of dual-task-related gait changes in idiopathic normal pressure hydrocephalus. *European Journal of Neurology*, 18(8):1081–1084, 2011.

[102] Ju Man and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2006.

[103] Yasushi Makihara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. Gait recognition using a view transformation model in the frequency domain. In *Computer Vision – ECCV 2006*, pages 151–163. Springer, Berlin, Heidelberg, 2006-05-07.

[104] R. T. Collins, R. Gross, and Jianbo Shi. Silhouette-based human identification from body shape and gait. In *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pages 366–371, 2002.

[105] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

[106] Haruyuki Iwama, Daigo Muramatsu, Yasushi Makihara, and Yasushi Yagi. Gait verification system for criminal investigation. *Information and Media Technologies*, 8(4):1187–1199, 2013.

[107] Mayu Okumura, Yasushi Makihara, Shinsuke Nakamura, Shigeo Morishima, and Yasushi Yagi. The online gait measurement for the audience-participant digital entertainment. In *in Invited Workshop on Vision Based Human Modeling and Synthesis in Motion and Expression*, 2009.

[108] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A Study on Gait-Based Gender Classification. *IEEE Transactions on Image Processing*, 18(8):1905–1910, August 2009.

[109] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi. Gait-based age estimation using a whole-generation gait database. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–6, 2011.

[110] Erik E Stone and Marjorie Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *2011 Annual international conference of the IEEE engineering in medicine and biology society*, pages 6491–6494. IEEE, 2011.

[111] Trong-Nguyen Nguyen, Huu-Hung Huynh, and Jean Meunier. Extracting silhouette-based characteristics for human gait analysis using one camera. In *Proceedings of the Fifth Symposium on Information and Communication Technology*, pages 171–177, 2014.

[112] Tsung-Yen Liao, Shaou-Gang Miaou, and Yu-Ren Li. A vision-based walking posture analysis system without markers. In *2010 2nd International Conference on Signal Processing Systems*, volume 3, pages V3–254. IEEE, 2010.

[113] Yoshihiko Kubo, Hiroaki Kazui, Tetsuihiko Yoshida, Yumiko Kito, Nobuyo Kimura, Hiromasa Tokunaga, Atsushi Ogino, Hiroji Miyake, Masatsune Ishikawa, and Masatoshi Takeda. Validation of grading scale for evaluating symptoms of idiopathic normal-pressure hydrocephalus. *Dementia and geriatric cognitive disorders*, 25(1):37–45, 2007.

[114] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

[115] Carsten Wikkelsø, Hugo Andersson, Christian Blomstrand, Goran Lindqvist, and Pal Svendsen. Predictive value of the cerebrospinal fluid tap-test. *Acta neurologica scandinavica*, 73(6):566–573, 1986.

[116] Lisa D. Ravdin, Heather L. Katzen, Anna E. Jackson, Diamanto Tsakanikas, Stephanie Assuras, and Norman R. Relkin. Features of gait most responsive to tap test in normal pressure hydrocephalus. *Clinical neurology and neurosurgery*, 110(5):455–461, 2008.

[117] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, volume 1, page 3, 2017.

References

[118] Ruochen Liao, Kousuke Moriwaki, Yasushi Makihara, Daigo Muramatsu, Noriko Takemura, and Yasushi Yagi. Health Indicator Estimation by Video-Based Gait Analysis. *IEICE Transactions on Information and Systems*, E104.D(10):1678–1690, 2021.

[119] Gilbert B. Forbes and Julio C. Reina. Adult lean body mass declines with age: Some longitudinal observations. *Metabolism*, 19(9):653–663, September 1970.

[120] Andrew S. Jackson and Michael L. Pollock. Practical Assessment of Body Composition. *The Physician and Sportsmedicine*, 13(5):76–90, May 1985.

[121] Richard Baumgartner. Body composition in healthy aging. *Annals of the New York Academy of Sciences*, 904:437–448, 2000.

[122] Kiyoshi Sanada, Motohiko Miyachi, Izumi Tabata, Katsuhiko Suzuki, Kenta Yamamoto, Hiroshi Kawano, Chiyoko Usui, and Mitsuru Higuchi. Differences in body composition and risk of lifestyle-related diseases between young and older male rowers and sedentary controls. *Journal of Sports Sciences*, 27(10):1027–1034, August 2009.

[123] Jennifer S. Brach, Eleanor M. Simonsick, Stephen Kritchevsky, Kristine Yaffe, and Anne B. Newman. The Association Between Physical Function and Lifestyle Activity and Exercise in the Health, Aging and Body Composition Study. *Journal of the American Geriatrics Society*, 52(4):502–509, 2004.

[124] R. F. Kushner and D. A. Schoeller. Estimation of total body water by bioelectrical impedance analysis. *The American Journal of Clinical Nutrition*, 44(3):417–424, September 1986.

[125] H. C. Lukaski, W. W. Bolonchuk, C. B. Hall, and W. A. Siders. Validation of tetrapolar bioelectrical impedance method to assess human body composition. *Journal of Applied Physiology*, 60(4):1327–1332, April 1986.

[126] Barbara Sternfeld, Long Ngo, William A. Satariano, and Ira B. Tager. Associations of Body Composition with Physical Performance and Self-reported Functional Limitation in Elderly Men and Women. *American Journal of Epidemiology*, 156(2):110–121, July 2002.

[127] Ali Özkan, Gürhan Kayıhan, Yusuf Köklü, Nevin Ergun, Mitat Koz, Gülfem Ersöz, and Alexandre Dellal. The Relationship Between Body Composition, Anaerobic Performance and Sprint Ability of Amputee Soccer Players. *Journal of Human Kinetics*, 35:141–146, December 2012.

[128] David Stodden, Stephen Langendorfer, and Mary Ann Roberton. The Association Between Motor Skill Competence and Physical Fitness in Young Adults. *Research Quarterly for Exercise and Sport*, 80(2):223–229, June 2009. Publisher: Routledge.

[129] P Suchanek, I Kralova Lesna, O Mengerova, J Mrazkova, V Lanska, and P Stavek. Which index best correlates with body fat mass: BAI, BMI, waist or WHR? *Neuro Endocrinology Letters*, 33 Suppl 2:78–82, January 2012.

[130] P. Brambilla, G. Bedogni, M. Heo, and A. Pietrobelli. Waist circumference-to-height ratio predicts adiposity better than body mass index in children and adolescents. *International Journal of Obesity*, 37(7):943–946, July 2013. Number: 7 Publisher: Nature Publishing Group.

[131] P. T. Katzmarzyk and C. Bouchard. Where is the beef? Waist circumference is more highly correlated with BMI and total body fat than with abdominal visceral fat in children. *International Journal of Obesity*, 38(6):753–754, June 2014. Number: 6 Publisher: Nature Publishing Group.

[132] Sarah M. Camhi, George A. Bray, Claude Bouchard, Frank L. Greenway, William D. Johnson, Robert L. Newton, Eric Ravussin, Donna H. Ryan, Steven R. Smith, and Peter T. Katzmarzyk. The Relationship of Waist Circumference and BMI to Visceral, Subcutaneous, and Total Body Fat: Sex and Race Differences. *Obesity*, 19(2):402–408, 2011.

[133] Yasushi Makihara, Hidetoshi Mannami, and Yasushi Yagi. Gait Analysis of Gender and Age Using a Large-Scale Multi-view Gait Database. In Ron Kimmel, Reinhard Klette, and Akihiro Sugimoto, editors, *Computer Vision–ACCV 2010*, Lecture Notes in Computer Science, pages 440–451. Springer Berlin Heidelberg, 2011.

[134] Taku Matsuura, Kazuhiro Sakashita, Andrey Grushnikov, Fumio Okura, Ikuhisa Mitsugami, and Yasushi Yagi. Statistical Analysis of Dual-task Gait Characteristics for Cognitive Score Estimation. *Scientific Reports*, 9(1):1–12, December 2019.

[135] R. Chang, L. Guan, and J. A. Burne. An automated form of video image analysis applied to classification of movement disorders. *Disability and Rehabilitation*, 22(1-2):97–108, January 2000.

[136] Ruochen Liao, Yasushi Makihara, Daigo Muramatsu, Ikuhisa Mitsugami, Yasushi Yagi, Kenji Yoshiyama, Hiroaki Kazui, and Masatoshi Takeda. A video-based gait disturbance assessment tool for diagnosing idiopathic normal pressure hydrocephalus. *IEEJ Transactions on Electrical and Electronic Engineering*, 15(3):433–441, 2020.

[137] Kota Aoki, Trung Thanh Ngo, Ikuhisa Mitsugami, Fumio Okura, Masataka Niwa, Yasushi Makihara, Yasushi Yagi, and Hiroaki Kazui. Early Detection of Lower MMSE Scores in Elderly Based on Dual-Task Gait. *IEEE Access*, 7:40085–40094, 2019. Conference Name: IEEE Access.

[138] Jerry Ajay, Chen Song, Aosen Wang, Jeanne Langan, Zhinan Li, and Wenyao Xu. A pervasive and sensor-free Deep Learning system for Parkinsonian gait analysis. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 108–111, March 2018.

[139] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

References

[140] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1302–1310. IEEE Computer Society, 2017.

[141] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing.

[142] Md. Zasim Uddin, Thanh Trung Ngo, Yasushi Makihara, Noriko Takemura, Xiang Li, Daigo Muramatsu, and Yasushi Yagi. The ou-isir large population gait database with real-life carried object and its performance evaluation. *IPSJ Transactions on Computer Vision and Applications*, 10(1):5, May 2018.

[143] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. Gaitpart: Temporal part-based model for gait recognition. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[144] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. Gait recognition via semi-supervised disentangled representation learning to identity and covariate features. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[145] Mark S. Nixon, Tieniu N. Tan, and Rama Chellappa. *Human Identification Based on Gait*. Int. Series on Biometrics. Springer-Verlag, Dec. 2005.

[146] Yasushi Makihara, Darko S. Matovski, Mark S. Nixon, John N. Carter, and Yasushi Yagi. *Gait Recognition: Databases, Representations, and Applications*, pages 1–15. John Wiley & Sons, Inc., 1999.

[147] Patrick Connor and Arun Ross. Biometric recognition by gait: A survey of modalities and features. *Computer Vision and Image Understanding*, 167:1–27, 2018.

[148] Changsheng Wan, Li Wang, and Vir V. Phoha. A survey on gait recognition. *ACM Comput. Surv.*, 51(5):89:1–89:35, August 2018.

[149] S. Niyogi and E. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. of the 7th IEEE Conf. on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[150] D. Cunado, M.S. Nixon, and J.N. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90(1):1–41, 2003.

[151] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, dec. 2003.

[152] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of the 9th European Conference on Computer Vision*, pages 151–163, Graz, Austria, May 2006.

[153] Worapan Kusakunniran, Qiang Wu, Jian Zhang, and Hongdong Li. Support vector regression for multi-view gait recognition based on local motion feature selection. In *Proc. of IEEE computer society conference on Computer Vision and Pattern Recognition 2010*, pages 1–8, San Francisco, CA, USA, Jun. 2010.

[154] Jiwen Lu and Yap-Peng Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5):382–393, 2010.

[155] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi. Gait-based age estimation using a whole-generation gait database. In *Proc. of the Int. Joint Conf. on Biometrics (IJCB2011)*, pages 1–6, Washington D.C., USA, Oct. 2011.

[156] J. Han and Bir Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, February 2006. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[157] Chen Wang, Junping Zhang, Liang Wang, Jian Pu, and Xiaoru Yuan. Human identification using temporal information preserving gait template. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(11):2164 –2176, nov. 2012.

[158] Toby H. W. Lam, K. H. Cheung, and James N. K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44:973–987, April 2011.

[159] K. Bashir, T. Xiang, and S. Gong. Gait recognition without subject cooperation. *Pattern Recognition Letters*, 31(13):2052–2060, Oct. 2010.

[160] J.D. Shutler, M.G. Grant, M.S. Nixon, and J.N. Carter. On a large sequence-based human gait database. In *Proc. of the 4th Int. Conf. on Recent Advances in Soft Computing*, pages 66–71, Nottingham, UK, Dec. 2002.

[161] Y. Makihara, H. Mannami, A. Tsuji, M.A. Hossain, K. Sugiura, A. Mori, and Y. Yagi. The ou-isir gait database comprising the treadmill dataset. *IPSJ Transactions on Computer Vision and Applications*, 4:53–62, Apr. 2012.

[162] Thomas Wolf, Mohammadreza Babaee, and Gerhard Rigoll. Multi-view gait recognition using 3D convolutional neural networks. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 4165–4169, September 2016. ISSN: 2381-8549.

[163] TzeWei Yeoh, Hernán E. Aguirre, and Kiyoshi Tanaka. Clothing-invariant gait recognition using convolutional neural network. In *2016 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–5, October 2016.

[164] Zifeng Wu, Yongzhen Huang, Liang Wang, Xiaogang Wang, and Tieniu Tan. A Comprehensive Study on Cross-View Gait Based Human Identification with Deep

References

CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):209–226, February 2017. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[165] Munif Alotaibi and Ausif Mahmood. Improved gait recognition based on specialized deep convolutional neural network. *Computer Vision and Image Understanding*, 164:103–110, November 2017.

[166] Wu Liu, Cheng Zhang, Huadong Ma, and Shuangqun Li. Learning Efficient Spatial-Temporal Gait Features with Deep Learning for Human Identification. *Neuroinformatics*, 16(3):457–471, October 2018.

[167] Maryam Babaee, Linwei Li, and Gerhard Rigoll. Gait Energy Image Reconstruction from Degraded Gait Cycle Using Deep Learning. In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision–ECCV 2018 Workshops*, Lecture Notes in Computer Science, pages 654–658, Cham, 2019. Springer International Publishing.

[168] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. Gait recognition invariant to carried objects using alpha blending generative adversarial networks. *Pattern Recognition*, 105:107376, September 2020.

[169] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. Gait-based human age estimation using age group-dependent manifold learning and regression. *Multimedia Tools and Applications*, 77(21):28333–28354, November 2018.

[170] Haiping Zhu, Yuheng Zhang, Guohao Li, Junping Zhang, and Hongming Shan. Ordinal distribution regression for gait-based age estimation. *Science China Information Sciences*, 63(2):120102, January 2020.

[171] S Zhang, SK Poon, K Vuong, A Sneddon, and Ct Loy. A Deep Learning-Based Approach for Gait Analysis in Huntington Disease. *Studies in Health Technology and Informatics*, 264:477–481, August 2019.

[172] Julià Camps, Albert Samà, Mario Martín, Daniel Rodríguez-Martín, Carlos Pérez-López, Joan M. Moreno Arostegui, Joan Cabestany, Andreu Català, Sheila Alcaine, Berta Mestre, Anna Prats, Maria C. Crespo-Maraver, Timothy J. Counihan, Patrick Browne, Leo R. Quinlan, Gearóid O Laighin, Dean Sweeney, Hadas Lewy, Gabriel Vainstein, Alberto Costa, Roberta Annicchiarico, Angels Bayés, and Alejandro Rodríguez-Molinero. Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit. *Knowledge-Based Systems*, 139:119–131, January 2018.

[173] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear CNN Models for Fine-Grained Visual Recognition. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1449–1457, December 2015. ISSN: 2380-7504.

[174] Angie K Reyes, Juan C Caicedo, and Jorge E Camargo. Fine-tuning Deep Convolutional Networks for Plant Recognition. *CLEF (Working Notes)*, 1391:467–475, 2015.

[175] Hakan Bilen, Basura Fernando, Efstratios Gavves, Andrea Vedaldi, and Stephen Gould. Dynamic Image Networks for Action Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3034–3042, June 2016. ISSN: 1063-6919.

[176] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-Tuning CNN Image Retrieval with No Human Annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1655–1668, July 2019. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[177] Tetsu Matsukawa and Einoshin Suzuki. Person re-identification using CNN features learned from combination of attributes. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2428–2433, December 2016.

[178] Ashnil Kumar, Jinman Kim, David Lyndon, Michael Fulham, and Dagan Feng. An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification. *IEEE Journal of Biomedical and Health Informatics*, 21(1):31–40, January 2017. Conference Name: IEEE Journal of Biomedical and Health Informatics.

[179] Nima Tajbakhsh, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway, and Jianming Liang. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, May 2016. Conference Name: IEEE Transactions on Medical Imaging.

[180] Zongwei Zhou, Jae Shin, Lei Zhang, Suryakanth Gurudu, Michael Gotway, and Jianming Liang. Fine-Tuning Convolutional Neural Networks for Biomedical Image Analysis: Actively and Incrementally. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4761–4772, July 2017. ISSN: 1063-6919.

[181] Xiaoqi Liu, Chengliang Wang, Jianying Bai, and Guobin Liao. Fine-tuning Pre-trained Convolutional Neural Networks for Gastric Precancerous Disease Classification on Magnification Narrow-band Imaging Images. *Neurocomputing*, April 2019.

[182] Guotai Wang, Wenqi Li, Maria A. Zuluaga, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sébastien Ourselin, and Tom Vercauteren. Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573, July 2018. Conference Name: IEEE Transactions on Medical Imaging.

[183] Kuang Gong, Jiahui Guan, Chih-Chieh Liu, and Jinyi Qi. PET Image Denoising Using a Deep Neural Network Through Fine Tuning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2):153–161, March 2019. Conference Name: IEEE Transactions on Radiation and Plasma Medical Sciences.

[184] J. J. Cunningham. Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation. *The American Journal of Clinical Nutrition*, 54(6):963–969, December 1991. Publisher: Oxford Academic.

[185] Zongyi Liu and S. Sarkar. Simplest representation yet for gait recognition: averaged silhouette. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, pages 211–214 Vol.4, August 2004. ISSN: 1051-4651.

References

[186] Sandy A. Ross and Jack R. Engsberg. Relationships Between Spasticity, Strength, Gait, and the GMFM-66 in Persons With Spastic Diplegia Cerebral Palsy. *Archives of Physical Medicine and Rehabilitation*, 88(9):1114–1120, September 2007.

[187] Hayato Nakao, Takahiro Yoshikawa, Tatsuya Mimura, Taketaka Hara, Katsuo Nishimoto, and Shigeo Fujimoto. Influence of Lower-extremity Muscle Force, Muscle Mass and Asymmetry in Knee Extension Force on Gait Ability in Community-dwelling Elderly Women. *Journal of Physical Therapy Science*, 18(1):73–79, 2006.

[188] Y. Makihara, T. Kimura, F. Okura, I. Mitsugami, M. Niwa, C. Aoki, A. Suzuki, D. Muramatsu, and Y. Yagi. Gait collector: An automatic gait data collection system in conjunction with an experience-based long-run exhibition. In *2016 International Conference on Biometrics (ICB)*, pages 1–8, June 2016.

[189] Laboratory for New Media 15th Exhibition "Let's Walk! The first step for innovation" | Miraikan Permanent Exhibition.

[190] D. Zhou, Y. Dai, and H. Li. Ground-plane-based absolute scale estimation for monocular visual odometry. *IEEE Transactions on Intelligent Transportation Systems*, 21(2):791–802, 2020.