Gait-Based Soft Biometric Estimation Using Wearable Inertial Sensing: Comparative Evaluation and Demographic Fairness Analysis

Shuo Li[1], Mohamed Elgendi[2], Carlo Menon[3], Rosa H. M. Chan[4*], Edmund Y. Lam[1]

***Abstract*—The ability to infer demographic patterns from passive inertial sensing raises important privacy considerations and has potential applications in user-aware systems. In this paper, we investigated whether natural walking patterns can provide indicative demographic insights. Building upon previous work in this field, we constructed an evaluation study to compare the performance of previous methods in IMU soft biometrics, random convolution models, and InceptionTime-based models to estimate subjects' age and gender under a controlled and fair experimental setup on the OU-ISIR Gait Database. InceptionTime and H-InceptionTime obtained the best performance among all three tasks. The InceptionTime model achieved a mean absolute error of 6.60 years in age value regression, suggesting sufficient accuracy for coarse-grained demographic-aware adaptation. It also yielded a correct classification rate of 88.88% in gender classification. H-InceptionTime achieved a correct classification rate of 59.46% in the challenging four-category age group classification (notably above the 25% chance level). The experimental results demonstrate the potential of using random convolution models like HYDRA, and advanced InceptionTime-based models, to effectively extract and identify subtle demographic signatures embedded in human movement patterns measured by inertial sensors.**

***Index Terms*—gait analysis, machine learning, deep learning, biometrics, wearable sensors**

I. Introduction

In modern times, identity authentication systems are offering more and more convenience to people's daily life. Since the security and convenience of traditional identity authentication methods are often constrained, human biometrics recognition techniques have gradually taken the place of traditional methods. Soft biometrics refers to the extraction of auxiliary, non-unique information about a person, such as gender, height, age, or ethnicity. Compared with hard biometrics which directly identify individuals based on their unique physical or behavioral traits, the data collection restrictions in soft biometrics are relatively loose, thus can be adopted as the supplementary means of identity verification. Inertial measurement unit (IMU), which is composed of an accelerometer and a gyroscope, is a widely used way for human gait measurement [1]. This pervasiveness highlights engineering challenges and opportunities in processing IMU-derived gait data. A six-dimensional gait signal can be measured by placing an IMU sensor on the human body during the walking process. Studies have revealed differences in gait characteristics among various age and gender groups from a statistical point of view [2], [3]. In 2014, Ngo *et al.* [4] proposed the OU-ISIR dataset for evaluating the algorithm performance of age and gender prediction based on IMU gait traces. Developing algorithms to recognize wearer's demographic information can assist in identifying the current user of stolen electronic devices. Its potential can also be tapped in health surveillance, such as biological age testing. A summary of previous algorithms is provided in Table I.

The time series classification (TSC) problem generally refers to developing a method which can map the input continuous time series signal into a discrete response value. Some researchers systematically summarized relevant technologies for TSC based on previous research background, datasets, and mainstream algorithms [5]. Bagnall *et al.* [6] conducted a comprehensive evaluation of popular TSC models in machine learning (ML). Middlehurst *et al.* [7] focused on new models developed after 2017 and expanded the evaluation scope further. The review of Foumani *et al.* [8] summarized TSC models using deep learning (DL) approaches, which can sometimes outperform current state-of-the-art models.

Previous methods developed for IMU soft biometrics were conducted under different experimental conditions with unaligned study purposes. Key variations include the specific subject cohorts selected, the methodologies for data segmentation and preprocessing, diverse definitions of demographic categories (e.g., age group categorization [10]), and a lack of standardized evaluation against a common set of contemporary TSC benchmarks. These limitations hinder meaningful parallel comparisons across studies. In this article, we focus on constructing a comparative evaluation of previous algorithms specifically developed for IMU soft biometrics and state-of-the-art TSC models under a controlled and fair experimental condition. Besides, we also preliminarily explore the underlying demographic fairness problem, which has attracted increasing concerns in the scientific community.

[1]Shuo Li and Edmund Y. Lam are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam 999077, Hong Kong (e-mail: shuoli_eee@connect.hku.hk; elam@eee.hku.hk).

[2]Mohamed Elgendi is with the Department of Biomedical Engineering and Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, UAE (e-mail: mohamed.elgendi@ku.ac.ae).

[3]Carlo Menon is with the Biomedical and Mobile Health Technology Laboratory, Department of Health Sciences and Technology, ETH Zürich, 8008 Zürich, Switzerland (e-mail: carlo.menon@hest.ethz.ch).

[4]Rosa H. M. Chan is with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong, and State Key Laboratory of Terahertz and Millimeter Waves, City University of Hong Kong, Hong Kong, China (e-mail: rosachan@cityu.edu.hk).

*Corresponding author: Rosa H. M. Chan.

979-8-3315-7336-2/25/$31.00 ©2025 IEEE
DOI 10.1109/BioCAS67066.2025.00058

TABLE I: Summary of previous research in IMU gait-based soft biometrics. Note: SVM = support vector machine, CNN = convolutional neural network, AE-GDI = angle embedded gait dynamic image, LSTM = long short-term memory. More details can be checked in the original papers accordingly.

Author	Approach	Model	Study Purpose
Khabir *et al.* [9]	Preprocessing, segmentation, feature extraction, prediction.	SVM	Gender classification
			Age value regression
Sun *et al.* [10]	Segmentation, prediction.	CNN	Gender classification
			Age group classification
Ahad *et al.* [11]	Two-dimensional feature extraction, prediction.	AE-GDI, CNN.	Age value regression
			Gender classification
Mostafa *et al.* [12]	Autocorrelation function, prediction.	CNN	Age value regression
			Gender classification
Pathan *et al.* [13]	Preprocessing, feature extraction, feature selection, prediction	Bagging	Gender classification
Davarci *et al.* [14]	Data augmentation, feature extraction, prediction	CNN, LSTM.	Gender classification

II. Methods

A. Problem Setup

According to the previous research summarized in Table I, the scope of IMU soft biometrics involves gender classification, age value regression, and age group classification. The specific definitions of these three types of soft biometrics recognition tasks are provided in Table II. The definition of the age group classification task ensures that the age intervals within each group are strictly consistent. Subjects with age higher than 45 years old are excluded due to insufficient data volume in this age range compared to other age groups. It is worth pointing out that the definition of age group classification in our work is more stringent than that of Sun *et al.* [10], which considers only two age groups.

TABLE II: Task details of IMU gait-based soft biometrics.

Task	Description				
Age value regression	Directly predict the age value of the subject.				
Age group classification	Categorize the subject into a pre-defined age group.				
	Class Number	0	1	2	3
	Age Range	[5, 15)	[15, 25)	[25, 35)	[35, 45)
Gender classification	Predict the gender of the subject.				
	Class Number	0		1	
	Gender	Female		Male	

B. Dataset

The dataset we use is based on the OU-ISIR Gait Database [4], which was collected by Ngo *et al.* from the Institute of Scientific and Industrial Research (ISIR) of Osaka University (OU). The original dataset comprises IMU gait data

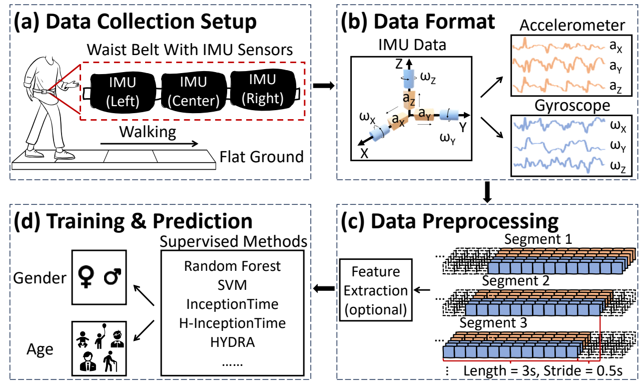

Fig. 1: The general workflow of age and gender estimation based on IMU gait traces.

of 495 subjects. The inertial sensor data was synchronously collected at three positions on the waist belt. The sampling frequency is 100Hz. For every IMU sensor of each subject, the dataset recorded two level-walk sequences. Each IMU gait sequence contains three-dimensional accelerometer data and three-dimensional gyroscope data, with the corresponding age and gender information of subjects. The data recording process in the OU-ISIR dataset is depicted in Fig. 1 (a) and (b).

To ensure that each included subject can provide sufficient data, subjects with IMU recordings shorter than 5 seconds are excluded. For each age group, we randomly sample 10 female and 10 male subjects, which is the maximum sample size while preserving data balance across different age and gender groups. From the two original level-walk sequences of each subject, the first 5 seconds are segmented into 3-second windows with a stride of 0.5 seconds to make sure the input lengths of follow-up algorithms are the same, which is shown in Fig. 1 (c). As a result, data of 80 subjects are included in the experiment, with each subject contributing 10 data segments. This approach maximally guarantees both demographic balance and sufficient data volume for analysis.

C. Previous Algorithms

There are mainly 6 methods proposed in recent years for age or gender recognition based on IMU gait data. A brief summary has been shown in Table I. Our reproduction of previous methods closely follow the implementation details in the reference papers. The method of Khabir *et al.* [9] was built on the support vector machine (SVM) algorithm with manually designed statistical features as the input. Pathan *et al.* [13] leveraged the Bagging algorithm for gender classification. Sun *et al.* [10] developed a one-dimensional convolutional neural network (1D-CNN) model with 4 convolutional layers and 2 fully connected layers to automatically extract features from the input IMU segments. Mostafa *et al.* [12] also developed a 1D-CNN model with 2 convolutional layers and 2 fully connected layers. But the model inputs were the autocorrelation features with a lag value of 10. Ahad *et al.* [11] first transformed the input IMU signals into angle embedded gait dynamics images (AE-GDI) and then developed a two-dimensional CNN (2D-CNN) with 3 convolutional layers and 2 fully connected layers. Davarci *et al.* [14] proposed a hybrid DL architecture combining CNN and long short-term memory (LSTM) layers with the input features extracted from the IMU segments using continuous wavelet transform (CWT).

In this study, the reproductions of traditional ML models are based on the Statistics and Machine Learning Toolbox in MATLAB [15]. Other four DL models are reproduced using the PyTorch Python package [24]. Due to the limited space, more details can be checked in the references accordingly.

D. Random Convolution Models

The random convolutional kernel transform (ROCKET) model [16] is a powerful approach that has been validated on the UCR TSC archive [5]. The basic idea of ROCKET is to generate a large number of convolutional kernels with random parameters. The extracted feature vectors are processed subsequently by pooling operations like computing the maximum value and the proportion of positive values (PPV). The resulting features are then concatenated into a feature vector. MiniROCKET [17] and MultiROCKET [18] are two extended models based on ROCKET. The MiniROCKET model is sped up by deleting redundant random components and only preserving a fixed set of 84 kernels. The MultiROCKET model only generates PPV features. And each kernel can produce multiple features using multiple bias values for PPV computing. The MultiROCKET model is a further extension based on ROCKET and MiniROCKET. The features are extracted using the original signal and its first order difference. Three additional pooling operations are added: mean of positive values (MPV), mean of indices of positive values (MIPV), and longest stretch of positive values (LSPV). The hybrid dictionary-ROCKET architecture (HYDRA) model [19] introduces the dictionary-based idea into ROCKET-based models to enhance both training speed and accuracy. To simplify the feature transform process, the settings of bias values are not considered in the original implementation of HYDRA so that it cannot flexibly adjust the number of extracted features.

In our experiment, each input IMU segment is a six-dimensional time sequence of 3 seconds. For the ROCKET, MiniROCKET, and MultiROCKET models, we extract 10000 features per dimension according to previous references [16]–[18]. The features extracted from each dimension are then concatenated into a feature vector of length 60000. In the HYDRA model, the number of features can not be manually adjusted, resulting in a feature vector length of 36864. The RidgeClassifierCV and RidgeCV functions in the scikit-learn Python package [20] are utilized for the classification and regression tasks, respectively. The regularization strength (alpha), is automatically determined by the built-in selection procedure, with a candidate range from 0.001 to 1000.

E. InceptionTime-based Models

According to a previous comprehensive evaluation [7] of DL models on the UCR TSC archive [5], the InceptionTime model [21] and its variant, hybrid InceptionTime (H-InceptionTime) [22], achieved the best accuracy. The InceptionTime model is a CNN model based on an ensemble of cascading Inception modules [23]. Compared with one single Inception network, ensembling Inception modules can make the training results more robust. H-InceptionTime, as the extension of InceptionTime, includes one-dimensional convolution filters with fixed parameters to make feature detection more stable and reduce the computational complexity.

We construct the InceptionTime and H-InceptionTime models based on the PyTorch Python package [24]. Considering the trade-off between model complexity and computational efficiency, the hyperparameter settings of InceptionTime and H-InceptionTime are: filter length = {10, 20, 40}, number of filters = 32, bottleneck size = 32, depth = 6. The construction of model structure is also referenced to the implementation details in the original papers to ensure that our model setup aligns with established best practices.

III. Experimental Results

In the experiment, we adopt 10-fold cross validation to guarantee the stability of results. Algorithms are assessed independently on the IMU data collected separately from each of the three waist-mounted sensor locations (left, center, and right), with mean performance reported. The pretreated data from 80 subjects in total are uniformly divided into 10 folds, with each fold containing data from 8 subjects. Within the 8 subjects of each fold, the demographic characteristics are uniformly distributed, resulting in 1 male and 1 female coming per age group. The purpose of subject-wise data stratification is to avoid data leakage problems. Each time 9 folds are used as the training set and the remaining 1 fold is used for testing. The process is repeated 10 times with different selections of training and test set. The average performance metrics are calculated over these 10 iterations. DL models are trained with the batch size of 300 and the training epochs of 1000 using the AdamX optimizer [25] with an initial learning rate of 0.0003. These training settings can ensure all models reach reliable convergence results. The random seed is fixed at 2025 for reproducibility.

Table III shows the performance of previous algorithms, random convolution models, and InceptionTime-based models evaluated in this study. For the age value regression task, we employ two performance metrics: mean absolute error (MAE) and Pearson correlation coefficient (PCC). For the age group classification and gender classification, we employ correct classification rate (CCR) as the performance metric. The method of Sun *et al.* was the most outstanding among previous algorithms. It ranked at the first place on age value regression and age group classification. For gender classification, the method of Mostafa *et al.* based on autocorrelation features achieved a slightly better accuracy than Sun *et al.*'s method. Although HYDRA achieved the best performance among random convolution models, it was falling behind Sun *et al.*'s method. Compared with other algorithms, InceptionTime-based models achieved the best performance on all three tasks.

TABLE III: Evaluation results of all included algorithms. Note: MAE = mean absolute error, PCC = Pearson correlation coefficient, CCR = correct classification rate. The “Left”, “Center”, and “Right” columns refer to the IMU sensor positions, while the “Mean” column indicates the mean performance across these sensor positions. The best mean performances among different types of algorithms are marked in bold.

Methods	Age Value Regression								Age Group Classification (CCR. %)				Gender Classification (CCR. %)			
	MAE				PCC											
	Left	Center	Right	Mean	Left	Center	Right	Mean	Left	Center	Right	Mean	Left	Center	Right	Mean
Previous Algorithms																
Khabir *et al.* [9]	10.83	11.36	12.04	11.41	0.25	0.20	0.10	0.18	33.88	29.13	36.13	33.04	63.75	67.13	61.63	64.17
Sun *et al.* [10]	7.15	7.17	6.68	**7.00**	0.52	0.53	0.67	**0.57**	58.75	55.50	61.25	**58.50**	80.25	83.38	82.63	82.08
Ahad *et al.* [11]	8.23	8.72	8.62	8.52	0.42	0.38	0.41	0.40	44.63	45.13	48.50	46.08	78.50	75.50	80.50	78.17
Mostafa *et al.* [12]	8.58	8.23	8.72	8.51	0.42	0.48	0.40	0.43	45.38	49.88	46.75	47.33	83.25	81.13	85.25	**83.21**
Pathan *et al.* [13]	8.86	9.45	9.42	9.24	0.37	0.30	0.32	0.33	39.88	34.88	37.75	37.50	58.25	64.38	59.25	60.63
Davarci *et al.* [14]	8.74	9.10	8.03	8.62	0.41	0.37	0.49	0.43	48.50	44.75	51.25	48.17	74.38	84.00	80.25	79.54
Random Convolution Models																
ROCKET	8.82	9.18	8.60	8.87	0.40	0.42	0.41	0.41	42.13	45.00	42.00	43.04	73.50	76.88	79.13	76.50
MiniROCKET	8.63	7.84	8.06	8.18	0.39	0.54	0.54	0.49	36.38	44.00	47.88	42.75	70.50	74.00	79.38	74.63
MultiROCKET	9.10	7.46	8.39	8.31	0.33	0.63	0.49	0.48	44.25	49.75	43.00	45.67	71.88	77.25	80.13	76.42
HYDRA	8.81	7.11	7.76	**7.89**	0.34	0.66	0.59	**0.53**	40.75	52.00	47.50	**46.75**	72.08	77.00	84.88	**77.99**
InceptionTime-based Models																
InceptionTime	7.24	6.52	6.04	**6.60**	0.57	0.69	0.73	**0.66**	54.63	55.00	62.13	57.25	88.13	88.63	89.88	**88.88**
H-InceptionTime	7.72	6.38	6.22	6.77	0.57	0.68	0.71	0.65	57.13	56.63	64.63	**59.46**	87.88	86.88	86.75	87.17

The mean MAE and PCC of InceptionTime on age value regression were 6.60 and 0.66. Its CCR on gender classification was 88.88%. The highest mean CCR on age group classification was achieved by H-InceptionTime at 59.46%.

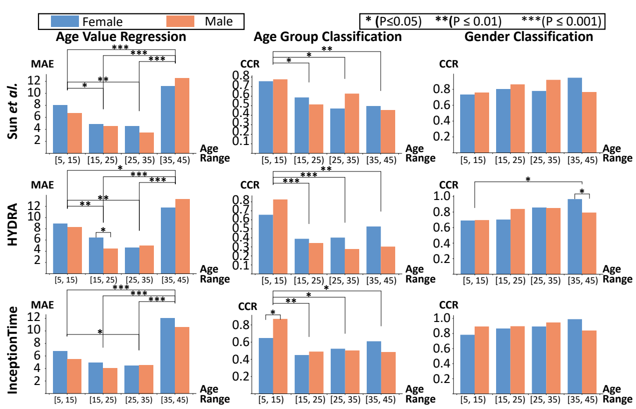

Fig. 2: Algorithm performance across demographic groups. *, **, and *** indicate P≤0.05, P≤0.01, and P≤0.001.

We further explore the performance variations across different age and gender groups to analyze the potential demographic fairness problems. Sun *et al.*'s method, HYDRA, and InceptionTime are included as representatives. The results are visualized in bar plots (Fig. 2). The two-sided Wilcoxon signed-rank tests are conducted to evaluate the significance of performance differences. In the age value regression task, algorithms generally obtained significantly worse MAE for subjects of age within [35,45) compared with subjects from other age ranges. Similar phenomenon was observed for the age group of [5,15) in the age group classification task. In comparison, the gender classification was relatively fair across different age groups. Furthermore, few significant performance differences emerged between male and female subjects.

IV. Conclusions

In this study, we conduct a comparative evaluation of age and gender recognition methods based on IMU gait signals. Together with previous algorithms, we also include new perspectives from random convolution models and InceptionTime-based models. The experimental results indicate the superior performance of InceptionTime and H-InceptionTime models in all three tasks: age value regression, age group classification, and gender classification, demonstrating the promising potential of leveraging advanced TSC models in IMU soft biometrics, with performance levels suggesting practical utility for certain demographic-aware applications, though multi-class age group classification (59.46% CCR) currently supports only coarse-grained system personalization. In application scenarios requiring precise demographic targeting, a higher accuracy would be desirable. We also quantitatively analyze the performance variation across different demographic groups, such as higher age prediction errors for the [35,45) cohort and reduced age group classification accuracy for the [5,15) range, revealing the potential concern for improving AI fairness. Future efforts of group-specific data augmentation and domain adaptation strategies for specialized subgroups are needed to address the ethical concerns of fairness. In addition, the problems of different gait styles, terrain types, footwear characteristics,weather conditions, and multi-modal identity authentication should be explored to ensure security in biometric systems.

REFERENCES

[1] I. A. Faisal, T. W. Purboyo, and A. S. R. Ansori, "A review of accelerometer sensor and gyroscope sensor in IMU sensors on motion capture," *J. Eng. Appl. Sci*, vol. 15, no. 3, pp. 826–829, 2019.

[2] E. Rowe, M. K. Beauchamp, and J. Astephen Wilson, "Age and sex differences in normative gait patterns," *Gait & Posture*, vol. 88, pp. 109–115, 2021.

[3] I. Hagoort, N. Vuillerme, T. Hortobágyi, and C. J. C. Lamoth, "Age and walking conditions differently affect domains of gait," *Human Movement Science*, vol. 89, p. 103075, 2023.

[4] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication," *Pattern Recognition*, vol. 47, no. 1, pp. 228–237, Jan. 2014.

[5] H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh, "The UCR time series archive," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293–1305, Nov. 2019.

[6] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. 31, pp. 606–660, May 2017.

[7] M. Middlehurst, P. Schäfer, and A. Bagnall, "Bake off redux: a review and experimental evaluation of recent time series classification algorithms," *Data Mining and Knowledge Discovery*, vol. 38, no. 4, pp. 1958–2031, 2024.

[8] N. Mohammadi Foumani, L. Miller, C. W. Tan, G. I. Webb, G. Forestier, and M. Salehi, "Deep learning for time series classification and extrinsic regression: A current survey," *ACM Computing Surveys*, vol. 56, no. 9, pp. 1–45, 2024.

[9] K. M. Khabir, M. S. Siraj, M. Ahmed, and M. U. Ahmed, "Prediction of gender and age from inertial sensor-based gait dataset," in *Proc. Joint 8th Int. Conf. Informatics, Electronics & Vision (ICIEV) and 3rd Int. Conf. Imaging, Vision & Pattern Recognition (icIVPR)*, 2019, pp. 371–376.

[10] Y. Sun, F. P.-W. Lo, and B. Lo, "A deep learning approach on gender and age recognition using a single inertial sensor," in *Proc. IEEE 16th Int. Conf. Wearable and Implantable Body Sensor Networks (BSN)*, 2019, pp. 1–4.

[11] M. A. R. Ahad, T. T. Ngo, A. D. Antar, M. Ahmed, T. Hossain, D. Muramatsu, Y. Makihara, S. Inoue, and Y. Yagi, "Wearable sensor-based gait analysis for age and gender estimation," *Sensors*, vol. 20, no. 8, p. 2424, Apr. 2020.

[12] A. Mostafa, S. A. Elsagheer, and W. Gomaa, "BioDeep: A Deep Learning System for IMU-based Human Biometrics Recognition," in *Proc. 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, 2021, pp. 620–629.

[13] R. K. Pathan, M. A. Uddin, N. Nahar, F. Ara, M. S. Hossain, and K. Andersson, "Gender classification from inertial sensor-based gait dataset," in *Proc. International Conference on Intelligent Computing & Optimization*, 2020, pp. 583–596.

[14] E. Davarci and E. Anarim, "Gender Detection Based on Gait Data: A Deep Learning Approach With Synthetic Data Generation and Continuous Wavelet Transform," in *IEEE Access*, vol. 11, pp. 108833-108851.

[15] The MathWorks Inc., *Statistics and Machine Learning Toolbox Documentation*, Natick, MA, USA, 2025. [Online]. Available: https://www.mathworks.com/help/stats/index.html

[16] A. Dempster, F. Petitjean, and G. I. Webb, "ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1454–1495, Sep. 2020.

[17] A. Dempster, D. F. Schmidt, and G. I. Webb, "Minirocket: A very fast (almost) deterministic transform for time series classification," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, 2021, pp. 248–257.

[18] C. W. Tan, A. Dempster, C. Bergmeir, and G. I. Webb, "MultiRocket: multiple pooling operators and transformations for fast and effective time series classification," *Data Mining and Knowledge Discovery*, vol. 36, no. 5, pp. 1623–1646, 2022.

[19] A. Dempster, D. F. Schmidt, and G. I. Webb, "Hydra: Competing convolutional kernels for fast and accurate time series classification," *Data Mining and Knowledge Discovery*, vol. 37, no. 5, pp. 1779–1805, 2023.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and others, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Nov. 2011.

[21] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P. A. Muller, and F. Petitjean, "Inceptiontime: Finding alexnet for time series classification," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, pp. 1936–1962, 2020.

[22] A. Ismail-Fawaz, M. Devanne, J. Weber, and G. Forestier, "Deep learning for time series classification using new hand-crafted convolution filters," in *Proc. IEEE International Conference on Big Data (Big Data)*, Dec. 19–22, 2022, pp. 972–981.

[23] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, 2017.

[24] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.

[25] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.