EGOCENTRIC 3D SKELETON LEARNING IN IDENTITY-AWARE DEEP LSTM NETWORK ENCODES OBESE-LIKE MOTION REPRESENTATIONS

Jea Kwon[†]
Institute for Basic Science
onlytojay@gmail.com

Moonsun Sa[†]
Institute for Basic Science
anstjsa@gmail.com

Hyewon Kim
Institute for Basic Science
wkd091937@gmail.com

Yejin Seong
Institute for Basic Science
sungyejin@ibs.re.kr

C. Justin Lee [*]
Institute for Basic Science
cjl@ibs.re.kr

ABSTRACT

Recent advancements in 3D motion capture technology are emerging as a crucial catalyst for future developments in healthcare. With obesity increasingly recognized as a significant health concern stemming from poor dietary habits, our research focuses on identifying early indicators of obesity-inducing dietary patterns using 3D time-series skeleton data. Initially, we gathered 3D time-series skeletons from mouse models with diet-induced obesity. Subsequently, we explored the effectiveness of different viewpoints for analyzing 3D skeleton data: allocentric versus egocentric perspectives. Finally, we sought to develop efficient deep recurrent network architectures. Our findings demonstrate that integrating the concept of an egocentric viewpoint into 3D skeleton data analysis, coupled with training deep LSTM networks to accurately classify identities, can effectively distinguish motion differences induced by diet between control and high-fat diet groups. This research offers a viable approach to leveraging deep learning for early detection of health risks, facilitating timely interventions and broadening the scope of healthcare technology.

1 INTRODUCTION

Recent advancements in deep learning for motion capture systems show promise in medical healthcare by accurately capturing 3D skeleton data over time (Redmon et al., 2016; Lee et al., 2021; Wang et al., 2021; Lam et al., 2023). These data are valuable for early diagnosis and remote monitoring, particularly for conditions related to motor movements (Monje et al., 2021; Delrobaei et al., 2018; Tian et al., 2024; Bruce et al., 2021). However, their potential for understanding health conditions not directly related to motor movements remains unclear.

Although 3D skeleton data offer spatiotemporal insights into dynamic movements, their complexity makes extracting meaningful information challenging (Kwon et al., 2023; Su et al., 2021). Employing allocentric and egocentric perspectives in time-series analysis enhances our understanding of behavioral patterns (Dhamanaskar et al., 2023; Grauman et al., 2022).

Obesity, a prevalent lifestyle disease, is primarily linked to poor dietary habits (Liberali et al., 2020; San-Cristobal et al., 2020). While the transition from an unhealthy diet to obesity diagnosis can take considerable time, the traditional Body Mass Index (BMI) system, which bases obesity diagnosis on weight and height, falls short in accuracy by failing to differentiate between muscle mass and body fat (Collazo-Clavell et al., 2008). Predicting the risk of diet-induced obesity (DIO) through early behavioral pattern observation presents a promising avenue for healthcare systems.

This study aims to detect early signs of obesity-like motion representations using time-series 3D skeleton data from DIO mouse models. Since diet tracking on a daily basis in clinical settings is

[*]Corresponding author [†] These authors contributed equally

challenging, we explore the potential of deep recurrent networks (DRNs) to capture motion representations for obese-like motion representations without dietary information. Inspired by previous works on deep convolutional networks (DCNs) (Kim et al., 2024; Zhou et al., 2022), we leverage DRNs to predict the identity of DIO models based on their motion data. This approach offers a promising way to identify obesity-related motion characteristics without invasive or continuous dietary monitoring.

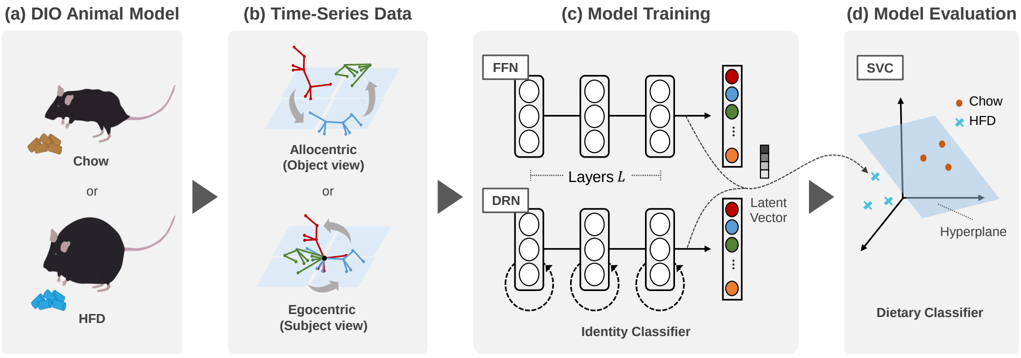

Figure 1: **Schematic illustration of our experimental design** (a) Mouse model: chow diet (top) and high-fat diet (bottom). (b) 3D skeleton data, collected from the AVATAR system, is processed into different viewpoints: allocentric (top) or egocentric (bottom) data. (c) FFN (top) and DRNs (bottom) were trained to perform identity classification tasks. (d) The latent vectors from the last hidden layer were used to test the separability of diet types(chow vs HFD) with a linear SVC.

2 Methods

DIO Animal Model The Diet-Induced Obesity (DIO) mouse model a widely utilized approach in obesity research, designed to mimic the metabolic and physiological characteristics of human obesity (Sa et al., 2023). Male C57BL/6J mice are fed with a a standard chow diet or high-fat diet (HFD) (Fig. 1a), typically consisting of 60% of total calories from fat, over a period of 6 to 15 weeks (See Appendix 5).

Time-Series Data For the acquisition of 3D obesity skeleton dataset, we used AVATAR system (Kim et al., 2022), a YOLO-based 3D pose estimation with multi-view images that extracts D [1] × V [2] time series data from multiple joint movements from freely moving mice. Both the chow group (12 mice) and the HFD group (12 mice) are subjected to weekly measurements at the same time for 9 weeks, including body weight measurements, and recorded for 10 minutes (20 FPS; 12,000 frames per session).

For generation of **allocentric** dataset (object view), the skeletons were adjusted on spatiotemporal centroid offset. For generation of **egocentric** dataset (subject view), the skeletons were adjusted on anus node offset (Fig. 1b). Each traces were randomly split with a chunk size of T. Then the dataset was randomly grouped into *train*, *valid*, and *test* datasets with 8:1:1 ratio, with different sequence lengths [3].

Model Training To address time series 3D skeleton data, we have explored the potential of DRNs (Fig. 1c): we compared feed-forward network (FFN), recurrent neural network (RNN (Bengio et al., 1994)), gated recurrent unit (GRU) (Cho et al., 2014) and long-short term memory (LSTM) (Hochreiter & Schmidhuber, 1997). With batch size of N, the inputs for FFN is given by 2D tensor $(N,\ T \times D \times V)$, and for DRNs are given by 3D tensor $(N,\ T,\ D \times V)$. Both FFN and DRNs were consist of $L = [1, 2, 3]$ number of hidden layers with 256 units per layer. The objective of these

[1] D: Number of dimensions (x, y and z). Here, $D = 3$

[2] V: Number of joints (head, limbs, tail and etc). Here, $V = 9$

[3] T: Number of frames (consecutive number of skeletons). Here, $T = [10, 20, 30, 40, 50]$

models is to predict the accurate identity (finding the source animal) from provided 3D skeleton sequences. During this phase, *train* and *valid* datasets were used.

Model Evaluation After training models, the last layer hidden activations (or latent vectors) were extracted to detect obese-like motion representations, similar to previous approach using support vector classifier (SVC)(Fig. 1d) (Zhou et al., 2022). The objective of SVC is to find a linear hyperplane that discriminates the diet group (chow vs HFD). During this phase, valid datasets were used to *train* and *test* datasets were used to evaluate.

3 Results

3.1 Obese-like motion representations emerge with identity classification

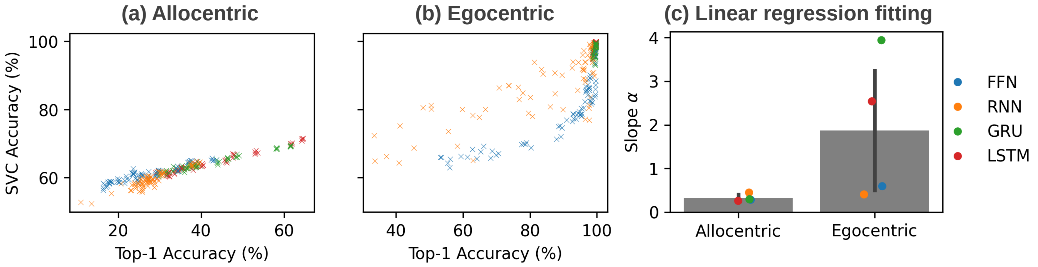

Figure 2: **Effect of allocentric and egocentric viewpoint differences** (a, b) Scatter plots of top-1 identity classification accuracy trained with (a) allocentric or (b) egocentric 3D skeletons and their diet group prediction accuracy from linear SVC; Each color represent FFN (blue), RNN (orange), GRU (green) or LSTM (red); Each dots represent a different number of hidden layers, sequences, and trials (Total 75 = 3 × 5 × 5). (c) Summary bar graphs for the slopes from linear regression ($y = \alpha x + \beta$) on scatter plots for each network.

In clinical settings, it is a complex task to monitor a patient's daily diet and simultaneously capture their motion data. Creating a comprehensive dataset for training models to identify dietary habits is challenging due to the intricacy and diversity of food intake. On the other hand, identifying which individual a particular motion belongs to—essentially, capturing the identity linked to each motion—is comparatively easier. Based on this premise, we explored the potential of deep learning models to match motion data with individual identities. This approach aimed to uncover motion representations induced by different diets.

In our study, we took advantage of SVC's ability to find optimal hyperplane that differentiates HFD and chow diet model. We show that deep neural networks, when models were trained to the task of identity classification, significantly enhance the ability to differentiate between two dietary groups, chow vs HFD (Fig. 2). These positive correlations (Pearson's correlation analysis, Top-1 Acc. vs SVC acc.; allocentric, r=0.94, $p < 0.001$; egocentric, $r = 0.78$, $p < 0.001$) imply that shared latent feature vectors encode both identity and dietary information necessary for differentiating data. Moreover, egocentric viewpoints of 3D skeletons consistently outperformed compared to allocentric ones, suggesting that subject viewpoint is beneficial in capturing underlying data structure of 3D skeletons. This finding can be seen with previous studies, supporting the idea that egocentric motion representations can contain effective information for understanding and classifying complex behaviors and traits (Dhamanaskar et al., 2023).

Next, we sought to explore the impact of viewpoint difference on the effectiveness of DRNs in time series data analysis. Through linear regression analysis, we demonstrated that allocentric data consistently exhibited strong linear relationships regardless of the model architecture used, with high goodness of fit values across different architectures (FFN, $r^2 = 0.85$; RNN, $r^2 = 0.84$; GRU, $r^2 = 0.96$; LSTM, $r^2 = 0.97$). In contrast, egocentric data generally showed a decreased effectiveness (FFN, $r^2 = 0.70$; RNN, $r^2 = 0.65$; GRU, $r^2 = 0.22$; LSTM, $r^2 = 0.74$), with a significantly larger variance in the slope values across models (Fig. 2c). These findings suggest that the choice of model architecture plays a more pronounced role when learning from egocentric 3D motion data. In summary, combining egocentric 3D skeleton data with modern DRNs can be effective in constructing a shared latent feature space for both identity and dietary classification, highlighting the delicate interplay between viewpoint and model architecture in analyzing time series data.

3.2 Identity-aware Deep LSTM encodes obese-like motion representations

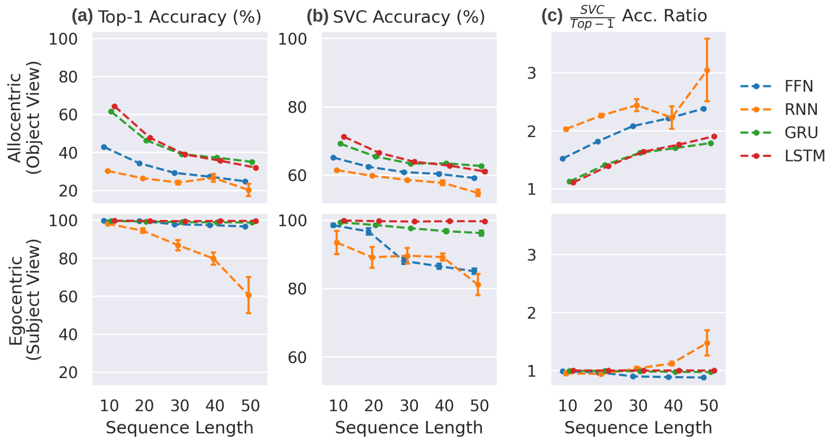

Figure 3: **Evaluation of various architectures with three hidden layers** Top-1 accuracy of identity classification task (a) or SVC accuracy of dietary classification task (b) in allocentric (top) or egocentric (bottom) viewpoints. (c) SVC to Top-1 accuracy ratio. Error bar, standard deviation of 5 trials.

While we found that an egocentric viewpoint plays a significant role in capturing shared features of both identity and dietary information with a combination of DRNs, it remains unclear how architectural difference affects the dietary classification. To investigate this, we have compared the dietary classification accuracy across models with different numbers of hidden layers and sequence lengths (Fig. 3, sFig. 6 and 7, See Appendix A.2 and A.3).

The most significant impact on identity accuracy was observed to be due to differences in viewpoint (Fig. 3a). Regarding sequence length, we noted a general trend where increasing lengths tended to decrease accuracy for both identity and dietary classifications. However, deep LSTM networks demonstrated a remarkable resilience to this performance degradation compared to other networks(Fig. 3b and c). In addition, this identity-aware LSTM was effective compared to end-to-end diet-aware LSTM (sFig. 4). These results collectively suggest that the memory cells in deep LSTM networks trained with identity may play a key role in capturing the underlying data structure which is beneficial for accurately predicting both identity and dietary habits.

Another intriguing observation is that the performance of the identity-aware LSTM, utilizing animals' 3D behavioral data, was more accurate than when using the animals' weight and period information (sFig. 5 and 8). This suggests that group distinctions based on behavioral changes induced by dietary habits manifest more quickly than changes in weight over time. This implies that in the progression of diet-induced obesity, behavioral changes may precede physical changes.

4 Conclusion

In this study, our focus centered on the detection and prediction of obesity, a growing concern in modern society. Employing novel methods, we aimed to uncover motion features indicative of dietary information through the process of identity classification. By developing a 3D skeleton identity classification network for both chow and HFD models, we extracted latent vectors and utilized an SVC to evaluate the representation of dietary features within these vectors. Our findings underscore the potential of identity-aware deep LSTM networks in identifying obese-like features from time-series 3D skeleton data, shedding light on the previously elusive association between dietary

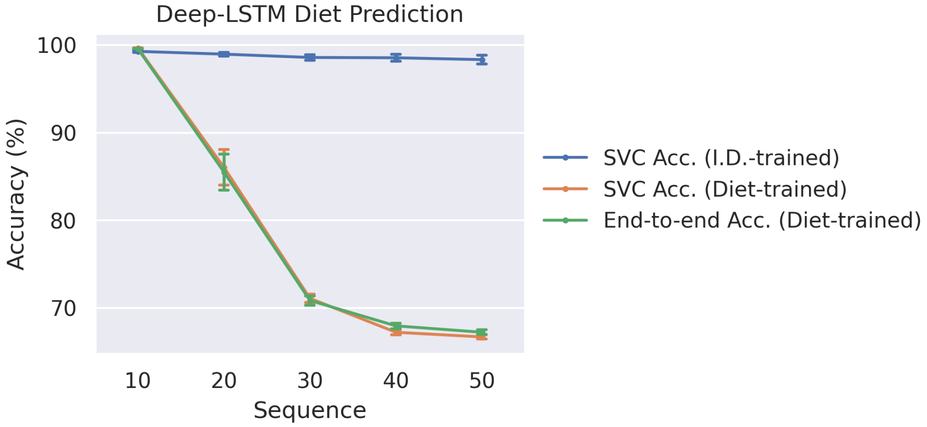

Figure 4: **Comparison with diet-aware LSTM** Diet-prediction with LSTM with 3 hidden layers trained with identity (I.D.) or diet. As shown in the figure, SVC accuracy of I.D.-trained LSTM outperforms the SVC accuracy of diet-trained LSTM or end-to-end test accuracy.

habits and skeletal movement. We foresee these models transforming clinical practice, fostering deep-learning solutions for obesity prevention and enhancing healthcare outcomes, thus promoting societal well-being.

Despite the remarkable performances demonstrated by deep LSTM networks in both identity and dietary classification tasks, our study acknowledges some limitations. Firstly, the absence of cross-subject validation. While our results showcase the LSTM's proficiency, questions persist regarding its reliance on memorization versus genuine generalization capabilities. Additionally, our dataset and task formulation have not yet been compared across various structures such as graph-based, convolutional, or transformer-based networks. Future research endeavors should prioritize investigating the LSTM's ability to generalize across subjects, or the potential of different model architectures, thus providing further validation of its effectiveness in discerning dietary-related motion patterns beyond the confines of our study's experimental scope.

Shorter sequences better represent obese-like activities In this study, we identified a notable trend in which the effectiveness of dietary predictions through Support Vector Classification (SVC) diminished as sequence lengths extended, a phenomenon consistent across several architectural frameworks. This observation indicates that short-term behavioral patterns offer a clearer reflection of dietary-induced changes than their long-term counterparts. However, the potential presence of additional interpretable aspects, such as cyclic behaviors within these patterns, remains an open question for further exploration. This area of inquiry holds promise for enriching our understanding of dietary impacts on behavior and underscores the need for future studies to delve into these complex dynamics.

Translational insights to human clinical applications Controlling genetic diversity, living environments, and dietary habits in clinical human studies is challenging, and within this context, simultaneously tracking 3D motion and dietary data to study diet-induced obesity's effects adds a significant layer of complexity. In contrast, animal models enable the acquisition of tightly controlled datasets for studying long-term obesity progression. In the context of human clinical settings, obtaining personal identity information is far simpler than accurately documenting dietary composition. This reality emphasizes the effectiveness of an identity-aware deep LSTM model in accurately identifying dietary patterns without directly specifying diet group classifications, laying a crucial groundwork for translating findings to human clinical practices.

ACKNOWLEDGMENTS

This work was supported by the Institute for Basic Science in Korea IBS-R001-D2.

REFERENCES

Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

XB Bruce, Yan Liu, Keith CC Chan, Qintai Yang, and Xiaoying Wang. Skeleton-based human action evaluation using graph convolutional network for monitoring alzheimer's progression. *Pattern Recognition*, 119:108095, 2021.

Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

ML Collazo-Clavell, JA Batsis, and FH Sert-Kuniyoshi. Accuracy of body mass index in diagnosing obesity in the adult general population. *International journal of obesity*, 32(6):959–966, 2008.

Mehdi Delrobaei, Sara Memar, Marcus Pieterman, Tyler W Stratton, Kenneth McIsaac, and Mandar Jog. Towards remote monitoring of parkinson's disease tremor using wearable motion capture systems. *Journal of the neurological sciences*, 384:38–45, 2018.

Ameya Dhamanaskar, Mariella Dimiccoli, Enric Corona, Albert Pumarola, and Francesc Moreno-Noguer. Enhancing egocentric 3d pose estimation with third person views. *Pattern Recognition*, 138:109358, 2023.

Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. Ego4d: Around the world in 3,000 hours of egocentric video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18995–19012, 2022.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Dae-Gun Kim, Anna Shin, Yong-Cheol Jeong, Seahyung Park, and Daesoo Kim. Avatar: ai vision analysis for three-dimensional action in real-time. *bioRxiv*, pp. 2021–12, 2022.

Gwangsu Kim, Dong-Kyum Kim, and Hawoong Jeong. Spontaneous emergence of rudimentary music detectors in deep neural networks. *Nature Communications*, 15(1):148, 2024.

Jea Kwon, Sunpil Kim, Dong-Kyum Kim, Jinhyeong Joo, SoHyung Kim, Meeyoung Cha, and C Justin Lee. Subtle: An unsupervised platform with temporal link embedding that maps animal behavior. *bioRxiv*, pp. 2023–04, 2023.

Winnie WT Lam, Yuk Ming Tang, and Kenneth NK Fong. A systematic review of the applications of markerless motion capture (mmc) technology for clinical measurement in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 20(1):1–26, 2023.

Lik-Hang Lee, Tristan Braud, Pengyuan Zhou, Lin Wang, Dianlei Xu, Zijun Lin, Abhishek Kumar, Carlos Bermejo, and Pan Hui. All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda. *arXiv preprint arXiv:2110.05352*, 2021.

Rafaela Liberali, Emil Kupek, and Maria Alice Altenburg de Assis. Dietary patterns and childhood obesity risk: a systematic review. *Childhood Obesity*, 16(2):70–85, 2020.

Mariana HG Monje, Sergio Domínguez, Javier Vera-Olmos, Angelo Antonini, Tiago A Mestre, Norberto Malpica, and Álvaro Sánchez-Ferro. Remote evaluation of parkinson's disease using a conventional webcam and artificial intelligence. *Frontiers in Neurology*, 12:742654, 2021.

Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

Moonsun Sa, Eun-Seon Yoo, Wuhyun Koh, Mingu Gordon Park, Hyun-Jun Jang, Yong Ryoul Yang, Mridula Bhalla, Jae-Hun Lee, Jiwoon Lim, Woojin Won, et al. Hypothalamic gabra5-positive neurons control obesity via astrocytic gaba. *Nature Metabolism*, 5(9):1506–1525, 2023.

Rodrigo San-Cristobal, Santiago Navas-Carretero, Miguel Ángel Martínez-González, José María Ordovas, and José Alfredo Martínez. Contribution of macronutrients to obesity: implications for precision nutrition. *Nature Reviews Endocrinology*, 16(6):305–320, 2020.

Yukun Su, Guosheng Lin, Ruizhou Sun, Yun Hao, and Qingyao Wu. Modeling the uncertainty for self-supervised 3d skeleton action representation learning. In *Proceedings of the 29th ACM International Conference on Multimedia*, pp. 769–778, 2021.

Haoyu Tian, Haiyun Li, Wenjing Jiang, Xin Ma, Xiang Li, Hanbo Wu, and Yibin Li. Cross-spatiotemporal graph convolution networks for skeleton-based parkinsonian gait mds-updrs score estimation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. Deep 3d human pose estimation: A review. *Computer Vision and Image Understanding*, 210:103225, 2021.

Liqin Zhou, Anmin Yang, Ming Meng, and Ke Zhou. Emerged human-like facial expression representation in a deep convolutional neural network. *Science advances*, 8(12):eabj4383, 2022.

A APPENDIX

A.1 WEIGHT PROGRESS OF ANIMAL MODEL THROUGH OBSERVED WEEKS

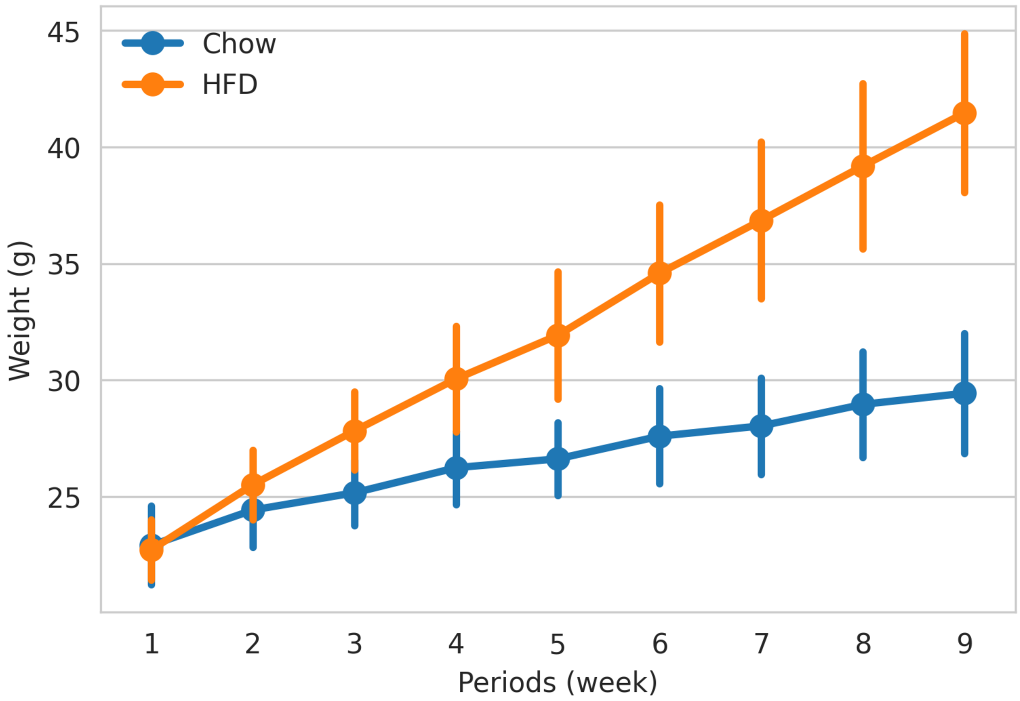

Figure 5: **Weekly weight changes in different animal groups** Comparison of weight between 12 mice in the chow diet (blue) and 12 mice in the HFD diet (orange) for duration of 9 weeks

A.2 Identity classification Top-1 accuracy

Table 1: **Allocentric identity classification top-1 accuracy (Mean±std)**

Sequence	Number of				
Length	Layers	FFN	RNN	GRU	LSTM
10	layer1	23.7±0.15	37.7±0.24	48.7±0.29	53.2±0.24
	layer2	37.3±0.35	39.2±0.65	58.2±0.10	61.8±0.17
	layer3	42.9±0.40	30.3±0.50	61.6±0.07	64.4±0.14
20	layer1	20.1±0.18	30.6±0.25	38.9±0.23	40.3±0.15
	layer2	29.9±0.21	27.1±0.76	44.0±0.21	46.0±0.17
	layer3	34.4±0.28	26.5±0.75	46.4±0.11	47.8±0.16
30	layer1	18.3±0.24	27.2±0.42	33.5±0.32	33.9±0.29
	layer2	25.6±0.32	28.5±0.86	38.0±0.15	38.1±0.26
	layer3	29.3±0.25	24.2±2.07	39.1±0.43	39.0±0.22
40	layer1	17.7±0.17	25.6±0.38	31.7±0.26	32.1±0.21
	layer2	23.8±0.28	27.2±0.46	36.2±0.23	35.0±0.45
	layer3	27.3±0.14	26.7±4.14	37.3±0.22	35.7±0.40
50	layer1	16.6±0.27	23.6±0.50	30.0±0.45	29.4±0.20
	layer2	22.0±0.38	25.4±0.65	36.4±0.71	32.0±0.36
	layer3	24.9±0.23	20.3±6.57	35.1±0.29	32.1±0.21

Table 2: **Egoocentric identity classification top-1 accuracy (Mean±std)**

Sequence	Number of				
Length	Layers	FFN	RNN	GRU	LSTM
10	layer1	88.9±0.58	99.4±0.14	99.8±0.02	99.7±0.02
	layer2	99.6±0.07	99.0±0.22	99.7±0.04	99.7±0.02
	layer3	99.8±0.03	98.2±0.45	99.7±0.04	99.8±0.03
20	layer1	78.8±0.95	98.7±0.14	99.7±0.03	99.4±0.02
	layer2	98.3±0.25	96.5±1.18	99.5±0.06	99.6±0.06
	layer3	99.7±0.06	94.5±2.44	99.3±0.04	99.7±0.03
30	layer1	68.6±1.84	97.5±0.71	99.5±0.07	98.9±0.12
	layer2	95.5±0.34	89.5±13.23	99.1±0.08	99.3±0.10
	layer3	97.9±0.19	86.9±5.39	99.1±0.13	99.5±0.05
40	layer1	62.8±2.33	82.4±14.05	99.5±0.10	99.0±0.08
	layer2	94.4±0.46	68.7±13.51	99.5±0.22	99.5±0.09
	layer3	97.5±0.19	79.9±6.44	99.0±0.16	99.7±0.08
50	layer1	55.0±1.24	47.8±10.41	99.5±0.10	98.9±0.10
	layer2	92.5±0.79	57.4±16.86	99.7±0.12	99.5±0.07
	layer3	96.7±0.15	60.6±19.15	98.9±0.16	99.7±0.05

A.3 SVC Obese-like behavior classification accuracy

Table 3: **Allocentric SVC Accuracy (Mean±std)**

Sequence Length	**Number of Layers**	**FFN**	**RNN**	**GRU**	**LSTM**
10	layer1	61.2±0.34	63.7±0.47	67.5±0.25	66.7±0.40
	layer2	64.5±0.17	63.9±1.08	69.8±0.09	68.7±0.13
	layer3	65.3±0.22	61.8±0.38	71.4±0.68	69.4±0.13
20	layer1	59.1±0.37	61.5±0.65	63.7±0.35	63.9±0.28
	layer2	62.6±0.43	60.3±0.64	65.4±0.40	65.0±0.44
	layer3	62.5±0.45	60.2±0.26	67.0±0.45	65.6±0.27
30	layer1	58.6±0.22	59.5±0.21	63.1±0.33	62.6±0.54
	layer2	61.4±0.33	59.7±0.40	63.1±0.39	64.0±0.38
	layer3	61.2±0.62	59.0±0.82	64.7±0.66	64.1±0.34
40	layer1	57.8±0.46	58.5±0.53	61.1±0.56	61.2±0.49
	layer2	60.4±0.52	58.3±0.48	61.6±0.53	61.9±0.28
	layer3	60.1±0.55	57.2±1.65	62.5±0.56	62.8±0.59
50	layer1	57.5±0.84	56.7±0.76	60.3±0.53	60.7±0.55
	layer2	59.4±0.25	56.5±0.53	60.6±0.45	62.9±0.64
	layer3	59.3±0.61	55.4±2.10	61.1±0.55	62.4±0.52

Table 4: **Egocentric SVC Accuracy (Mean±std)**

Sequence Length	**Number of Layers**	**FFN**	**RNN**	**GRU**	**LSTM**
10	layer1	73.8±1.66	99.0±0.15	98.7±0.14	98.6±0.40
	layer2	85.8±1.62	97.2±3.92	99.3±0.16	98.7±0.11
	layer3	98.5±0.82	93.4±6.79	99.8±0.01	99.3±0.12
20	layer1	69.8±0.60	95.1±2.63	97.7±0.17	97.9±0.20
	layer2	82.4±0.68	86.2±6.59	99.4±0.09	97.1±0.28
	layer3	97.0±1.79	89.3±5.73	99.8±0.08	98.5±0.39
30	layer1	67.6±1.10	91.9±4.97	97.1±0.49	96.5±0.37
	layer2	79.2±1.01	83.6±10.52	99.5±0.10	96.2±0.66
	layer3	88.0±1.84	90.1±4.32	99.7±0.03	97.8±0.15
40	layer1	66.9±1.09	83.6±10.83	96.6±0.44	95.0±0.61
	layer2	78.1±0.60	82.3±3.62	99.4±0.24	97.0±1.56
	layer3	86.3±1.62	89.1±2.15	99.7±0.04	96.6±1.12
50	layer1	64.6±1.41	69.5±5.24	95.8±1.36	94.8±1.49
	layer2	77.3±0.89	77.4±6.39	99.5±0.05	99.4±0.16
	layer3	84.9±1.20	81.5±5.91	99.6±0.10	96.4±1.27

A.4 Identity Top-1 & SVC Obese-like behavior classification accuracy ratio

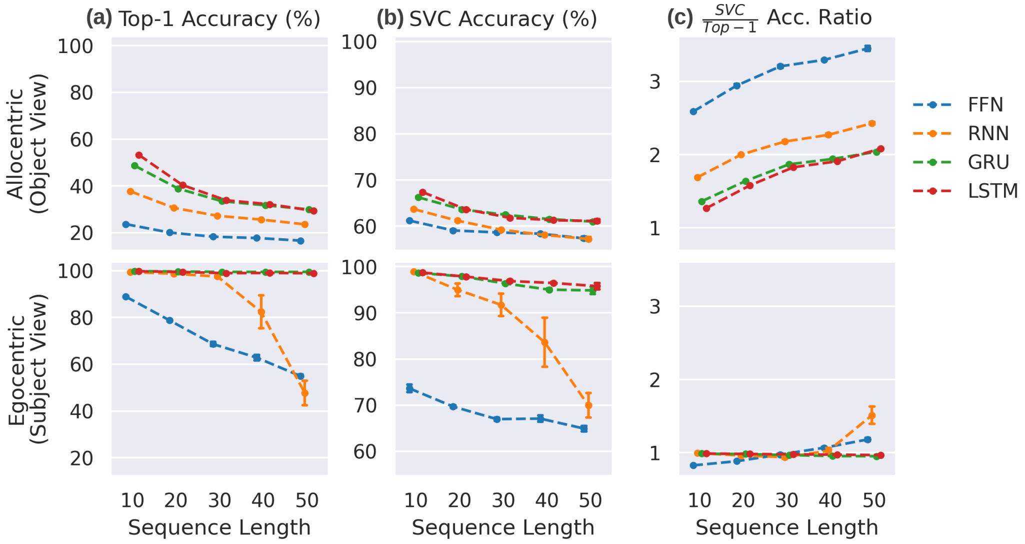

Figure 6: **Evaluation of various architectures with one hidden layer** Top-1 accuracy of identity classification task (a) or SVC accuracy of dietary classification task (b) in allocentric (top) or egocentric (bottom) viewpoints. (c) SVC to Top-1 accuracy ratio. Error bar, standard deviation of 5 trials.

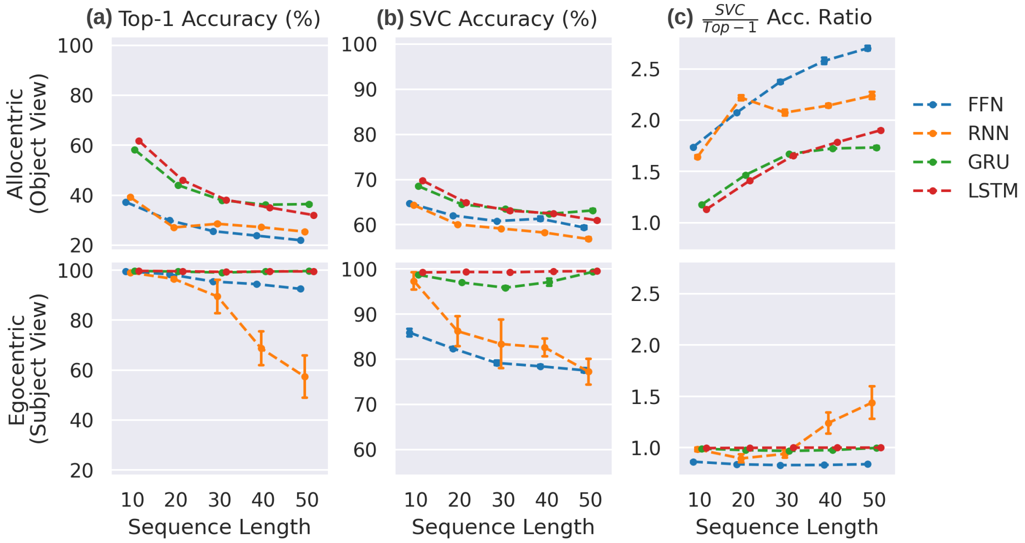

Figure 7: **Evaluation of various architectures with two hidden layers** Top-1 accuracy of identity classification task (a) or SVC accuracy of dietary classification task (b) in allocentric (top) or egocentric (bottom) viewpoints. (c) SVC to Top-1 accuracy ratio. Error bar, standard deviation of 5 trials.

A.5 Comparison of Obese-like behavior SVC classification Accuracy

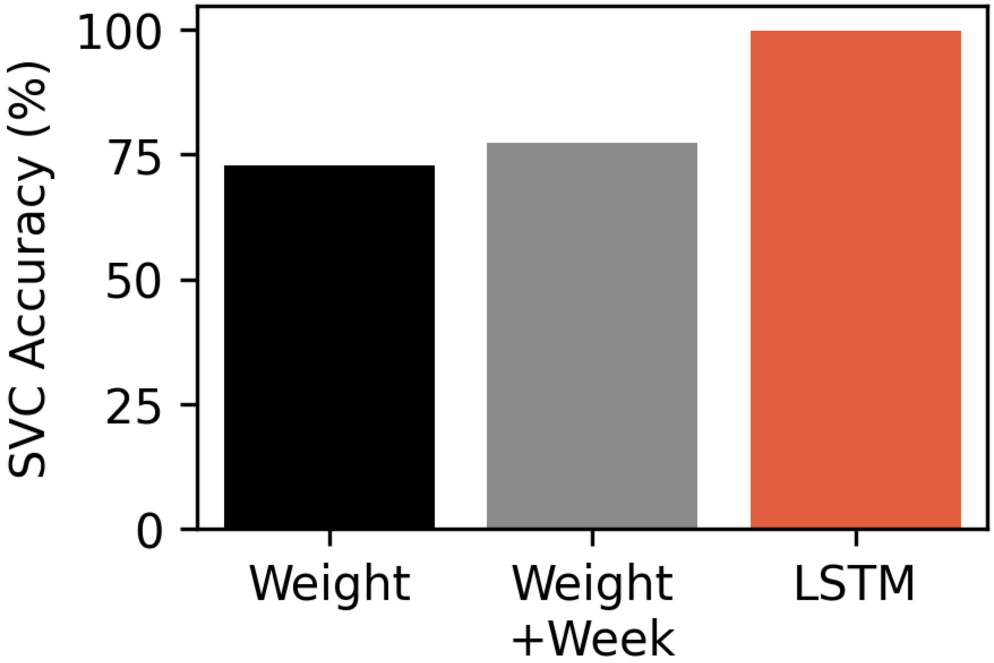

Figure 8: **LSTM effectively distinguish different dietary groups with linear SVC** Identity trained LSTM with 3 hidden layers (red) outperforms dietary group prediction when using only weight information (black) or weight and period information (grey).