Available online at
https://mijuitm.com.my

e-ISSN: 2735-0703

Mathematical Sciences and Informatics Journal 6(2) 2025, 305 – 316.

Mathematical Sciences and Informatics Journal

A Contactless Computer Vision System for Underwater Walking and Jogging Gait Analysis Using YOLO-Pose and Multi-CNN BiLSTM Architecture

Tong Bao Cheng[1*], Uswah Khairuddin[1]

[1]*Department of Mechanical Precision Engineering, Malaysia-Japan International Institute of Technology UTM, Centre for AI and Robotics UTM*

ARTICLE INFO

Article history:
Received 15 August 2025
Revised 27 August 2025
Accepted 22 September 2025
Online first
Published 31 October 2025

Keywords:
Hydrotherapy
Vision-based Gait Analysis
Deep Learning
Temporal Gait Parameters
Hyperparameter Optimization

DOI:
10.24191/mij.v6i2.9665

ABSTRACT

Buoyancy-assisted hydrotherapy exercise has been shown to reduce joint loading and accelerate functional recovery. However, conventional marker or sensor-based approaches are costly and impractical for underwater use due to water interference and setup constraints when monitoring recovery progress monitoring. To overcome these challenges, a computer vision-based gait analysis model was trained for jogging sessions in hydrotherapy pools. In this study, 2D coordinates extracted using You Only Look Once (YOLO) 11m-pose served as the model input without noise filtration to validate their robustness. A comparison of hyperparameter optimization algorithms was conducted, with the combination of multivariate tree-structured Parzen estimators (MultiTPE) and Hyperband identified as the optimal approach. Two convolutional bidirectional long short-term memory architectures, i.e., single vs. multiple convolutional layers (CNNs) per pooling were applied and compared in multi-head and single-head regression settings. Result indicated that multi-CNNs per pooling with multi-task learning best exploit inter-parameter correlations. On a 45-sample test set, the model achieved an intraclass correlation coefficient (ICC) with two-way random effects, absolute agreement, single rater model of 0.8999, Pearson's correlation coefficient (PCC) of 0.9066, mean absolute error (MAE) of 0.0954 s for swing, stance, and stride time, while 3.5141 steps/min for cadence. The developed system thus achieves precise analysis for underwater leg movements.

[1*] Corresponding author. *E-mail address*: tongbaocheng1117@gmail.com
https://doi.org/10.24191/mij.v6i2.9665

1. INTRODUCTION

Gait analysis is the study of human locomotion (Badiye et al., 2022) and is important for assisting various clinical decisions. Rehabilitation, for example, benefits from gait analysis to enable continuous monitoring so essential adjustments can be made if required (Hulleck et al., 2022). Among rehabilitation techniques, hydrotherapy which offers constant hydrostatic pressure and reduced gravitational weight, is beneficial for promoting tissue healing, improving blood circulation, relieving pain and stiffness, and enhancing mobility and posture when compared to land-based exercise (Carere & Orr, 2016). These benefits make hydrotherapy particularly useful for patient with spinal cord injuries (Ellapen et al., 2018) and stroke patients (Silva et al., 2022).

However, a continuous and efficient monitoring of hydrotherapy participants is limited because conventional sensor approaches are inapplicable due to hydrodynamics properties. For instance, optoelectronics methods are challenged by attenuation, refraction and reflection of electromagnetic waves crossing the air-water boundary, especially for the lower infrared wavelengths on which most commercial systems operate (Bernardina et al., 2016; Monoli et al., 2021). Electromyography (EMG) and ground reaction force (GRF) sensors are also constrained by subtle signal variations across the gait cycle, respectively due to water drag and buoyancy force (Barela et al., 2005). The approaches are also time consuming due to the need of calibration, which reflects the concern of increased wage expenditure, with the capture of gait is restricted only in laboratory settings due to portability concern (Ganguli, 2024; Simon, 2004). Finally, yet importantly, the accessibility of gait analysis is limited in Malaysia, when the ratio of practising physiotherapist per 10,000 residents is only 1.14 (World Physiotherapy, 2024b), about one-third of world average (World Physiotherapy, 2024a). This indicates that an automated system which is capable to reduce the physiotherapist workloads and at the same time addressing the limitations of conventional approaches when are applied in underwater settings is highly demanding.

By leveraging computer vision, 2D pose estimation models themselves are capable to extract raw keypoints coordinates, but these coordinates alone have limited clinical value, in which case they still need to be fed into a downstream module for various purposes, for example movement pattern classification, spatiotemporal gait parameters derivation, and abnormal gait flagging. In other words, pose extraction alone is not able yet to provide insight into that high-level interpretive metrics, but it must be integrated with an analytics layer that can translate those joint trajectories into actionable metrics. In this study, among the hydrotherapy exercise jogging was given focus as it can be performed effectively by individuals with varying levels of fitness and experience (Wang et al., 2021). The key contribution of this study lies in integrating pose-estimation outputs with a deep-learning predictive framework while systematically evaluating model performance, architectural variations, and optimization strategies to derive clinically meaningful gait parameters.

2. RELATED WORKS

There have already a few past studies that have integrated AI into gait analysis, using various model input such as vision-based RGB cameras, and sensor-based sensors like depth sensors, wearable IMUs and surface electromyography (sEMG). Among the choices, many studies have leveraged 2D RGB camera inputs due to their accessibility and cost-effectiveness, with different approaches used to extract essential information from videos captured in predicting gait parameters. Kidziński et al. (2020) used OpenPose to extract landmarks and use CNN to classify the gait events; Lonini et al. (2022) fine-tuned ResNet 50 on DeepLabCut platform and used CNN to predict gait parameters directly; and Barzyk et al. (2024) developed deep learning models which respectively detects individual person, predicts his landmarks' relative location to camera in 3D, and finally performs gait cycle detection. Despite of simple 2D RGB camera implementation, Boborzi et al. (2025) utilized an Azure Kinect with keypoints extracted using a combination of YOLO8 detection and Real-Time Multi-Person Pose Estimation (RTMPose), and 3D

https://doi.org/10.24191/mij.v6i2.9665

skeleton reconstructed using the depth information. The authors then performed step detection using a gait event detection algorithm, and gait parameters were determined from the detected steps. Jing et al. (2023), on other hand, used Kinect, with 3D skeleton constructed using the built-in Kinect Body Tracking software development kit and using the Kinect data to classify gait events using Bi-LSTM.

Similar patterns in how gait parameters are determined, either via calculation from classified gait events or direct prediction, can be observed with sensor-based approaches. Hartmann et al. (2025) used a single channel depth sensor as the only input to extract the required pose information and compared multiple approaches to classify gait events and estimate the gait parameters. Xu et al. (2025) used a BiTCN-BiGRU-Cross-Attention model trained using IMUs data to predict heel strike and toe-off gait events. On the other hand, Verbiest et al. (2023) and Zadka et al. (2024) used IMUs as input features, paired with CNN and XGBoost models respectively, while Liu et al. (2024) used sEMG features to compare decision tree, random forest and XGBoost, all predicting gait parameters directly.

For model validation and evaluation, most of the research are comparing their predicted gait parameters with gold-standard systems like optical motion capture, instrumented walkways, and inertial sensors, with metrics computed based on the comparison, while this research compared predicted results with annotated ground truth. The evaluation metrics that are used across the research includes MAE, RMSE and coefficient of determination (R^2), which is a common practice in model performance evaluation as performed by Mohamad Razi et al. (2022) and Sani Abdul Rahman et al. (2023); with additional PCC and ICC(2,1) included to assess the reliability and consistency of the predicted outcomes against ground-truth measurements, thereby ensuring both accuracy and agreement in model performance evaluation.

Although there are studies integrate computer vision and artificial intelligence for gait analysis, there is a gap of study when most of them do not validate the feasibility of such application for underwater settings. The limited available underwater pose data for the training of pose estimation model serves as the main research gap, as on-land pose data is easier to be obtained and hence gains more attention. Consequently, pose estimation models are trained based on on-land pose data thereby are lacked performance validation when they are applied for underwater pose estimation. Meanwhile, most aforementioned studies lack a systematic way in determining their models' hyperparameter configuration, which is concerned to have sub-optimal performance. When some of them performed hyperparameter optimization (HPO), they dealt with low-dimensional search space, that the maximum search space dimension is not more than 9, in resource-extensive ways such as grid search as done by Hartmann et al. (2025), Kidziński et al. (2020), and Verbiest et al. (2023), which is unapplicable in high-dimensional search space because it is time and resource intensive.

3. METHODOLOGY

3.1 Dataset development

The dataset was developed by recording 450 underwater jogging clips using REDBUFFALO SPARK waterproof monocular camera at 3840×2160 pixels, 30 Hz in University of Technology Malaysia Kuala Lumpur's residence pool. The data set consists of 7 male adult participants, with their movement unconstrained and unguided, and moments when participants turned their body to change direction or temporarily moved out of frame were retained to validate the model robustness. The video length ranges from 8.17 to 33.81 seconds (14.08 ± 3.63 seconds), and the videos were broadly categorized into two groups, i.e., Diagonal Jogging (DJ) and Lateral Jogging (LJ) based on their relative heading direction toward camera throughout the video clip. Examples included in Fig. 1.

https://doi.org/10.24191/mij.v6i2.9665

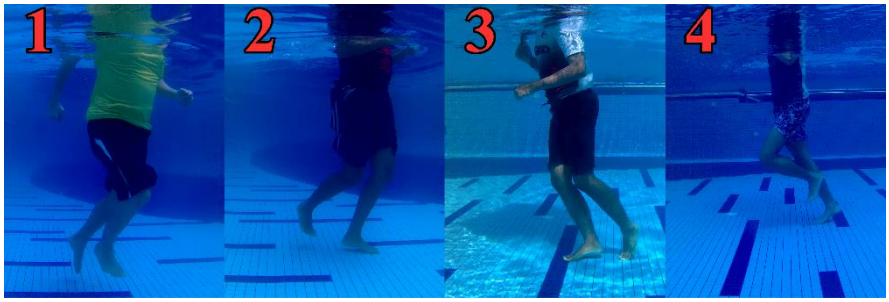

Fig. 1. Examples of frames extracted from video clips recorded. (DJ: 1 & 2; LJ: 3& 4)

A frame-to-frame motion analysis was then performed to obtain the ground truths of average temporal gait parameters: swing, stance, stride time and cadence for each video, with the reference foot used to compute the gait parameters to be right foot without considering the variability between the legs. The convention in determining the gait parameters, as shown in Table 1, follows what were described by Perry & Burnfield (2024) and Prakash et al. (2018).

Table 1. Description of temporal gait parameter

Temporal Gait Parameter	Description
Swing time	Frame number difference between when right knee starts to flex T_{flex} and when right tibia is perpendicular to floor $T_{\perp}$.
Stance time	Frame number difference between when right heel is in contact with ground T_{heel} and when right toe is in initial contact T_{toe}.
Stride time	Sum of swing time and stance time.
Cadence	Number of steps per minute (spm).

3.2 Experiment setup

Captured clips were resized to 960×540 pixels to accelerate the inference speed of pose estimation models, while BlazePose full model from Mediapipe (Lugaresi et al., 2019) and YOLO11m-pose from Ultralytics (Jocher et al., 2023) are compared in 25 randomly selected videos in term of Euclidean distance difference (EDD) between manually labelled and predicted 2D-coordinates of lower-body keypoints: bilateral hips, knees and ankles. Lower body joints were only compared and included in model training as Hartmann et al. (2025) found out including only best contributing features obtain the best prediction result. After comparison the model with better performance was used to extract the lower-body keypoints' 2D-coordinates for the remaining videos as the input features to the deep learning algorithm. Linear interpolation on missing data was only performed without noise filtration.

The CNN-BiLSTM network was built using Keras library (Watson Matthew et al., 2024), with HPO performed using Optuna library (Akiba et al., 2019). Three types of comparisons were performed throughout HPO: first is HPO algorithms comparison which involved Gaussian Process (GP), Univariate Tree-structured Parzen Estimator with Hyperband (UniTPE + Hyperband) and Multivariate Tree-structured Parzen Estimator with Hyperband (MultiTPE + Hyperband); second is CNN stacking method comparison between single CNN and multiple CNNs before each pooling layer; and last is model learning paradigm comparison between single-task learning (STL) and multi-task learning (MTL). Table 2 summarizes the 15 hyperparameters to be optimized with respective boundaries.

https://doi.org/10.24191/mij.v6i2.9665

Table 2. Optuna HPO search space

Category	Hyperparameter	Suggest type	Boundary
Data level	Sliding window size	Integer	[70, 250], step=10
	Sliding window stride	Integer	[1, 5]
	Batch size	Categorical	32, 64, 128
1D-CNNs	Number of 1D-CNNs stacks	Integer	[1, 3]
	Number of filters	Integer	[32, 128], step=32
	Number of kernels	Categorical	3, 5, 7
	Pool size	Categorical	2, 5
	CNN dropout rate	Float	[0.0, 0.5], step=0.1
	Use of batch normalization layer	Categorical	0, 1
Bi-LSTM	Number of LSTM layers	Integer	[1,3]
	Number of LSTM units	Integer	[32, 128], step=32
	Dropout rate	Float	[0.0, 0.5], step=0.1
	Dense units	Integer	[32, 128], step=32
Optimizer	Learning rate	Float	[log 1×10^{-5}, log 1×10^{-3}]
	Weight decay	Float	[log 1×10^{-6}, log 1×10^{-3}]

In HPO and as the parameter required by Optuna, the GP used *n_startup_trial*=45, following the default setting of Keras Tuner, that 3×dimensionality of hyperparameter space is used (Watson Matthew et al., 2024). In contrast, UniTPE + Hyperband and MultiTPE + Hyperband used *n_startup_trial=75*, *min_recource=1*, *max_resource= "auto"*, and *reduction_factor=3* to have a recommended Hyperband bracket number of 5 (Li et al., 2016).

Throughout this research, the prediction pipeline followed the flow described in Fig. 2, and the configuration listed in Table 3 was used. The comparisons to determine best model training strategy began by setting single CNN per pooling layer with MTL to determine the best HPO algorithm in terms of the least value of loss function obtained, where each algorithm was given 70 trials and a maximum of 100 epochs per trial for the 15-dimensional mixed-type HPO problem. After that, the best HPO algorithm was used to determine the better CNN stacking method under MTL settings, and finally comparison between MTL and STL was performed using the better CNN stacking method. These two comparisons were performed under stratified 10-fold cross-validation to estimate model generalization. Fig. 3 visualizes the comparison hierarchy.

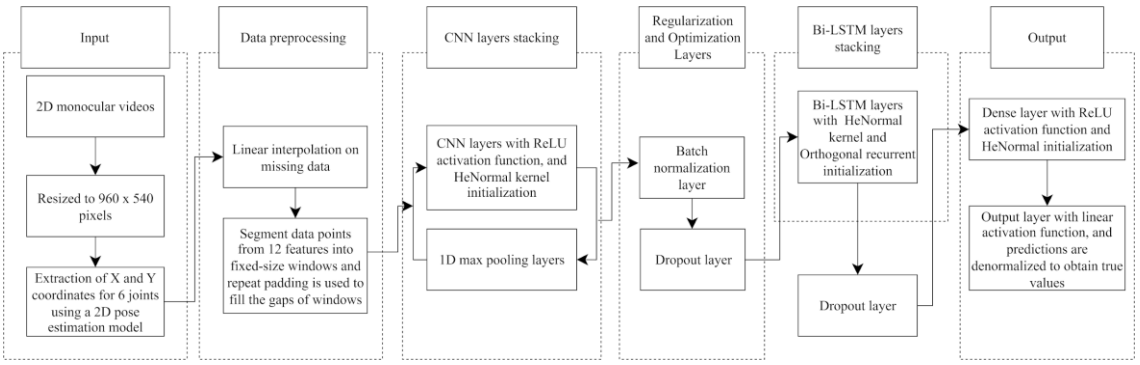

Fig. 2. Block diagram of the prediction pipeline

https://doi.org/10.24191/mij.v6i2.9665

Table 3. Configuration used for model training

Model training configuration	Setting
Train-validation-test data splitting ratio	81:9:10
Loss function	Mean squared error (MSE)
Optimizer	Adaptive Moment Estimation with Decoupled Weight Decay (ADAMW)
Normalization method	Maximum normalization
Pooling type	Maximum pooling
Early stopping	Patience of 10 epochs based on validation root mean squared error
Reduced learning rate on plateau	50%, patience of 5 epochs based on validation root mean squared error

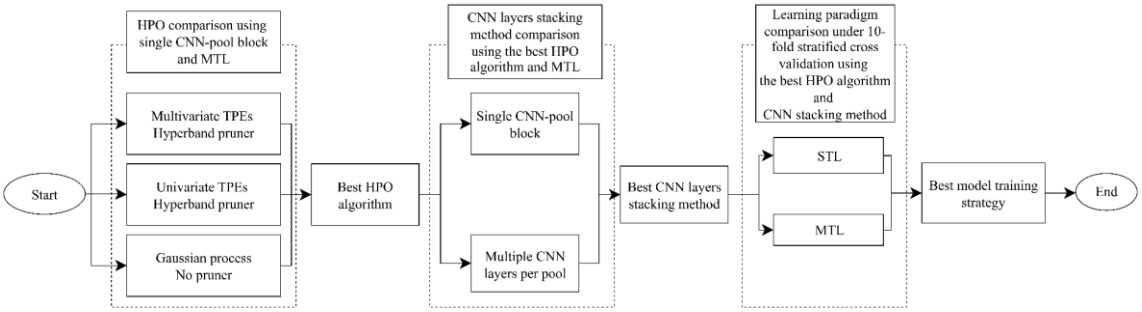

Fig. 3. Visualization of the hierarchy of comparisons

4. RESULTS AND DISCUSSION

Fig. 4 shows an example of the pipeline's output, in which the input videos are saved as copies with the predicted results displayed at the bottom. Meanwhile, the combination of pose estimator model of YOLO11m-pose, HPO algorithm of MultiTPE + Hyperband, CNN stacking method of multiple CNNs per pooling layer, and output structure of multi-head output was identified to perform the best, with a test set evaluation result of R^2 score of 0.8208; average ICC(2,1) of 0.8999, average PCC of 0.9066, and average MAE and RMSE of 0.0954 seconds and 0.1418 seconds, respectively, for swing, stance, and stride time. Meanwhile, MAE and RMSE for cadence are 3.5141 and 4.9464 spm respectively. Breakdown of the result is as shown in Table 4.

Fig. 4. Examples of the pipeline's output extracted using YOLOv11m-pose are also displayed to aid verification, as inaccurately extracted keypoints would degrade prediction quality

https://doi.org/10.24191/mij.v6i2.9665

Table 4. Detailed model performance evaluated on test set of 45 video clips

Temporal gait parameter	MAE	RMSE	PCC	ICC(2,1)	R^2 score
Swing time (seconds)	0.0758	0.0946	0.9323	0.9315	0.8208
Stance time (seconds)	0.0974	0.1580	0.8342	0.8138	
Stride time (seconds)	0.1130	0.1729	0.9078	0.9027	
Cadence (spm)	3.5141	4.9464	0.9520	0.9513	

The validity and reliability of the developed system for temporal gait parameter measurement demonstrated a high degree of agreement and reliability with the ground truth annotated by the rater, proved by high PCC) and ICC values. Specifically, the system achieved excellent reliability, with ICC values greater than or equals to 0.90 (Koo & Li, 2016), and excellent positive correlation, with PCC values greater than or equals to 0.9 (Mukaka, 2012), for three of the four parameters: cadence (ICC=0.9513, PCC=0.9520), swing time (ICC=0.9315, PCC=0.9323), and stride time (ICC=0.9027, PCC=0.9078). These results are comparable with, and in several cases surpass values reported in past literature employing different approaches in determining the gait parameters. For instance, the cadence PCC of 0.9520 and ICC of 0.9513 are competitive with the results obtained by Barzyk et al. (2024) who reported to have cadence ICC of 0.9870, and significantly exceeding the PCC of 0.79 found by Kidziński et al. (2020) and the 0.91 by Liu et al. (2024). The consistent performance of the developed model across multiple parameters confirms its capability for relative measurement, indicating its efficacy in tracking changes in gait over time, which is critical for longitudinal clinical assessment.

Apart from relative agreement measured by ICC and PCC values, the absolute accuracy of the system, which is measured using MAE, also demonstrated clinical utility. The cadence measurement achieved an MAE of 3.51 spm, which is superior to the mean error of 3.88 spm reported by Lonini et al. (2022). Similarly, the stance time MAE of 0.0974 s was lower than the mean error of 0.12 s published by Lonini et al. (2022). These small MAE values for two fundamental parameters confirm the system's high level of calibration and low systematic bias. Although the stride time RMSE of 0.1729 s was notably higher than RMSE of 0 s, achieved by Boborzi et al. (2025), the stride time ICC of 0.9027 still confirms excellent reliability, suggesting the value may be influenced by a few larger, non-systematic errors which are penalized more heavily by the metric, and are most probably caused by completely wrong detection of landmarks coordinates in certain frames due to the water disturbances and sudden lightning change.

A particularly noteworthy finding is the exceptionally high reliability observed for swing time (ICC=0.9315, PCC=0.9323) which significantly surpasses the moderate reliability values reported across the cited literature, including the PCC of 0.66 (Lonini et al., 2022) and the ICC of 0.663 (Boborzi et al., 2025). This high agreement suggests the system's event detection algorithm is highly robust in precisely identifying the toe-off and heel-strike events that define the swing phase. Conversely, the stance time showed the lowest ICC of 0.8138 among the reliable parameters and lagged behind the PCC of 0.93 as reported by Lonini et al. (2022). Stance time which is defined by the period between two complex load-bearing events is inherently challenging to capture accurately without a force plate. The slightly lower ICC may reflect the authors' fine-tuned ResNet 50 captures the subtle movement artifacts during weight transfer or push-off better, which is beneficial because minimal variability was introduced in the stance time calculation.

For the HPO algorithm comparison, GP obtained MSE of 0.11341; UniTPE + Hyperband obtained 0.004987; followed by MultiTPE + Hyperband which obtained the least MSE of 0.004981. It was expected that GP performs worse in mixed-type search space and in high-dimensional search space, whose number of hypermeters equals and more than 15. It was also validated that UniTPE + Hyperband performs worse than MultiTPE + Hyperband, as univariate TPE cannot capture the interaction effects across hyperparameters, since it assumes independence across hyperparameters and models the joint density as a product of one-dimensional kernel density estimators. Although there was no significant difference of MSE

https://doi.org/10.24191/mij.v6i2.9665

values between UniTPE + Hyperband and MultiTPE + Hyperband, a larger difference is expected in scenarios with higher hyperparameter search space dimensionality. Nevertheless, under the same settings of Hyperband as pruner, MultiTPE + Hyperband pruned unpromising trials more aggressively when compared to UniTPE + Hyperband, as shown in Table 5.

Table 5. Comparison between UniTPE + Hyperband and MultiTPE + Hyperband in term of pruned trials

HPO algorithm	Trial at optimal	Number of pruned trials before optimal	Number of total pruned trials	Number of total trials
UniTPE + Hyperband	142nd	82	96	166
MultiTPE + Hyperband	157th	109	145	215

The difference may be due to how different Optuna models the univariate TPE and multivariate TPE's kernel density estimation, where univariate TPE determines each kernel's bandwidth based on the local gap to respective nearest neighbours in the sorted data; while multivariate TPE uses a global, single kernel bandwidth based on the formula:

$$\sigma = 0.2 \times n^{-1/(d+4)} \times R \quad (1)$$

σ: kernel bandwidth
n: number of past observations
d: dimensionality of hyperparameter search space
R: parameter range normalization term

For that reason, multivariate TPE tends to have comparatively smaller bandwidth of joint density of good trials, indicating that there will be potentially more trials falls outside the promising region, and the pruner, Hyperband immediately eliminates them based on their unpromising intermediate values, hence more pruned trials. The stacking of multiple CNNs before each pooling layer achieved better result because pooling layers are destructive and by stacking up multiple CNNs can help learning complex features of input vector. In terms of model learning paradigm, MTL constantly obtained better result compared to STL, in both stratified 10-fold cross validation and test set evaluation, except for swing time when evaluated using stratified 10-fold cross validation, as shown in Table 6 and Table 7.

Table 6. Comparison between MTL and STL when evaluated using stratified 10-fold cross validation

Model learning paradigm	Swing time MAE	Stance time MAE	Stride time MAE	Cadence MAE
MTL	0.0713	0.0942	0.1109	3.6807
STL	0.0692	0.1008	0.1144	3.8628

Table 7. Comparison between MTL and STL when evaluated on test set of 45 video clips

Model learning paradigm	Swing time MAE	Stance time MAE	Stride time MAE	Cadence MAE
MTL	0.0758	0.0974	0.1130	3.5141
STL	0.0774	0.1121	0.1304	3.7939

This may be mainly due to the sharing of hidden layers, which is beneficial when the predicted tasks are closely related, as task correlation can be learnt. Through the sharing also the essential features might be amplified with noises suppressed and harder task can eavesdrop on the internal representation when one task can learn particular feature more readily. These promote learning and accuracy of the model, while with multiple objectives stochastic gradient descent will follow a path in parameter space that balances all tasks rather than overfitting on a single task, hence better generalization.

In the lower-body keypoints EDD comparison, BlazePose full model achieved an average EDD of 35.22 ± 16.94 pixels, where YOLO11m-pose achieved an average EDD of 11.52 ± 2.89 pixels when they were evaluated on 25 random video clips; indicating YOLO11m-pose constantly predicts more accurate.

https://doi.org/10.24191/mij.v6i2.9665

Although both models utilize top-down approach, the key difference is that BlazePose separates the inference pipeline into detection which uses a lightweight face detector, then followed by pose estimation performed using a pose tracker network based on the detected region of interest (ROI); while YOLO11m-pose features the combined detection and pose estimation as a single-stage head on top of its detection model which is trained on full-body boxes, making it independent on the presence of facial views. In contrast, BlazePose identifies the ROI based on the facial views, the absence of facial views especially when videos were captured underwater might misguided the ROI and this might be the root cause of having worse performance compared to YOLO11m-pose.

5. CONCLUSION

This research validated the feasibility of underwater gait analysis by incorporating computer vision with deep learning method, under a scenario which requires only minimal data preprocessing, that is, only linear interpolation for missing data was performed without input data noise filtration and removal of undesired frames. Comparisons on HPO algorithms, CNN stacking method and model learning paradigms were also performed, with the result indicating that the configuration of MultiTPE + Hyperband, stacking of multiple CNNs before pooling and MTL settings performs the best. In the case of 2D pose estimation model comparison, YOLO11m-pose shows better estimation accuracy for not relying on facial views to determine the ROI.

For the future directions, there are few aspects to be considered. Firstly, the dataset should be expanded by including participants of various health conditions and ages to introduce diverse gait patterns, thereby improves model generalizability and supports more inclusive analysis for targeted rehabilitation protocol. Moreover, multiple raters should also be involved in determining gait parameters to mitigate subjectivity, as there is no gold-standard system for gait analysis in underwater settings. Stride-level gait parameter annotation should also be adopted instead of per-video level annotation to have real-time gait analysis enabled and the granularity of gait prediction enhanced. Noise filtration methods such as Butterworth high-pass filter, median filter, and gaussian filter may be applied to reduce input noise, which is potential to ease the model training and obtain better prediction accuracy. Furthermore, although stacking of multiple CNNs before each pooling is proved to perform better, the number of CNNs stacking was predetermined to be two as recommended, which may require further studies to validate the optimal number of CNNs stacks for this specific task.

6. ACKNOWLEDGEMENTS/FUNDING

The authors would like to acknowledge Sealbotics Sdn Bhd for their technical collaboration and Dr Nadia Mustafah from the rehabilitation unit, Hospital Al Sultan Abdullah UiTM Puncak Alam for their clinical support in validating the hydrotherapy rehabilitation monitoring system.

7. CONFLICT OF INTEREST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

8. AUTHORS' CONTRIBUTIONS

Tong Bao Cheng designed, carried out the research, wrote and revised the article. Uswah Khairuddin conceptualised the central research idea, supervised research progress, anchored the review, revisions and approved the article submission.

https://doi.org/10.24191/mij.v6i2.9665

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework*. http://arxiv.org/abs/1907.10902

Badiye, A., Kathane, P., & Krishan, K. (2022). *Forensic Gait Analysis*.

Barela, A. M. F., Stolf, S. F., & Duarte, M. (2005). Biomechanical characteristics of adults walking in shallow water and on land. *Journal of Electromyography and Kinesiology*, *16*(3), 250–256. https://doi.org/10.1016/j.jelekin.2005.06.013

Barzyk, P., Boden, A. S., Howaldt, J., Stürner, J., Zimmermann, P., Seebacher, D., Liepert, J., Stein, M., Gruber, M., & Schwenk, M. (2024). Steps to facilitate the use of clinical gait analysis in stroke patients: The validation of a single 2D RGB smartphone video-based system for gait analysis. *Sensors*, *24*(23). https://doi.org/10.3390/s24237819

Bernardina, G. R. D., Cerveri, P., Barros, R. M. L., Marins, J. C. B., & Silvatti, A. P. (2016). Action sport cameras as an instrument to perform a 3D underwater motion analysis. *PLoS ONE*, *11*(8). https://doi.org/10.1371/journal.pone.0160490

Boborzi, L., Bertram, J., Schniepp, R., Decker, J., & Wuehr, M. (2025). Clinical whole-body gait characterization using a single RGB-D sensor. *Sensors*, *25*(2). https://doi.org/10.3390/s25020333

Carere, A., & Orr, R. (2016). The impact of hydrotherapy on a patient's perceived well-being: a critical review of the literature. In *Physical Therapy Reviews* (Vol. 21, Issue 2, pp. 91–101). Taylor and Francis Ltd. https://doi.org/10.1080/10833196.2016.1228510

Ellapen, T. J., Hammill, H. V., Swanepoel, M., & Strydom, G. L. (2018). The benefits of hydrotherapy to patients with spinal cord injuries. In *African Journal of Disability* (Vol. 7, pp. 1–8). AOSIS OpenJournals Publishing AOSIS (Pty) Ltd. https://doi.org/10.4102/AJOD.V7I0.450

Ganguli, P. (2024). Gait Analysis Revolution: Bridging the Gap between Forensic Science and Technology. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4878540

Hartmann, Y., Paul, R. E., Klöckner, J., Deichsel, L., & Schultz, T. (2025). Gait parameter estimation from a single depth sensor. *Journal of Smart Cities and Society*, *4*(1), 35–61. https://doi.org/10.1177/27723577251320237

Hulleck, A. A., Menoth Mohan, D., Abdallah, N., El Rich, M., & Khalaf, K. (2022). Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. In *Frontiers in Medical Technology* (Vol. 4). Frontiers Media S.A. https://doi.org/10.3389/fmedt.2022.901331

Jing, Y., Qin, P., Fan, X., Qiang, W., Wencheng, Z., Sun, W., Tian, F., & Wang, D. (2023). Deep Learning–Assisted Gait Parameter Assessment for Neurodegenerative Diseases: Model Development and Validation. *Journal of Medical Internet Research*, *25*. https://doi.org/10.2196/46427

Jocher, G., Qiu, J., & Chaurasia, A. (2023). *Ultralytics YOLO*. https://github.com/ultralytics/ultralytics

Kidziński, Ł., Yang, B., Hicks, J. L., Rajagopal, A., Delp, S. L., & Schwartz, M. H. (2020). Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-17807-z

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

https://doi.org/10.24191/mij.v6i2.9665

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2016). *Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization*. http://arxiv.org/abs/1603.06560

Liu, S. H., Ting, C. E., Wang, J. J., Chang, C. J., Chen, W., & Sharma, A. K. (2024). Estimation of Gait Parameters for Adults with Surface Electromyogram Based on Machine Learning Models. *Sensors*, *24*(3). https://doi.org/10.3390/s24030734

Lonini, L., Moon, Y., Embry, K., Cotton, R. J., McKenzie, K., Jenz, S., & Jayaraman, A. (2022). Video-Based Pose Estimation for Gait Analysis in Stroke Survivors during Clinical Assessments: A Proof-of-Concept Study. *Digital Biomarkers*, *6*(1), 9–18. https://doi.org/10.1159/000520732

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). *MediaPipe: A Framework for Building Perception Pipelines*. http://arxiv.org/abs/1906.08172

Mohamad Razi, N. F., Baharun, N., & Omar, N. (2022). Machine Learning Predictive Model of Academic Achievement Efficiency based on Data Envelopment Analysis. *Mathematical Sciences and Informatics Journal*, *3*(1), 86–99. https://doi.org/10.24191/mij.v3i1.18284

Monoli, C., Fuentez-Perez, J. F., Cau, N., Capodaglio, P., Galli, M., & Tuhtan, J. A. (2021). Land and Underwater Gait Analysis Using Wearable IMU. *IEEE Sensors Journal*, *21*(9), 11192–11202. https://doi.org/10.1109/JSEN.2021.3061623

Mukaka, M. M. (2012). Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. In *Malawi Medical Journal* (Vol. 24, Issue 3).

Perry, J., & Burnfield, J. (2024). *Gait Analysis*. CRC Press. https://doi.org/10.1201/9781003525592

Prakash, C., Kumar, R., & Mittal, N. (2018). Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review*, *49*(1), 1–40. https://doi.org/10.1007/s10462-016-9514-6

Sani Abdul Rahman, A., Yusrina Idris, A., & Abdul Rahman, S. (2023). Evaluation of Machine Learning in Predicting Air Quality Index. *Mathematical Sciences and Informatics Journal*, *4*(1), 1–10. https://doi.org/10.24191/mij.v4i1.21889

Silva, G. M. J., Dei Tos, D., & Fabiano, L. C. (2022). Effects of aquatic therapy on gait and balance in patients with brain vascular accident: a literature review. *Revista Uningá*, *59*(1), eUJ4134. https://doi.org/10.46311/2318-0579.59.eUJ4134

Simon, S. R. (2004). Quantification of human motion: Gait analysis - Benefits and limitations to its application to clinical problems. *Journal of Biomechanics*, *37*(12), 1869–1880. https://doi.org/10.1016/j.jbiomech.2004.02.047

Verbiest, J. R., Bonnechère, B., Saeys, W., Van de Walle, P., Truijen, S., & Meyns, P. (2023). Gait Stride Length Estimation Using Embedded Machine Learning. *Sensors*, *23*(16). https://doi.org/10.3390/s23167166

Wang, W., Qu, F., Li, S., & Wang, L. (2021). Effects of motor skill level and speed on movement variability during running. *Journal of Biomechanics*, *127*. https://doi.org/10.1016/j.jbiomech.2021.110680

Watson Matthew, Chollet, F., Sreepathihalli Divyashree, Saadat, S., Sampath Ramesh, Rasskin, G., and Zhu, S., Singh, V., Wood, L., Tan, Z., Stenbit, I., Qian Chen, Bischof, J., & others. (2024). *KerasHub*. https://github.com/keras-team/keras-hub

World Physiotherapy. (2024a). *Annual Membership Census 2024: Global Report*. https://world.physio/resources/publications

https://doi.org/10.24191/mij.v6i2.9665

World Physiotherapy. (2024b, June 30). *Annual Membership Census 2024: Malaysia*. https://world.physio/membership/malaysia

Xu, K., Yu, W., Yu, S., Zheng, M., & Zhang, H. (2025). The Detection of Gait Events Based on Smartphones and Deep Learning. *Bioengineering*, *12*(5). https://doi.org/10.3390/bioengineering12050491

Zadka, A., Rabin, N., Gazit, E., Mirelman, A., Nieuwboer, A., Rochester, L., Del Din, S., Pelosin, E., Avanzino, L., Bloem, B. R., Della Croce, U., Cereatti, A., & Hausdorff, J. M. (2024). A wearable sensor and machine learning estimate step length in older adults and patients with neurological disorders. *Npj Digital Medicine*, *7*(1). https://doi.org/10.1038/s41746-024-01136-2

https://doi.org/10.24191/mij.v6i2.9665