



Yeong-Hyeon Byeon¹, Dohyung Kim², Jaeyeon Lee² and Keun-Chang Kwak^{1,*}

- ¹ Interdisciplinary Program in IT-Bio Convergence System, Department of Electronics Engineering, Chosun University, Gwangju 61452, Korea; pride@chosun.kr
- ² Intelligent Robotics Research Division, Electronics Telecommunications Research Institute, Daejeon 34129, Korea; dhkim008@etri.re.kr (D.K.); leejy@etri.re.kr (J.L.)
- * Correspondence: kwak@chosun.ac.kr; Tel.: +82-062-230-6086

Abstract: In modern society, the population has been aging as the lifespan has increased owing to the advancement in medical technologies. This could pose a threat to the economic system and, in serious cases, to the ethics regarding the socially-weak elderly. An analysis of the behavioral characteristics of the elderly and young adults based on their physical conditions enables silver robots to provide customized services for the elderly to counter aging society problems, laying the groundwork for improving elderly welfare systems and automating elderly care systems. Accordingly, skeleton sequences modeling the changes of the human body are converted into pose evolution images (PEIs), and a convolutional neural network (CNN) is trained to classify the elderly and young adults for a single behavior. Then, a heatmap, which is a contributed portion of the inputs, is obtained using a gradient-weighted class activation map (Grad-CAM) for the classified results, and a skeletonheatmap is obtained through a series of processes for the ease of analysis. Finally, the behavioral characteristics are derived through the difference matching analysis between the domains based on the skeleton-heatmap and RGB video matching analysis. In this study, we present the analysis of the behavioral characteristics of the elderly and young adults based on cognitive science using deep learning and discuss the examples of the analysis. Therefore, we have used the ETRI-Activity3D dataset, which is the largest of its kind among the datasets that have classified the behaviors of young adults and the elderly.

Keywords: behavioral characteristics; convolutional neural network; grad-cam; skeleton

1. Introduction

The aging population is increasing in several countries worldwide because of the advancement in medical technologies and low birth rates. Because a constant increase in the number of young adults becoming elderly has increased the interest in the elderly welfare. Furthermore, the increase in the elderly population decreases the number of adults who can engage in economic activities, and the productivity cannot keep up with it to support them. This may jeopardize the economic systems of certain countries, and in serious cases, an economic crisis may worsen the treatment of the socially-weak elderly [1–3].

Unlike when they were young, the elderly undergo deterioration of physical functions such as muscle reduction and the stooping of the back. These changes slow down the movements and responses of the elderly and cause a hindrance when performing social-functional tasks. The elderly automatically learn motion strategies to adapt to such physical changes and to increase the efficiency of activities, which finally become their behavioral characteristics. In other words, the motion strategies of young adults and the elderly are different because their physical conditions are different, that is, the behavioral characteristics are different between these two groups [4–10].

Two types of applications can be considered in the analysis of the behavioral characteristics to address the population aging problem. First, customized designing is possible



Citation: Byeon, Y.-H.; Kim, D.; Lee, J.; Kwak, K.-C. Explaining the Unique Behavioral Characteristics of Elderly and Adults Based on Deep Learning. *Appl. Sci.* 2021, *11*, 10979. https:// doi.org/10.3390/app112210979

Academic Editors: Rung-Ching Chen, Hui Yu, Hendry Hendry and Qiangfu Zhao

Received: 7 September 2021 Accepted: 17 November 2021 Published: 19 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). in the product development process for the elderly welfare; the young adults mingling with the elderly in daily life can understand, be kind, and respect the elderly. Second, a customized behavioral recognizer can be designed based on the behavioral characteristics of the elderly, and it can provide important information for the context-awareness of silver robots, improving the automation of elderly care systems. In other words, silver robots can always remain around the elderly to automatically recognize and respond to the arising situations, and this will help alleviate the population aging problem by reducing the social costs of caring for the elderly [11,12].

Horst [13] used two force plates and infrared cameras to measure the body joint angles and ground reaction forces in gaits and used deep learning for individual identifications. Furthermore, the trained model was decomposed backward using the layer-wise relevance propagation (LRP) method, thereby visualizing the variables and the time of their contributions and analyzing the results based on the gait characteristics of individual persons. Notthoff [14] systematically reviewed and analyzed the literature considering the demographic characteristics, health, and psychological factors to determine the reasons for significant differences in physical activities between individual persons over the age of 50 years. To compare the physical activities of people more than 90 years old as well as young people and the elderly, Johannsen [15] conducted a statistical analysis by measuring their total energy expenditure and the resting metabolic rate over 14 days. To clarify the relationships between the daily non-exercise activities of the young people and that of the elderly, Harris [16] measured the lean body mass and body posture over ten days and estimated the energy consumption, which was then statistically analyzed. To compare the ability to perform complex bimanual tasks between the elderly and young adults, Goble [17] analyzed the activated parts in the functional magnetic resonance imaging results. The earlier studies on the analysis of the behavioral characteristics of the elderly and young adults focused primarily on the comparisons of statistical figures on the data of the elderly and young adults. There are, however, very few studies on the analysis of behavioral characteristics, where the behaviors of the elderly and young adults are classified using deep learning in terms of cognitive science, and then the trained models are analyzed.

In this paper, we propose a method of analyzing the behavioral characteristics using explainable artificial intelligence (AI). First, we convert the skeleton data, in which the behaviors of young adults and elderly people are distinguished, into pose evolution images (PEIs) to design a model that classifies young adults and elderly people for the same behavior. This model self-learns the behavioral characteristics through the learning process that allows distinguishing young adults and the elderly. Second, if an explainable AI method is used to analyze the method of classification, the behavioral characteristics of the class can be understood. However, the skeleton data are simply a list of the joint coordinates of humans, and it is difficult to analyze the features directly because the images contain patterns that cannot be interpreted objectively, unlike regular images. Third, the feature information that contributed to the classification based on the PEI is displayed again on the skeleton space for each frame, and then the features on the skeleton space of each frame are superimposed and outputted to display the temporal features. Finally, the visualized results are compared between the data of young adults and the elderly to analyze the behavioral characteristics.

We conducted experiments using the ETRI-Activity3D dataset to investigate whether it is appropriate to analyze the behavioral characteristics using an explainable AI through the behavior classification of young adults and the elderly using the latest machine learning. In this dataset, 55 behaviors have been obtained from 50 elderly and 50 young adults through Kinect v2 sensors, and it is the second-largest dataset in the world among the related datasets. It is the first large-scale data, in which the behavioral data were collected from Kinect sensors by classifying the elderly and young adults. Analyzing behaviors by classifying the data of the elderly and young adults is important in developing silver robots that operate through human-robot interactions, which implies that they perform analysis and provide services based on the behavioral characteristics of the elderly [18].

A major contribution of our study is that the behavioral characteristics of young adults and the elderly were analyzed in terms of cognitive science using three-dimensional (3D) skeleton data by applying an explainable AI technique of a two-dimensional convolutional neural network (2D-CNN). The special features of the proposed method are as follows. First, 3D skeleton data can be easily obtained using Microsoft Kinect without a complex installation process. Second, the latest deep CNN-based explainable AI technique with significantly improved classification performance is used to learn good features and to accurately determine the features that contributed to the classification. Third, the skeleton data are data that simply model a whole human body, and the global characteristics are analyzed rather than the local characteristics for a certain joint. Finally, the obtained videobased behavior recognition database is applied in human-robot interactions in the home service robot environment.

2. Methods

2.1. Skeleton Sequence

The skeleton is a data type that facilitates efficient storage of human movements by reconstructing the human skeleton with coordinate points based on the sensor data. Kinect v2 models a human body with 25 joints including the head, arms, legs, and hips. In this study, we used a method of creating a sequence for the skeletons for multiple moments in chronological order, similar to a video, because a skeleton at a particular moment did not contain the behavioral information of the person [19,20].

2.2. Pose Evolution Image

Because temporal information, as well as spatial information, is important for an effective analysis of skeleton sequence data, conversion methods have been studied to effectively extract the appropriate information [21–23]. Here, we introduced PEI, which is a method of converting a skeleton sequence into a color image. PEI is a simple and fast transform but has less readability for humans. Although it is less readable, spatial and temporal information could be analyzed with 2D-CNN effectively. Because the body skeleton changes when a person moves in a timeline for a certain behavior, the skeleton should be continually captured in certain time intervals. A skeleton sequence generated for behavior in this manner has a 3D data type. The method to convert the 3D data into a 2D image involves projecting the 3D coordinates, as they are, on the RGB space. If the skeleton sequence is represented by $(J \times D \times T)$ as 3D data, J denotes the number of joints showing the body skeleton, D denotes the dimensionality of the coordinates representing the joints, and T denotes the number of skeleton frames over time in a temporal dimension. The dimension (D) of joint coordinates was substituted with the temporal dimension (T) to convert the skeleton sequence into an image. When the dimensionality (D) of joint coordinates was 3, one frame of the color image $(J \times T \times 3)$ was obtained through the substitution process. If the color image was normalized for each channel and the image size was linearly converted, a skeleton image was generated. Because the pre-trained 2D-CNN was designed to primarily receive three channels of RGB as inputs for image recognition, the skeleton sequence was converted into a PEI to be immediately used in the pre-trained 2D-CNN. Furthermore, by converting the skeleton sequence into a PEI, the spatial and temporal characteristics were considered using only 2D filters [23,24]. Figure 1 shows the diagram of the PEI transform.



Figure 1. Diagram of the PEI transform.

2.3. Convolutional Neural Network

The conventional image processing methods implement the signal processing process for feature extraction based on the knowledge of experts and classify the extracted features using a classifier, but CNN is an algorithm that extracts and classifies the features by itself from data. It comprises convolutional filter-applied convolutional layers that pass through the 2D space to effectively extract the features of an image, sub-sampling layers that are stable to the movements and the size changes, fully connected layers for classification, and softmax [25]. The reason we use deep learning for analyzing behavioral characteristics is that recently deep learning has shown good performance in behavioral recognition. Conventional machine learning has had worse performance than deep learning because its large scale and complex data make it difficult to analyze features and compute-heavy processing. The challenge of the black box of deep learning makes many researchers pour their interests into explainable AI. The trials and errors result in advances in explainable AI of 2D-CNN [26]. And machine learning-based or deep learning-based behavior analysis is less studied though people need it to cope with aging society in the future.

2.4. Grad-Class Activation Map

Machine learning performs recognition by determining an optimal solution through calculations over a long period of time with numerous parameters, and humans do not know the reason for the machine's good recognition results. An explainable AI is a technology that decomposes the decision-making process of such machine-learning techniques to a level that humans can understand [26].

As the machine learning model is trained using the behavioral data, the behavioral characteristics are automatically absorbed in the model. The analysis of the behavioral characteristics is facilitated by comparatively analyzing the classification process of the model for each domain. This process has the potential to analyze even the parts that humans may miss or may not know because the results are produced through complex calculation processes using the data.

A CNN is a method of extracting and classifying features by receiving 2D images as inputs, and as it has been widely used in image recognition, several studies have been conducted on an explainable AI of a CNN. A CAM shows the part of the input image that was used for classification by the neural network model. Because a fully connected layer is located at the last layer in conventional CNNs, backward estimation is difficult because the spatial information is lost for the input image. Therefore, the fully connected layer was replaced with a global average pooling layer to construct the feature vectors while maintaining the spatial information, and the weights were multiplied to obtain the output. The weights were important for the channels of the feature map immediately before the global average pooling layer, and they were used for the creation of a heatmap that displayed the information contributing to the classification [27]. The Grad-CAM used the gradient of the last convolutional layer when determining an important part of the output was differentiated using the activation function of the feature map to calculate the gradient, and the gradient propagated backward to obtain the important weights, thereby



performing the global average pooling [28]. Figure 2 presents a diagram for the analysis method of behavioral characteristics.

Figure 2. Diagram for the analysis method of behavioral characteristics.

3. Results

This section describes sequentially the overall experiments conducted for deep learningbased analysis of the behavioral characteristics of the elderly and young adults. First, the dataset used is described, and second, the behavior classification model is trained. Third, the contribution to the classification is determined through the gradient-weighted class activation map (Grad-CAM), an explainable AI, and fourth, conversion to a skeletonheatmap is performed to facilitate easy analysis based on the contribution level. Finally, the skeleton-heat maps are compared for the behavioral data of the elderly and young adults to analyze their unique behavioral characteristics.

3.1. ETRI-Activity3D Dataset

There are many datasets for action analysis such as MSR-Action3D [29], CAD-60 [30], Multiview 3D Event [31], and more described in [32]. The largest dataset for action analysis at this time is NTU RGB+D 120 [32] which includes 114,480 samples. Only the ETRI-Activity3D dataset is configured largely with the division of domains of young and elderly, which gives a chance to analyze differences between domains. For that reason, we used the ETRI-Activity3D dataset, which was constructed by the Korean Electronics and Telecommunications Research Institute (ETRI) to explain the behavioral characteristics of the elderly and young adults based on deep learning. These data are the second-largest dataset composed of 112,620 samples, which were obtained from 50 elderly and 50 young adults. The elderly included 17 males and 33 females between the ages of 64 and 88 years with an average age of 77.1 years. The young adults included 25 males and 25 females between the ages of 21 and 29 years with an average age of 23.6 years. They performed 55 behaviors in their daily life in the living room, kitchen, and bedroom of the residential apartment environment, and the data were collected using the Kinect v2. The 55 behaviors were defined by observing the frequently occurring behaviors in the daily lives of the elderly. Assuming a home service situation, four Kinect sensors were each installed at the heights of 70 cm and 120 cm, thereby obtaining the data from eight directions. The distance between the cameras and the subject was between 1.5 m and 3.5 m. The obtained data had a resolution of 1920×1080 for the color images and 512×424 for the depth images, and the skeleton information included the positions of 25 joints on a 3D space. The frame rate of data was 20. For the diversity of data, the data for each behavior, which was performed by one person at a place (living room, bedroom, or kitchen at home) or performed two or more times by changing the facing direction, were simultaneously collected from four or eight sensors for 100 persons [18].

3.2. Transfer Learning with Input of PEI

To analyze the behavioral characteristics based on cognitive science, we converted the skeleton sequences obtained using the Kinect v2 into PEIs and trained the 2D-CNN. Here, the classification of the elderly and young adults for every single behavior was performed in the training of the neural network to analyze the behavioral characteristics. ResNet-101, a pre-training model was used as the 2D-CNN, which was trained using the training options such as the Adam optimization method, a mini-batch size of 30, an initial learning rate of 0.0001, and an epoch of 20. Data index of triple increment in ETRI-Activity3D are used for test and the rest are used for training as described in [18]. When the training was completed, the 2D-CNN outputted the results of classifying whether the input data correspond to the behavior of the young adults or that of the elderly for the validation data. Figure 3 presents the classification accuracy of the elderly and young adults for each behavior. For simplicity, among the classification results, the unique behavioral characteristics of the elderly and young adults were explained for the behaviors of eating food with a fork (behavior no. 1) and pouring water into a cup (behavior no. 2). We focused on domain analysis. The accuracies are considered as a scale of how many differences are between young and elderly adults here. We consider action with higher accuracy may have more distinct differences between domains than with lower accuracy.



Figure 3. Classification accuracy (y-axis) of the elderly and young adults for 55 behaviors (x-axis). 1-eating food with a fork, 2-pouring water into a cup, 3-taking medicine, 4-drinking water, 5—putting (taking) food in (from) the fridge, 6—trimming vegetables, 7—peeling fruit, 8—using a gas stove, 9—cutting vegetables on the cutting board, 10—brushing teeth, 11—washing hands, 12—washing face, 13—wiping face with a towel, 14—putting on cosmetics, 15—putting on lipstick, 16—brushing hair, 17—blow drying hair, 18—putting on a jacket, 19—taking off a jacket, 20—putting (taking) on (off) shoes, 21-putting (taking) on (off) glasses, 22-washing the dishes, 23-vacuuming the floor, 24—scrubbing the floor with a rag, 25—wiping off the dining table, 26—rubbing up furniture, 27-spreading (folding) bedding, 28-washing a towel by hand, 29-hanging laundry, 30—looking around for something, 31—using a remote control, 32—reading a book, 33—reading a newspaper, 34—writing, 35—talking on the phone, 36—playing with a mobile phone, 37—using a computer, 38-smoking, 39-clapping, 40-rubbing face with hands, 41-doing freehand exercises, 42-doing neck roll exercises, 43-massaging a shoulder oneself, 44-taking a bow, 45-talking to each other, 46—handshaking, 47—hugging each other, 48—fighting each other, 49—waving a hand, 50—flapping a hand up and down, 51—pointing with a finger, 52—opening the door and walking in, 53—falling on the floor, 54—sitting (standing) up, 55—lying down.

3.3. Skeleton-Heatmap from Grad-CAM

The output results for the inputs of the trained ResNet are shown in a heatmap to identify the part of the input considered for the classification using the Grad-CAM method. The heatmap was displayed in warm colors (red) and cold colors (blue), and the warm colors indicated the parts that mainly contributed to the classification result. Because it was difficult to analyze the behavioral characteristics in the PEI and heatmap states, the PEI was changed back to the skeleton, and the heatmap was superimposed on the skeleton in the output. Because the meanings of axes on the heatmap were the same as those of the PEI, the *x*-axis corresponded to the frame and the *y*-axis to the joints. Based on this, a certain frame and a certain joint were represented by a corresponding color of the heatmap. Here, outputting a skeleton heatmap for each frame decreased the visibility in identifying the characteristics of the whole behavior, making the behavior analysis difficult. Therefore, the heatmap colors were outputted for only the frames and joints of warm color parts, which are important, and all the frames from the beginning to the end were superimposed to output the trajectory. This is defined as a skeleton-heatmap. Furthermore, because a skeleton-heatmap outputs all frames on one screen, there was a possibility of losing the contour of the skeleton shape, making the analysis difficult. Therefore, only the red color parts of the heat map were outputted while distinguishing the color at each joint, thereby improving the ease of analysis on determining the joint that corresponded to the heated trajectory. Figure 4 illustrates the color assignment of a skeleton on a skeleton-heatmap. The spine line is displayed in red, the right arm in green, the left arm in blue, the right leg in yellow, and the left leg in cyan. Furthermore, considering the hips as a center, the terminal parts are represented in light colors, and the center part is in dark colors. Furthermore, the output skeleton-heatmap for a single data was comparatively analyzed with the RGB video to analyze the characteristics. The behavioral characteristics of the elderly, as well as individual persons, were analyzed by comparing the same behavior between the elderly persons and comparing the same behavior between the elderly and the young adults.



Figure 4. Color assignment of skeleton on a skeleton-heatmap.

3.4. Interpreting and Understanding the Behavioral Characteristics by Comparing the Skeleton-Heatmaps between Domains

Figure 5 demonstrates the difference matching analysis of skeleton-heatmaps for the behavior of an elderly person eating food with a fork. The upper skeleton-heatmap and the lower skeleton-heatmap show similar results as far as the right hand (green) part is concerned, but the left hand (blue) is detected only on the upper skeleton-heatmap. Whereas the left hand is detected apart from the knee on the upper skeleton-heatmap, the left hand touching the left knee is detected on the lower part. Because of this difference,

the left-hand part appears heated only on the lower skeleton-heatmap, which implies that this elderly person tends to place the left hand motionless on the left knee while using the fork with the right hand. Furthermore, it can be interpreted that there is no tendency of leaving the hand in the air away from the knee.



Figure 5. Difference analysis of skeleton-heatmaps for the behavior of an elderly person eating food with a fork.

Figure 6 depicts a video matching analysis of the skeleton-heatmap for the behavior of an elderly person eating food with a spoon. Considering that the left hand appears heated following the right hand on the skeleton-heatmap, it can be interpreted that there is a tendency of using the spoon with the right hand while supporting it with the left to avoid the spilling of food.

Figure 7 illustrates a video matching analysis of the skeleton-heatmap for the behavior of an elderly person eating fruits with a fork. The left hand, right hand, and neck (red) appear heated on the skeleton-heatmap. It can be interpreted that there is a tendency of using the fork with the left hand while holding the plate with the right hand and eating with the head slightly leaned forward when the food comes near the mouth.



Figure 6. Video matching analysis of the skeleton-heatmap for the behavior of an elderly person eating food with a spoon.



Figure 7. Video matching analysis of the skeleton-heatmap for the behavior of an elderly person eating fruits with a fork.

Figure 8 demonstrates a difference matching analysis of skeleton-heatmaps between the elderly and the young adults for the behavior of eating food with a fork. The left column is for the elderly and the right column is for the young adults. Because mainly the neck and both hands appear heated on the skeleton-heatmap for the elderly, it can be noted that they tend to mainly use the neck and both hands when eating food. However, because the elbow and the shoulders also appear heated along with the hands and the neck on the skeleton-heatmap for the young adults, it can be noted that they tend to move more using the upper body compared to the movement of the elderly.

Figure 9 demonstrates a video and difference matching analysis of skeleton-heatmaps between the elderly and the young adults for the behavior of pouring water into a cup. The left box illustrates the analysis picture for the adults and the right box for the elderly. The results of analyzing the RGB videos by referring to the heated parts on the skeleton heatmaps indicate that, in general, the young adults tend to pour water with their heads straight or slightly tilted, whereas the elderly tend to pour water with their head stooping very low. It is speculated that in the process of checking the height of the water, the elderly have a habit of tilting the head in association with the poor eyesight caused by aging.



Figure 8. Difference matching analysis of skeleton-heatmaps between elderly people and young adults for the behavior of eating food with a fork.



Figure 9. Video and difference matching analysis of skeleton-heatmaps between elderly people and young adults for the behavior of pouring water into a cup.

Overall, the classification accuracy was higher for the behavior of pouring water into a cup than the behavior of eating food with a fork. In other words, the difference of behavioral characteristics between the young adults and the elderly was more evident in the behavior of pouring water into a cup in the CNN model-based cognitive science analysis.

Figure 10 demonstrates a difference matching analysis of skeleton-heatmaps between elderly people and young adults for the behavior of falling on the floor. These data contain already fallen persons on the floor without any motion. We couldn't expect the motion analysis from this data. Its classification accuracy was 64.47% which could lead us to assume that they have few differences from each other. However, we can find falling posture based on skeleton heatmaps. The elderly adults tend to lie on their side, but the young adults tend to lie on their front. This makes their hand position different as shown in skeleton heatmaps below. This tendency may be caused by instruction for actors when capturing data.



Figure 10. Difference matching analysis of skeleton-heatmaps between elderly people and young adults for the behavior of falling on the floor.

Unlike the previous 1D or 2D explanations [13,28], our method tries to explain volume data with a temporal axis using deep learning. In our knowledge, there is a similar approach using deep learning to analyze motion patterns [13]. It analyzes numerical data gotten from physical sensors using deep learning and explainable AI. It shows heated data feature lines one by one as a local body part. It concentrates on specific part detail. However, our method is to present total body parts and tracks which is more useful for understanding the entire motion scenario.

4. Discussion

We defined the characteristics that can be calculated using skeleton coordinates to analyze the difference in the behavioral characteristics between the elderly and the young adults. Next, we calculated and compared the characteristics of the elderly and the young adults for only the warm color parts of heatmaps but could not show the expected results. Therefore, we attempted a method of showing the heatmap again on the skeleton and improved the ease of analysis by creating skeleton heatmaps through a series of processes. Microsoft Kinect can be used for the analysis of the behavioral characteristics of a person, and an image-based CNN and its explainable AI technique can be used to produce the localization of features with high contributions. Furthermore, it analyzes the global characteristics for all regions of a skeleton rather than analyzing the local characteristics such as the angle and displacement of a specific joint. However, in certain behavioral data, skeletons have excessive noise or movements are extremely dynamic, making the analysis of the characteristics difficult. Because the analysis was performed by comparing the skeleton-heatmaps, the results were not absolute, and the analysis was performed for only the parts showing tendencies. For example, it was analyzed that the elderly tended to tilt the head when pouring water, but if this was because of visual factors, the tendency of tilting the head might be low in some cases because some elderly persons had good eyesight. Figure 11 shows the example of noisy and dynamic skeletons.





Dynamic motion

Dynamic skeleton

Figure 11. Example of noisy and dynamic skeletons.

5. Conclusions

In this paper, we proposed a method of analyzing the behavioral characteristics using explainable artificial intelligence. The aging population is increasing in several countries worldwide because of the advancement in medical technologies and low birth rates. It is important to analyze behavioral characteristics between elderly and young people because the aging society should be prepared scientifically to block the economic crisis. First, we converted the skeleton data, in which the behaviors of young adults and elderly people were distinguished, into PEIs to design a model that classifies young adults and elderly people for the same behavior. This model learned the behavioral characteristics through the learning process that allowed to distinguish young adults and the elderly. Second, explainable AI method was used to analyze the method of classification, the behavioral characteristics of the class could be understood. Third, the feature information that contributed to the classification based on the PEI was displayed. Finally, the visualized results were compared between the data of young adults and the elderly to analyze the behavioral characteristics. Further, we will study the method to analyze behavioral characteristics between domains more explainable to human understanding.

Author Contributions: Conceptualization, Y.-H.B. and K.-C.K.; Methodology, Y.-H.B. and K.-C.K.; Software, Y.-H.B. and K.-C.K.; Validation, Y.-H.B., D.K. and K.-C.K.; Formal Analysis, Y.-H.B. and K.-C.K.; Investigation, Y.-H.B., D.K., J.L. and K.-C.K.; Resources, D.K., J.L. and K.-C.K.; Data Curation, J.L. and D.K.; Writing—Original Draft Preparation, Y.-H.B.; Writing—Review and Editing, D.K. and K.-C.K.; Visualization, Y.-H.B., D.K. and K.-C.K.; Supervision, K.-C.K.; Project Administration, D.K. and J.L.; Funding Acquisition, D.K. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the ICT R&D program of MSIT/IITP. [2017-0-00162, Development of Human-care Robot Technology for Aging Society] (70%). This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2017R1A6A1A03015496) (30%).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Jeong, H.-S. A study on the philosophical to the problems of Korean society's aged man. J. New Korean Philos. Assoc. 2013, 71, 335–354.
- Kim, N.S.; Shi, X. A study on the effective hyo cultural analysis of elderly care of Korea. *Jpn. China Jpn. Mod. Assoc.* 2019, 66, 335–354.
- 3. Kim, J.K. A study on senior human rights in an aging society. J. Soc. Welf. Manag. Soc. 2014, 1, 1–18.
- 4. Chung, J.-S. Changes of the physical performance and bone mineral density by aging. Korean J. Phys. Educ. 2008, 47, 489–499.
- 5. Khosla, S. Pathogenesis of age-related bone loss in humans. *J. Gerontol.* **2012**, *68*, 1226–1235. [CrossRef] [PubMed]
- 6. Na, B.R.; Oh, B.-S. Aging and muscular strength in the lower limbs. Korean J. Res. Gerontol. 2020, 29, 1–24. [CrossRef]
- Kim, J.-M.; Seong, J.-S.; Seo, E.-S.; Kho, E.-K.; Lee, S.-J.; Yoo, G.-C. Anatomical and physiological changes in the aging eye. *Korean* Ophthalmic Optic. Soc. 2004, 9, 135–143.
- 8. Rosemann, S.; Thiel, C.M. The effect of age-related hearing loss and listening effort on resting state connectivity. Sci. Rep. 2019, 9, 2337.
- 9. Peters, R. Aging and the brain. *Postgrad. Med.* 2006, *82*, 84–88. [CrossRef] [PubMed]
- 10. Global Medical Knowledge. Available online: https://www.msdmanuals.com/ (accessed on 10 November 2020).
- 11. Kim, M.-K.; Cha, E.-Y. Using skeleton vector information and RNN learning behavior recognition algorithm. *J. Broadcast. Eng.* **2018**, *23*, 598–605.
- 12. Chang, J.-Y.; Hong, S.-M.; Son, D.; Yoo, H.; Ahn, H.-W. Development of real-time video surveillance system using the intelligent behavior recognition technique. *J. Inst. Internet Broadcast. Commun.* **2019**, *19*, 161–168.
- 13. Horst, F.; Lapuschkin, S.; Samek, W.; Muller, K.-R.; Schollhorn, W.I. Explaining the unique nature of individual gait patterns with deep learning. *Sci. Rep.* **2019**, *9*, 2391. [CrossRef] [PubMed]
- 14. Notthoff, N.; Reisch, P.; Gerstorf, D. Individual characteristics and physical activity in older adults: A systematic review. *Gerontology* **2017**, *63*, 443–459. [CrossRef] [PubMed]
- 15. Johannsen, D.L.; DeLany, J.P.; Frisard, M.I.; Welsch, M.A.; Rowley, C.K.; Fang, X.; Jazwinski, S.M.; Ravussin, E. Physical activity in aging: Comparison among young, aged, and nonagenarian individuals. *J. Appl. Physiol.* **2018**, *105*, 495–501. [CrossRef] [PubMed]
- Harris, A.M.; Lanningham-Foster, L.M.; McCrady, S.K.; Levine, J.A. Nonexercise movement in elderly compared with young people. Am. J. Physiol. Endocrinol. Metabol. 2007, 292, E1207–E1212. [CrossRef] [PubMed]
- Goble, D.J.; Coxon, J.P.; Impe, A.V.; Vos, J.D.; Wenderoth, N.; Swinnen, S.P. The neural control of bimanual movements in the elderly: Brain regions exhibiting age-related increases in activity, frequency-induced neural modulation, and task-specific compensatory recruitment. *Hum. Brain Mapp.* 2010, *31*, 1281–1295. [CrossRef] [PubMed]
- Jang, J.; Kim, D.; Park, C.; Jang, M.; Lee, J.; Kim, J. ETRI-activity3D: A large-scale RGB-D dataset for robots to recognize daily activities of the elderly. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 24 October–24 January 2021; pp. 10990–10997.
- Li, S.; Li, W.; Cook, C.; Zhu, C.; Gao, Y. Independently recurrent neural network (indrnn): Building a longer and deeper rnn. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 24–26 June 2018; pp. 5457–5466.
- 20. Li, C.; Zhong, Q.; Xie, D.; Pu, S. Skeleton-based action recognition with convolutional neural networks. In Proceedings of the IEEE International Conference Multimedia & Expo Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 597–600.
- Duan, H.; Zhao, Y.; Chen, K.; Shao, D.; Lin, D.; Dai, B. Revisiting skeleton-based action recognition. *arXiv* 2021, arXiv:2104.13586.
 Yan, S.; Xiong, Y.; Lin, D. Spatial temporal graph convolutional networks for skeleton-based action recognition. *arXiv* 2018, arXiv:1801.07455.
- Liu, M.; Yuan, J. Recognizing human actions as the evolution of pose estimation maps. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 24–26 June 2018; pp. 1159–1168.
- 24. Github. Available online: https://github.com/nkliuyifang/Skeleton-based-Human-Action-Recognition (accessed on 15 October 2020).
- 25. Tian, Y. Artificial intelligence image recognition method based on convolutional neural network algorithm. *IEEE Access* **2020**, *8*, 125731–125744. [CrossRef]
- 26. Ahn, J.H. XAI Dissecting Artificial Intelligence; Wikibooks: Paju, Korea, 2020; pp. 30–50.
- 27. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2921–2929.
- Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient based localization. In Proceedings of the IEEE Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
- Li, W.; Zhang, Z.; Liu, Z. Action recognition based on a bag of 3d points. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 9–14.
- Sung, J.; Ponce, C.; Selman, B.; Saxena, A. Human activity detection from rgbd images. In Proceedings of the Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 7–11 August 2011; pp. 47–55.

- 31. Wei, P.; Zhao, Y.; Zheng, N.; Zhu, S.-C. Modeling 4d human object interactions for event and object recognition. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 3272–3279.
- 32. Liu, J.; Shahroudy, A.; Perez, M.; Wang, G.; Duan, L.-Y.; Kot, A.C. *NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding*; IEEE: Piscataway, NJ, USA, 2019; Volume 42, pp. 2684–2701.