IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 28, 2022, accepted April 13, 2022, date of publication April 18, 2022, date of current version April 27, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3168019

Gait Recognition With Wearable Sensors Using Modified Residual Block-Based Lightweight CNN

MD. AL MEHEDI HASAN[1,2], FUAD AL ABIR[2], MD. AL SIAM[2], AND JUNGPIL SHIN[1], (Senior Member, IEEE)

[1]School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Fukushima 965-8580, Japan
[2]Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi 6204, Bangladesh

Corresponding author: Jungpil Shin (jpshin@u-aizu.ac.jp)

ABSTRACT Gait recognition with wearable sensors is an effective approach to identifying people by recognizing their distinctive walking patterns. Deep learning-based networks have recently emerged as a promising technique in gait recognition, yielding better performance than template matching and traditional machine learning methods. However, most recent studies have focused on improving gait detection accuracy while neglecting model complexity in the deep learning domain, making them unsuitable for low-power wearable devices. Therefore, inference from these models results in latency due to calculation overhead. This study proposes an efficient network suitable for wearable devices without sacrificing prediction performance. We have modified the residual block and accumulated it in shallow convolutional neural networks with five weighted layers only for gait recognition and proved the efficacy of all the architectural components with extensive experiments over publicly available IMU-based datasets: whuGait and OU-ISIR. Our proposed model outperforms all the state-of-the-art methods regarding recognition accuracy and is more than 85 percent efficient on average in terms of model parameters and memory consumption.

INDEX TERMS Computational efficiency, gait recognition, lightweight CNN, memory-usage reduction, parameter reduction, residual learning, wearable sensors.

I. INTRODUCTION

Biometrics is the process of automatically identifying an individual based on physiological or behavioral characteristics that are highly unique, stable, and easily obtained [1]. Physiological biometrics are concerned with the shape of the body, such as the human face [2], fingerprints [3], iris [4], etc., while behavioral biometrics are concerned with the pattern of behavior of a person, such as keystrokes, gait, signatures, etc. Many physiological biometrics have been commercially deployed. However, some of these biometrics are intrusive to users since they rely on the active participation of the users to collect data [5]. For example, users may be asked to place a finger on a gadget to take their fingerprints or stare at a camera close enough to have their irises photographed. In such circumstances, a user may feel insulted and quickly understand that his or her identity is being scrutinized [6]. Moreover, physiological biometrics has several insurmountable flaws. First, sensors for acquiring physiological characteristics (e.g., fingerprint scanners, cameras) are costly and large. Second, biometrics can be falsified and hacked in some cases. Face recognition, for example, may be tricked by using a picture or video of the target face [7], [8]. Finally, if any unlocked device is lost, there is a risk of revealing private data to strangers [9].

Gait is a behavioral biometric that refers to the walking posture of a person [10], which is very difficult to duplicate or copy [1], [2], [11]. The identity identification method based on gait is dynamic, real-time, and continuous in nature and does not require direct user participation and has a high level of security [5], [9]. In addition, as microelectronics technology has advanced, practically all the wearable intelligent devices (WIDs) have been integrated with the inertial measurement units (IMUs) because of their low cost, compact size, and low power consumption, which enables the researchers to collect gait information using the built-in IMUs in the WIDs and authenticate the users [12]. Inertial sensors, such as accelerometers and gyroscopes, are used to record the inertial data created by the movement of a walking body in inertia-based gait recognition approaches. As the sensors record gait dynamics, the inertial data effectively extracts walking patterns [13].

The associate editor coordinating the review of this manuscript and approving it for publication was Mohamad Forouzanfar.

Using IMU sensors, the effective use of wearable devices for gait recognition requires efficient recognition networks with minimal computation overhead. Therefore, we had to encounter two major challenges: (a) designing a lightweight model suitable for low-powered wearable devices; and (b) achieving state-of-the-art performance in gait recognition, which is particularly difficult for datasets with a large number of subjects. Our study overcame all these challenges, and the main contributions are as follows:

- We proposed an efficient residual convolutional neural network for gait recognition, which is suitable for wearable devices and outperforms all the state-of-the-art methods on multiple publicly available IMU-based datasets.
- We modified the residual block by using non-linear activation function before the batch normalization layer and showed the superiority of the proposed residual block by comparing it with the existing residual blocks. Finally, we designed a novel shallow convolutional neural network using the proposed modified residual block to build the lightweight model.
- Furthermore, we demonstrated the efficacy of all architectural components in the proposed lightweight model through extensive ablation, insertion, and modification experiments.

The rest of this paper is organized in the following manner: Section 2 presents previous works related to this study. The dataset is described in Section 3, along with the proposed residual learning recognition methodology. The experimental analysis and performance evaluation of the models are presented in Section 4. Finally, this paper concludes by emphasizing our contributions in Section 5.

II. RELATED WORKS

Gait recognition with sensors can be done in several ways, including with sensors on the floor, shoes, and body [14]–[16]. The inertia-sensor-based approaches are the most appealing among these methods and their variations because inertial sensors can be put on the body to collect movement characteristics in great detail [17]–[23], and the acquired time-series gait data can be used to identify and authenticate people [24], [25]. Template matching and machine learning-based methods are the two primary ways of IMU-based gait recognition [26], [27]. The user is identified using template matching methods by comparing the gait templates stored in WIDs [28]. If the resemblance exceeds a predefined level, the user is accepted as authentic. Dynamic time warping (DTW) [29], Pearson correlation coefficient (PCC) [30], and cross-correlation [31] are commonly used methods for calculating resemblance. Many earlier studies have looked at various template matching algorithms [32]–[38], and good results have been produced under controlled laboratory circumstances [39].

In their research, Ailisto *et al.* presented a signal-correlation approach for inertia-based gait identification where the recognition was done using template matching and cross-correlation computing [40], [41]. Following this research, Gafurov *et al.* made numerous significant improvements [16], [37], [42], [43]. In [43], they looked at the gait biometrics of the minimal-effort impersonation attack and the closest person attack. By inserting an accelerometer sensor into the pocket of the user, they were able to collect 300 gait sequences from 50 participants and achieve an equal error rate (EER) of 7.3% [37]. In [16], they tested user authentication using the foot, pocket, arm, and hip and discovered that a sideways motion of the foot makes the most difference and that a different portion of the gait cycle often leads to a different level of discrimination. DTW was used by Liu *et al.* to match gait curves [44]. As an improvement to this work, the wavelet denoising and gait-cycle segmentation techniques were introduced in their later work [45]. Trivino *et al.* proposed a method for modeling the perception of signal evolution using a fuzzy finite state machine (FFSM) [46]. Zhang *et al.* presented a method for avoiding cycle detection failures and phase misalignment between cycles [36]. Derawi *et al.* improved the gait-based authentication by providing a stable cycle detection mechanism [47] along with thorough comparisons.

Due to the fast growth of mobile devices in recent years, the accelerometer and gyroscope have become increasingly available on smartphones [48] and smartwatches [49]. In a variety of scenarios, such as person authentication [50]–[52], medical analysis [53]–[55], and impersonation-attack defense [56], smartphones have been used for gait recognition [57]–[59]. Data can be collected by keeping smartphones in participants' pockets [50]. However, template matching methods need to detect the gait cycles to construct the gait template, and test samples [39]. Gait cycle identification is difficult since it is sensitive to noise and device placements. Changes in pace, road conditions, and device position are all likely to produce gait cycle detection failures or inter-cycle phase misalignment, resulting in incorrect recognition results [36], [57], [60], [61]. Though there is currently no standard for manually extracting distinguishing gait features [62], Xu *et al.* proposed an adaptive preprocessing algorithm for extracting the effective components from gait data, and tested it on four publicly available datasets and three different neural networks [63]. To acquire good results, researchers must have extensive professional knowledge and experience in related domains, as well as go through data preprocessing, feature engineering, and continual experimental verification and improvement, which takes time and effort [64].

Gait recognition was performed using machine learning approaches that extracted and classified the unique properties of gait signals into separate classes [65]–[67]. For gait identification, previous research employed support vector machines (SVM) [68]–[70], k-nearest neighbors (KNN) [71]–[73], and random forests (RF) [68], [69] and found that these performed better than the template matching approaches. The manually derived features used in machine learning-based approaches had a significant effect on the

recognition accuracy of these models. In recent years, deep learning has seen much success in the fields of secure computing [74], [75], and activity recognition [76]. Gait identification based on deep learning approaches performed better than classic machine learning-based methods [57], [77], [78]. According to recent studies, the application of deep learning approaches to research gait identification has become a promising new trend [6], [26], [27], [61], [62], [79], [80]. Gadaleta and Rossi [26] used convolutional neural network (CNN) for gait recognition. They created an IDNet framework based on CNN and one-class SVM [81] for user identification and authentication, utilizing data obtained by smartphones' accelerometers and gyroscopes. The PCA [82] approach was used to lower the dimension of the gait features after they were extracted using a three-layer CNN. The features were then used to identify and authenticate users using the one-class SVM. Compared to manual feature extraction, their findings demonstrated that CNN automatically acquired more useful features and performed better. In [83], three deep CNNs were built for gait detection, utilizing the users' gait energy images (GEIs) as input. To boost classification accuracy, feature maps from different convolutional stages were combined. Deep CNNs with contrastive loss and triplet ranking loss were proposed by Takemura *et al.* for cross-view gait recognition, and better performance in person authentication and identification were obtained [84]. Elharrouss *et al.* could be able to recognize the gait with high accuracy using extracted GEIs and multitask convolutional neural network models [85]. Gul *et al.* proposed a 3D CNN architecture for gait recognition using a holistic approach in the form of GEIs [86]. Liu *et al.* presented a lightweight double-channel depthwise separable convolutional neural network (DC-DSCNN) model for gait recognition for wearable devices which could classify gait with high accuracy using a lightweight model [87].

Traditional feature selection and machine-learning methods like PCA, Bayesian classifier, and SVM can also be integrated with CNN [26], [88], [89]. In [90], CNN was used to process three-dimensional data that included pictures and optical flow information in order to recognize gait and activity. In [76], to take the temporal characteristics into account, a series of 2D images were combined into 3D data, and 3D convolution kernels were used to derive activity recognition characteristics. In the experiment performed by Donahue *et al.*, LSTMs and CNNs were integrated for activity recognition [91]. Yu *et al.* built a gait feature extractor using a generative adversarial network (GAN) to decrease the influence of view angle, weight, and clothes [92]. Chen *et al.* used the Multi-view Gait Generative Adversarial Network (MvGGAN) to produce synthetic gait samples in order to augment existing gait datasets, which provide sufficient gait samples for deep learning-based cross-view gait recognition methods [93].

Deep learning-based gait recognition methods yield better performance than template matching and machine learning methods. However, there is an apparent flaw with them: the models are too sophisticated with a large number of model parameters for wearable intelligent devices with limited computational power and capacity [6], [39]. Therefore, inference from these models result in latency due to calculation overhead. It is evident that we have to make the models more efficient by constructing simpler models with minimal computation overhead, i.e., fewer parameters without sacrificing prediction performance to make them suitable for wearable devices. Residual learning was firstly proposed for computer vision tasks to train very deep convolutional neural networks. We have successfully adopted residual learning for gait recognition, accumulated in shallow networks with a minimal number of parameters, and proved the efficacy of all the components of the architecture with extensive experiments. Our proposed model has achieved state-of-the-art accuracy in multiple widely used IMU sensor-based gait datasets and is more than 85% efficient on average in terms of parameter and memory consumption than the latest gait recognition study [39].

III. MATERIALS AND METHODS

A. DATASET DESCRIPTION

We used two publicly available gait datasets, namely the whuGait dataset and the OU-ISIR dataset. Both of them are sensor data collected using accelerometer and gyroscope. Ngo *et al.* from Osaka University released the OU-ISIR dataset, which is the most extensive public gait dataset with the most number of subjects [94]. Zou *et al.* from Wuhan University provided the whuGait dataset. They also preprocessed both of the datasets, benchmarked with train set and independent test set and shared them in their research [6].

The whuGait dataset comprises inertial data from 118 persons acquired using smartphones in an unconstrained environment with no knowledge of when, where, or how the subjects walked. The sensors sampled at 50 Hz, with each sample containing 3-axis accelerometer and 3-axis gyroscope data [6]. In this study, we have used four whuGait datasets namely Dataset #1 to Dataset #4 for gait recognition. The number of participants, data segmentation method, whether or not the samples overlap, and sample size are the most significant variations across the four datasets.

The OU-ISIR dataset consists of 744 subjects. Among them, 389 are males, and 355 are females, with a broad age range of 2 to 78 years. This dataset recorded Gait signals at 100 Hz via a waist belt-mounted central IMU. On a flat area, each of the 744 participants walked for 9 meters [6]. The 3-axis acceleration and 3-axis angular velocity data were acquired from accelerometer and gyroscope. Summary of the datasets used in this study is provided in Table 1.

B. PROPOSED METHODS FOR GAIT RECOGNITION

1) MODIFIED RESIDUAL BLOCK BASED LIGHTWEIGHT CNN

Since the inception of the residual learning, residual blocks and skip connections have been an integral part of training very deep neural networks with hundred to thousand

TABLE 1. Summary description of the datasets used in this study [6], [39].

Dataset Name	No. of Subjects	Data Segmentation Method	Overlap of Samples	Samples for Training	Samples for Test
Dataset #1	118	Gait cycle based segmentation (two gait cycles as a sample)	50%	33,104	3740
Dataset #2	20	Gait cycle based segmentation (two gait cycles as a sample)	0%	44,339	4936
Dataset #3	118	Fixed length based segmentation (sample length = 128)	50%	26,283	2991
Dataset #4	20	Fixed length based segmentation (sample length = 128)	0%	35,373	3941
OU-ISIR	744	Fixed length based segmentation (sample length = 128)	61%	13,212	1409

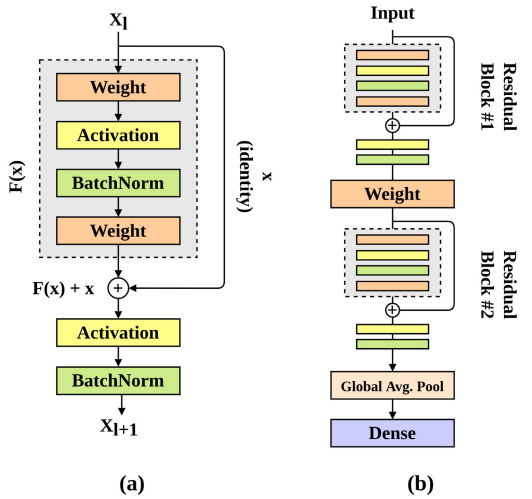


FIGURE 1. High-level architecture of the proposed (a) residual block and (b) the residual convolutional neural network.

layers [95], [96]. Besides very deep architectures, residual learning can be effective and efficient in shallow architectures in many causes. In this study, we have designed a lightweight residual convolutional neural network with two residual blocks using a total of five convolutional layers only. The proposed residual block is created by modifying the original residual block [95]. Contrasting with [95], we have used non-linear activation before the batch normalization operation. We assume that if we insert the batch normalization layer before an activation layer, the batch normalization layer may fully control the statistics of the input going into the next layer and yields better accuracy. Therefore, we have proposed our residual block with batch normalization after non-linearity. In order to incorporate non-linearity, we have investigated the performance of four widely used activation functions: Rectified Linear Unit (ReLU) [97], Exponential Linear Unit (ELU) [98], Leaky Rectified Linear Unit (LeakyReLU) [99] and Parametric Rectified Linear Unit (PReLU) [100] in our study.

The high-level architecture of the proposed residual block and the model is depicted in Fig. 1. In the model, two subsequent residual blocks incorporate a weight layer in the middle to increase the feature maps. Here, the weight layers are the one-dimensional convolution layers, and the activation refers to the various non-linearity from the ReLU family. The

TABLE 2. Network architecture of the proposed residual convolutional neural network.

Operational Group	Layer Name	Operation*	Output Shape*	No. of Params.*	Connected to
-	input	-	128 × 6	0	-
Residual Block #1	conv1_1	3 conv, 6	128 × 6	114	input
	leakyReLU1_1	-	128 × 6	0	conv1_1
	batchNorm1_1	-	128 × 6	12	leakyReLU1_1
	conv1_2	3 conv, 6	128 × 6	114	batchNorm1_1
	add_1	-	128 × 6	0	input, conv1_2
	leakyReLU1_2	-	128 × 6	0	add_1
	batchNorm1_2	-	128 × 6	12	leakyReLU1_2
Increase Feature Maps	conv2_0	3 conv, 64	128 × 64	1,216	batchNorm1_2
Residual Block #2	conv2_1	3 conv, 64	128 × 64	12,352	conv2_0
	leakyReLU2_1	-	128 × 64	0	conv2_1
	batchNorm2_1	-	128 × 64	128	leakyReLU2_1
	conv2_2	3 conv, 64	128 × 64	12,352	batchNorm2_1
	add_2	-	128 × 64	0	conv2_0, conv2_2
	leakyReLU2_2	-	128 × 64	0	add_2
	batchNorm2_2	-	128 × 64	128	leakyReLU2_2
Classification	globalAveragePool	/ 64	64	0	batchNorm2_2
	dense	-	118	7,670	globalAveragePool
			Total	34,098	

* Operation, Output Shape and No. of Params. are based on Dataset #1 and Dataset #3.

number of perceptrons with softmax activation in the dense layer is likely to change with respect to the number of subjects in the dataset.

Table 2 contains further architectural details of the proposed model. The operation, *x conv*, *y* refers that the residual field size is *x* for convolution operation and *y* number of filters and */z* refers to global average pooling across *z* channels. The operations, output shape and the number of parameters for all the layers are about to be changed with respect to datasets as the complexity of dataset and the number of subjects varies (see Table 1). In Table 2, the Operation, Output Shape and No. of Params. column has been populated depending on Dataset #1 and Dataset #3 for better understanding. Note that we have permuted the axis of the signals beforehand so that the data shape of 6×128 is converted into 128×6, which is suitable for our architecture.

2) PROPOSED AND OTHER VARIANTS OF RESIDUAL BLOCK

From the inception of the brilliant idea of residual learning [95], there have been numerous proposals for the residual block architecture. Along with the original residual block, He *et al.* proposed another four different configurations of residual block [96]. Moreover, with different settings of weighted layers, batch normalization, activation functions, and skip connections, there are some other residual blocks adopted to solve some other specific problems [101]–[104]. Fig. 2 contains a few of the depictions of the residual block architectures found in the literature. The sequence of the weighted and normalization layers and the skip connection varies among the blocks, whereas some have tried to replace the ReLU activation function with the other ones.

IV. EXPERIMENTAL ANALYSIS

A. EXPERIMENTAL SETTINGS AND EVALUATION METRICS

Zou *et al.* provided the benchmark datasets split into train and independent test set [6]. We have further split the train set into 90%–10% ramdomly, where 90% of the data was

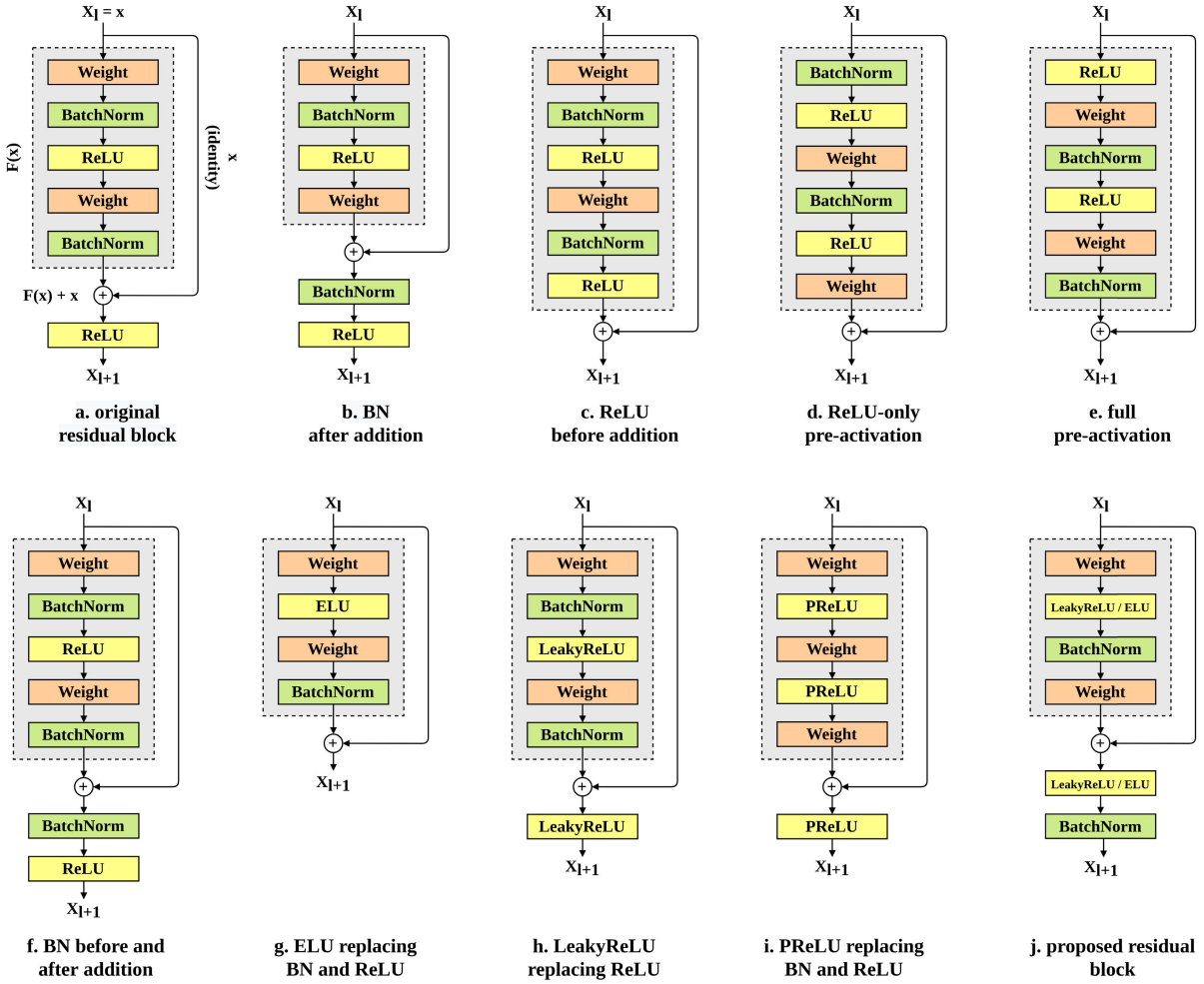


FIGURE 2. Different residual blocks used in the literature with the proposed one.

used for training and the other 10% as a validation set. In all experiments, during training, we trained the models for 1000 epochs with an early stopping mechanism to monitor validation loss using the Adam optimizer with a learning rate of 0.001. After training, we evaluated the performance of the models using the independent test set provided by Zou *et al.* [6].

We have reported accuracy (Acc) as the primary metric of measuring gait recognition performance and comparing the methods (see (1)). To measure the efficiency with respect to the previous studies in terms of parameter reduction (PR) and memory-usage reduction (MR), we have used (2) and (3), respectively. Moreover, the performance gain (PG) was computed as the difference between the performance with the existing methods and ours (see (4)). All of the experiments conducted in this study were backed by a tensor processing unit (TPU v2 with 8 cores) provided by Google Colab. The Keras API over TensorFlow backend was used to construct the models [105].

$$Acc\ (\%) = \frac{no.\ of\ correct\ predictions}{total\ no.\ of\ test\ cases} \times 100 \tag{1}$$

$$PR\ (\%) = \left[1 - \frac{no.\ of\ params.\ (our\ model)}{no.\ of\ params.\ (other\ model)}\right] \times 100 \tag{2}$$

$$MR\ (\%) = \left[1 - \frac{memory\ usage\ (our\ model)}{memory\ usage\ (other\ model)}\right] \times 100 \tag{3}$$

$$PG\ = Acc\ (our\ model) - Acc\ (other\ model) \tag{4}$$

Here, Acc, PR, MR and PG refers to accuracy, parameter reduction, memory-usage reduction and performance gain, respectively.

B. PERFORMANCE OF PROPOSED MODIFIED RESIDUAL BLOCK

We have evaluated the performance of the proposed residual architecture (see Fig. 1 (b) and Table 3) in different configurations. Varying the activation functions and the number of filters in the convolution layers in the second residual block, the performance of the model is shown in Table 3. We have fixed the number of feature maps in the first residual block of the model to 6 as the number of

TABLE 3. Performance of proposed residual block in different configurations.

# Filters	Activation	Accuracy (%)				
		Dataset #1	Dataset #2	Dataset #3	Dataset #4	OU-ISIR
16	ReLU	91.44	97.51	92.14	97.41	84.10
	ELU	91.84	97.81	92.21	97.82	85.73
	LeakyReLU	92.17	97.53	92.91	98.22	83.11
	PReLU	90.51	96.62	91.68	96.27	83.04
32	ReLU	93.74	97.65	93.75	98.27	93.12
	ELU	93.61	97.53	93.61	97.84	93.75
	LeakyReLU	93.96	98.10	94.02	98.86	92.05
	PReLU	92.59	97.37	93.05	98.55	92.97
64	ReLU	94.25	97.69	95.09	98.60	95.46
	ELU	94.39	97.69	94.58	98.45	96.81
	LeakyReLU	95.16	97.85	95.22	98.58	95.67
	PReLU	94.25	97.57	94.32	98.68	94.82
128	ReLU	94.65	97.75	95.22	98.50	97.16
	ELU	94.87	97.71	95.05	97.94	97.66
	LeakyReLU	95.04	97.77	95.39	98.55	97.52
	PReLU	94.65	97.71	94.92	98.60	96.88
256	ReLU	94.55	97.83	95.09	98.73	97.44
	ELU	95.03	97.81	95.35	98.43	98.30
	LeakyReLU	94.97	97.85	95.22	98.93	97.87
	PReLU	95.01	97.79	94.92	98.63	97.30

TABLE 4. Baseline configurations of the proposed architecture.

Criteria	Dataset #1	Dataset #2	Dataset #3	Dataset #4	OU-ISIR
# Filters	64	32	64	32	256
Activation	LeakyReLU	LeakyReLU	LeakyReLU	LeakyReLU	ELU

features in the input data. Convolving the output from the first residual block using 1D convolution operation, we have increased the number of feature maps into 16, 32, 64, 128, and 256, which is continued to the second residual block. Furthermore, the activation functions play an important role in signal propagation inside the model.

From Table 3, we can see that the LeakyReLU activation performs best with respect to the other activation functions in most of the cases. From the models that outperformed the state-of-the-art, we have identified the configurations based on the number of parameters and calculations overhead (marked in bold in Table 3). We call it the *baseline configuration* for our proposed model. The baseline configuration differs in terms of number of filters in the second residual block (# filters) and the non-linear activation function keeping rest of the architecture identical. For dataset #1 and dataset #3, we have selected the models with 64 filters in the second residual block using LeakyReLU as the activation function. We have selected the models with identical settings for dataset #2 and dataset #4 for 32 filters. Interestingly, the ELU activation works well for the OU-ISIR dataset. Therefore, we have selected the model with 256 feature maps and ELU activation while the LeakyReLU activation also outperformed all the state-of-the-art methods. Table 4 contains the summary of the baseline configurations of the proposed model. Although our proposed architecture is shallow, the number of parameters can increase if we use larger number of filters. Increasing the number of filters causes the better accuracy, but our goal is to make lightweight model with acceptable performance. Thus, we have selected lower number filters, when the difference of accuracy between the higher and lower # filters is negligible.

C. COMPARISON WITH OTHER RESIDUAL BLOCKS

We have incorporated all the prominent residual block architectures into our proposed architecture (see Section IV-B) and measured their performances. We have reported the results in Table 5. The table shows that our proposed residual block performs better than any other residual blocks for all datasets. The notable change in our residual block architecture is to perform non-linear activation before the batch normalization layer, which yields better performance.

D. ABLATION, INSERTION AND MODIFICATION STUDY ON PROPOSED ARCHITECTURE

To prove the efficacy and stability of our proposed architecture and evaluate partial importance of the components of the model, we have examined our model in three different ways. Along with the widely performed experiments called *Ablation* [106], [107], which was done by removing some module or portion of the proposed models (see Section IV-B) and measuring the performance to get the notion of partial importance of that module, we have also done some *Insertions* and *Modifications* in the model. *Insertions* are cases where we incorporate some other modules to the proposed model, and *modifications* are done by making some changes in some portions of the proposed model. In Table 6, we have listed all the ablation, insertion, and modification experiments and presented the performance of the models. To ease of understanding, we have depicted these models in Fig. 3.

We have *ablated* skip connections from the residual block to make them simple feed-forward CNN (see Fig. 3 (a)) and batch normalization layers to observe the effect of normalization (see Fig. 3 (g)). *Insertion* of the new modules have been done for most of the times, e.g., new skip connections from the input (head) (see Fig. 3 (b)), additional residual block (repetition of second residual block) with and without multi-headed skip connections (see Fig. 3 (e-f)), additional convolutional layer (with # filters = # filters with baseline configuration / 2) before the first residual block (see Fig. 3 (d)), introducing dropout in batch normalization layer [108] and tested with different dropout rates (see Fig. 3 (i)). In *modifications*, we have modified the first layer receptive field and tested with different kernel sizes (see Fig. 3 (c) i), used 1 convolution operation to increase feature maps in the convolution layer between the residual blocks (see Fig. 3 (c) II) and replaced the global average pooling with one and two layers of fully connected perceptron layers incorporated with batch normalization and dropout with $p = 0.3$ (see Fig. 3 (j)). Note that all the *ablations*, *insertions*, and *modifications* are independent of each other, i.e., we have performed all these experiments over the proposed model with baseline configurations defined in Section IV-B.

From Table 6, we can see that, the skip connections plays an important role as the performance decreased for no skip connection (Fig. 3 (a)) for all the datasets compared to the proposed lightweight model whereas, additional

TABLE 5. Performance comparison with other residual blocks incorporated with the proposed model.

Residual Blocks	Fig.	Accuracy (%)				
		Dataset #1	Dataset #2	Dataset #3	Dataset #4	OU-ISIR
Original Residual Block [95]	Fig. 2 (a)	93.98	97.39	93.58	97.72	94.46
BN after addition [96]	Fig. 2 (b)	94.52	97.51	94.75	98.25	95.81
ReLU before addition [96]	Fig. 2 (c)	94.28	97.45	94.48	97.67	96.10
ReLU-only pre-activation [96]	Fig. 2 (d)	94.28	97.45	94.02	98.43	95.32
Full pre-activation [96]	Fig. 2 (e)	94.06	97.37	94.08	98.73	92.41
BN before and after addition [101]	Fig. 2 (f)	94.36	97.59	94.85	97.36	95.46
ELU replacing BN and ReLU [102]	Fig. 2 (g)	94.17	97.55	94.62	97.89	96.74
LeakyReLU replacing ReLU [103]	Fig. 2 (h)	93.72	97.59	93.98	98.81	95.67
PReLU replacing BN and ReLU [104]	Fig. 2 (i)	91.02	95.48	92.44	95.28	70.69
Proposed Residual Block	Fig. 2 (j)	**95.16**	**98.10**	**95.22**	**98.86**	**98.30**

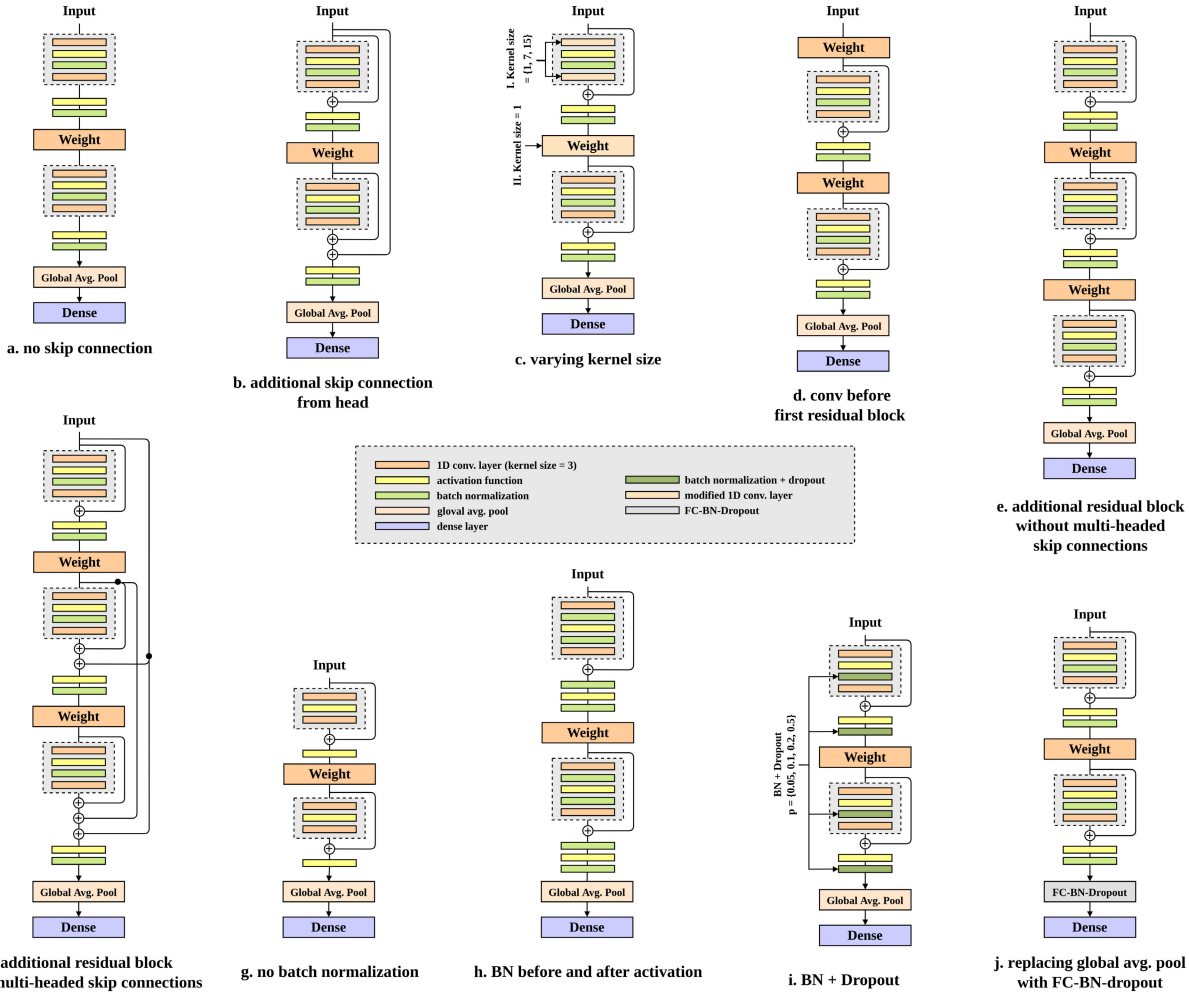


FIGURE 3. Experimental architectures for ablation, insertion and modification studies.

skip connection from head (Fig. 3 (b)) shows similar kind of results except for OU-ISIR dataset. Wei *et al.* [101] have chosen their first layer receptive field to cover the 10-millisecond duration of the signal, similar to the window size for many MFCC computations. The sampling rate of the signals they used was 8000 Hz; therefore, they have used 80 as their receptive field size. The sampling rate of our datasets are 50Hz and 100Hz. So, the optimal receptive field according to [101] will be $50/1000 * 10 = 0.5$ and $100/1000 * 10 = 1$. The receptive field size of 0.5 is invalid; therefore, we have experimented with the receptive field size of 1 along with 7 and 15 to observe the impact of increasing receptive field size. Nevertheless, for most of the cases, the proposed model (with receptive field size $= 3$) performs

TABLE 6. Performance comparison of the ablation, insertion and modification on proposed architecture.

Type	Description	Params.	Fig.	Accuracy (%)				
				Dataset #1	Dataset #2	Dataset #3	Dataset #4	OU-ISIR
Ablation	No skip connection		Fig. 3 (a)	94.25	97.69	94.42	98.63	98.15
Insertion	Additional skip connection from head		Fig. 3 (b)	94.52	97.73	94.75	98.55	**98.37**
Modification	First layer receptive field (kernel size)	1	Fig. 3 (c) I	94.76	97.81	**95.22**	98.76	**98.58**
		7	Fig. 3 (c) I	94.41	97.59	94.72	98.50	97.80
		15	Fig. 3 (c) I	94.17	97.61	94.65	98.45	96.59
Modification	Increase # filters using kernel size = 1		Fig. 3 (c) II	94.76	97.69	94.82	98.50	97.59
Insertion	Conv before first residual block		Fig. 3 (d)	**95.19**	97.91	94.92	98.73	**99.08**
Insertion	Additional residual block (with/without) multi-headed skip connections	without	Fig. 3 (e)	**95.35**	**98.34**	**95.29**	**98.91**	**98.86**
		with	Fig. 3 (f)	**95.45**	97.79	**95.22**	98.63	**98.86**
Ablation	No batch normalization (BN)		Fig. 3 (g)	94.79	97.69	95.15	98.76	97.94
Insertion	BN before and after activation		Fig. 3 (h)	93.69	97.63	93.71	98.71	88.79
Insertion	BN + dropout (p)	0.05	Fig. 3 (i)	94.60	97.87	**95.29**	**98.91**	**98.30**
		0.1	Fig. 3 (i)	94.60	97.79	94.92	**98.88**	97.59
		0.2	Fig. 3 (i)	94.04	97.57	94.65	98.50	95.53
		0.5	Fig. 3 (i)	87.89	93.70	87.30	94.44	34.63
Modification	Replacing global avg. pool with FC-BN-dropout (p=0.3)	1x	Fig. 3 (j)	92.11	97.02	91.74	97.56	44.02
		2x	Fig. 3 (j)	92.33	97.27	92.08	97.21	49.40
—	Proposed Lightweight Model		Fig. 1 (b)	**95.16**	**98.10**	**95.22**	**98.86**	**98.30**

Performances greater or equal to the proposed lightweight model are marked in bold.

better, and increasing or decreasing the field size decreases the performance.

Similarly, using 1 convolution operation, i.e., the receptive field size of 1 at the middle convolution layer (Fig. 3(c) II) performs worse than the proposed model. Additional convolutional layer before the first residual block (Fig. 3(d)) slightly improved the performance for dataset #1 whereas this model acquired the best performance of (99.08%) for OU-ISIR dataset. The additional residual block without multi-headed skip connection (Fig. 3(e)) performed better for all the datasets. Without the batch normalization operation (Fig. 3(g)), the performance decreased drastically; therefore, the batch normalization operation should be an integral part of the residual models. As it is an evergreen debate: where to put the batch normalization operation - before or after the non-linear activation, contrasting with the proposed models with batch normalization after the activation function and the original residual block with batch normalization before the activation function, we have experimented with batch normalization - both before and after activation function (Fig. 3 (h)); although it does not improve the performances. Dropout in batch normalization layer was proposed in [108]. A very minimal dropout percentage $p = 0.05$ can improve performance slightly, whereas increasing dropout percentage reduces accuracy due to too much regularization. Replacing global average pooling by fully connected layers with batch normalization and dropout (Fig. 3 (h)) results in drastic overfit. Increasing the number of fully connected layers increases the performance a bit, but none of them are up to the mark while increasing the number of parameters a lot. Our proposed model maximizes the representation learning in the convolutional layers without the use of fully connected layers.

TABLE 7. Performance comparison with the existing methods.

Method	Accuracy (%)				
	Dataset #1	Dataset #2	Dataset #3	Dataset #4	OU-ISIR
DeepConvLSTM [109]	92.25	96.80	-	-	-
IDNet [26]	92.91	96.78	-	-	-
CNN+LSTM $_{fix}$ [6]	93.52	97.33	-	-	72.32
LSTM & CNN [61]	94.15	-	-	-	89.79
CNN+CEDS [39]	94.71	97.67	95.09	98.58	97.16
Proposed Lightweight Model	**95.16**	**98.10**	**95.22**	**98.86**	**98.30**
Proposed Lightweight Model + ARB*	**95.35**	**98.34**	**95.29**	**98.91**	**98.86**

* ARB stands for Additional Residual Block.

From the above experimental result analysis, the performance of our proposed model with baseline configurations can be improved by additional weighted layers and residual blocks and proper choice of dropout in the batch normalization layers. Specially, insertion of an additional residual block shows marginal better performance than the lightweight model in all datasets. Since this insertion costs more than the proposed lightweight model, we considered to stick to the proposed model in further efficiency analysis in terms of number of parameters and memory-usage as our study aims to propose an efficient state-of-the-art model.

E. COMPARISON WITH THE STATE-OF-THE-ART METHODS

We have listed the performances of all the recent studies on gait recognition utilized the whuGait datasets and OU-ISIR dataset in Table 7 along with that of our proposed method.

We have reported the performances of the lightweight models described in Section IV-B and proposed model with an additional residual block from Section IV-D (see Figure 3(e)). Although the proposed model with additional residual block produces a little bit better accuracy than the proposed model, it increases the number of parameters. Since both of our models perform comparatively better than all

TABLE 8. Number of trainable parameters and memory-usage (in MB) comparison with the existing methods.

Method	Dataset #1		Dataset #2		Dataset #3		Dataset #4		OU-ISIR	
	# Params.	Mem.	# Params.	Mem.	# Params.	Mem.	# Params.	Mem.	# Params.	Mem.
Proposed Lightweight Model	**34,098**	**0.49**	**7,856**	**0.19**	**34,098**	**0.49**	**7,856**	**0.19**	**591,333**	**6.87**
CNN+CEDS [39]	344,055	4.24	168,341	2.13	343,543	~4.24*	168,341	~2.13*	1,468,266	17.70
CNN+LSTM$_{fix}$ [6]	4,716,406	56.70	4,415,252	53.10	-	-	-	-	4,986,601	59.90

* Memory-usage was not reported by the authors [39]; approximated from the number of parameters.

TABLE 9. Relative efficiency achieved by our proposed model.

Dataset	Compared Methods with Proposed Model	Achieved by Proposed Model		
		PR (%)	MR (%)	PG (%)
Dataset #1	CNN+CEDS [39]	90.09	88.44	+0.45
	CNN+LSTM$_{fix}$ [6]	99.28	99.14	+1.64
Dataset #2	CNN+CEDS [39]	95.33	91.08	+0.43
	CNN+LSTM$_{fix}$ [6]	99.82	99.64	+0.77
Dataset #3	CNN+CEDS [39]	90.07	88.44	+0.13
Dataset #4	CNN+CEDS [39]	95.33	91.08	+0.28
OU-ISIR	CNN+CEDS [39]	59.73	61.19	+1.14
	CNN+LSTM$_{fix}$ [6]	88.14	88.53	+25.98

PR, MR and PG refers to parameter reduction, memory-usage reduction and performance gain, respectively.

the previous methods and our goal is to produce lightweight model, the rest of the analysis will be performed on only the proposed lightweight model.

Simplifying the deep learning model is essential to propose a real-time gait recognition system and reduce the computation overhead. We have reduced the number of trainable parameters by a considerable margin. Our proposed model has a significantly lower number of parameters, and therefore, the inference time is decreased. The average inference time for Dataset#1 and Dataset#3 is 0.32 ms and 0.56 ms for GPU and CPU, respectively, whereas the time for Dataset#2 and Dataset#4 is 0.29 ms (GPU) and 0.43 ms (CPU). For the OU-ISIR dataset, the average inference time is 0.57 ms for GPU and 3.71 ms for CPU. We have calculated all these inference times in milliseconds (ms) on the Intel (R) Xeon(R) 2.20GHz CPU and the Tesla K80 GPU. Consumption of memory for inference has also dropped drastically for our model compared to the previous ones. Comparison with the existing methods in terms of the number of trainable parameters and memory-usage is listed in Table 8.

Using more than 99% fewer parameters than Zou *et al.* [6] we have achieved better performance in dataset #1 and dataset #2 whereas we have passed their performance by nearly 26% for the OU-ISIR dataset with 88% fewer parameters. On the other hand, using 92% fewer parameters on average than Huang *et al.*, we have achieved better performance in all four whuGait datasets. Our model had to use a greater number of feature maps in the second residual block, as discussed earlier, to surpass Huang *et al.* [39] in the OU-ISIR dataset. Still, the number of parameters is nearly 60% less than the mentioned study. Reduction in parameters and memory-usage (see (2) and (3)) is presented in Table 9 along with the performance gain (see (4)). Though the performance gain with respect to Huang *et al.* is marginal, our model costs much lesser computation overhead. Moreover, as we discussed in the ablation, insertion, and modification study section, the performance of our model can further be improved considering some tread-offs.

V. CONCLUSION

The primary intention of residual learning was to train very deep architectures. Nevertheless, we have successfully adopted the residual block with some modifications and efficiently created shallow convolutional neural networks for gait recognition. We have evaluated the performance of our methodology with two publicly available datasets collected in the wild and with the largest population and acquired state-of-the-art accuracy while reducing more than 85% of the parameters on average compared to the recent works. Our model can predict better than any other methods to date with minimum latency as the computation overhead reduces with the number of parameters that are suitable in practical applications using wearable devices. In future studies, we will explore and evaluate our methodology in other domains using IMU sensor datasets. Furthermore, usage of mobile computing power in smartphone-based recognition systems, response time, storage usage, energy consumption, etc. can also be evaluated in real-life scenarios. There are endless applications yet to be studied using gait patterns in medicine—classification of gait abnormalities that can utilize a similar setting. Abnormal gait patterns such as spastic, scissors, propulsive, steppage, etc., can be classified using wearable sensors. Moreover, abnormal gait patterns that develop over time due to some musculoskeletal or neurologic diseases can be predicted before they become life-threatening, potentially saving lives.

ACKNOWLEDGMENT

(Md. Al Mehedi Hasan and Fuad Al Abir contributed equally to this work.)

REFERENCES

[1] W. Shen and T. Tan, "Automated biometrics-based personal identification," *Proc. Nat. Acad. Sci. USA*, vol. 96, no. 20, pp. 11065–11066, Sep. 1999.

[2] R. He, X. Wu, Z. Sun, and T. Tan, "Wasserstein CNN: Learning invariant features for NIR-VIS face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1761–1773, Jul. 2019.

[3] G. Zheng, W. Yang, C. Valli, L. Qiao, R. Shankaran, M. A. Orgun, and S. C. Mukhopadhyay, "Finger-to-heart (F2H): Authentication for wireless implantable medical devices," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1546–1557, Jul. 2019.

[4] K. Wang and A. Kumar, "Toward more accurate iris recognition using dilated residual features," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3233–3245, Dec. 2019.
[5] S. Zhang and J. Liu, "Analysis and optimization of multiple unmanned aerial vehicle-assisted communications in post-disaster areas," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12049–12060, Dec. 2018.
[6] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li, "Deep learning-based gait recognition using smartphones in the wild," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3197–3212, 2020.
[7] R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, S. Ricerche, and F. Roli, "Fusion of multiple clues for photo-attack detection in face recognition systems," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–6.
[8] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, no. 12, pp. 22471–22499, Nov. 2014.
[9] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Proc. 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Oct. 2010, pp. 306–311.
[10] S. V. Stevenage, M. S. Nixon, and K. Vince, "Visual analysis of gait as a cue to identity," *Appl. Cogn. Psychol.*, vol. 13, pp. 513–526, Dec. 1999.
[11] A. Peinado-Contreras and M. Munoz-Organero, "Gait-based identification using deep recurrent neural networks and acceleration patterns," *Sensors*, vol. 20, no. 23, p. 6900, Dec. 2020.
[12] W. Zang, S. Zhang, and Y. Li, "An accelerometer-assisted transmission power control solution for energy-efficient communications in WBAN," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3427–3437, Dec. 2016.
[13] Q. Zou, L. Ni, Q. Wang, Q. Li, and S. Wang, "Robust gait recognition by integrating inertial and RGBD sensors," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1136–1150, Apr. 2018.
[14] M. D. Addlesee, A. Jones, F. Livesey, and F. Samaria, "The ORL active floor [sensor system]," *IEEE Pers. Commun.*, vol. 4, no. 5, pp. 35–41, Oct. 1997.
[15] B. Huang, M. Chen, P. Huang, and Y. Xu, "Gait modeling for human identification," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 4833–4838.
[16] D. Gafurov and E. Snekkenes, "Gait recognition using wearable motion recording sensors," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–16, Dec. 2009.
[17] T. Plötz, N. Y. Hammerla, and P. L. Olivier, "Feature learning for activity recognition in ubiquitous computing," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1729–1734.
[18] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognit. Lett.*, vol. 29, no. 16, pp. 2213–2220, Dec. 2008.
[19] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep activity recognition models with triaxial accelerometers," in *Proc. Workshops 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.
[20] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1553–1567, Oct. 2006.
[21] B. Shrestha, D. Ma, Y. Zhu, H. Li, and N. Saxena, "Tap-wave-rub: Lightweight human interaction approach to curb emerging smartphone malware," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2270–2283, Nov. 2015.
[22] R. E. Mayagoitia, A. V. Nene, and P. H. Veltink, "Accelerometer and rate gyroscope measurement of kinematics: An inexpensive alternative to optical motion analysis systems," *J. Biomech.*, vol. 35, no. 4, pp. 537–542, 2002.
[23] H. Zhao, Z. Wang, S. Qiu, Y. Shen, L. Zhang, K. Tang, and G. Fortino, "Heading drift reduction for foot-mounted inertial navigation system via multi-sensor fusion and dual-gait analysis," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8514–8521, Oct. 2019.
[24] S. Sprager and M. B. Juric, "An efficient HOS-based gait authentication of accelerometer data," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1486–1498, Jul. 2015.
[25] J. Frank, S. Mannor, and D. Precup, "Activity and gait recognition with time-delay embeddings," in *Proc. 24th AAAI Conf. Artif. Intell.*, 2010, pp. 1–6.
[26] M. Gadaleta and M. Rossi, "IDNet: Smartphone-based gait recognition with convolutional neural networks," *Pattern Recognit.*, vol. 74, pp. 25–37, Feb. 2018.
[27] Z. Qin, G. Huang, H. Xiong, Z. Qin, and K.-K.-R. Choo, "A fuzzy authentication system based on neural network learning and extreme value statistics," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 3, pp. 549–559, Mar. 2021.
[28] S. Deb, Y. O. Yang, M. C. H. Chua, and J. Tian, "Gait identification using a new time-warped similarity metric based on smartphone inertial signals," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 10, pp. 4041–4053, Oct. 2020.
[29] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007, pp. 69–84, 10.1007/978-3-540-74048-3_4.
[30] J. Adler and I. Parmryd, "Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient," *Cytometry A*, vol. 77A, pp. 733–742, Aug. 2010.
[31] T. A. Wren, K. P. Do, S. A. Rethlefsen, and B. Healy, "Cross-correlation as a method for comparing dynamic electromyography signals during gait," *J. Biomech.*, vol. 39, no. 14, pp. 2714–2718, Dec. 2005.
[32] H. Sun and T. Yuao, "Curve aligning approach for gait authentication based on a wearable accelerometer," *Physiol. Meas.*, vol. 33, no. 6, p. 1111, 2012.
[33] F. Sun, C. Mao, X. Fan, and Y. Li, "Accelerometer-based speed-adaptive gait authentication method for wearable IoT devices," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 820–830, Feb. 2019.
[34] F. Sun, W. Zang, R. Gravina, G. Fortino, and Y. Li, "Gait-based identification for elderly users in wearable healthcare systems," *Inf. Fusion*, vol. 53, pp. 134–144, Jan. 2020.
[35] C. Nickel, C. Busch, S. Rangarajan, and M. Mobius, "Using hidden Markov models for accelerometer-based biometric gait recognition," in *Proc. IEEE 7th Int. Colloq. Signal Process. Appl.*, Mar. 2011, pp. 58–63.
[36] Z. Yuting, P. Gang, K. Jia, M. Lu, Y. Wang, and Z. Wu, "Accelerometer-based gait recognition by sparse representation of signature points with clusters," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1864–1875, Sep. 2015.
[37] D. Gafurov, E. Snekkenes, and P. Bours, "Gait authentication and identification using wearable accelerometer sensor," in *Proc. IEEE Workshop Autom. Identificat. Adv. Technol.*, Jun. 2007, pp. 220–225.
[38] C. Xu, J. He, X. Zhang, C. Wang, and S. Duan, "Template-matching-based detection of freezing of gait using wearable sensors," *Proc. Comput. Sci.*, vol. 129, pp. 21–27, Jan. 2018.
[39] H. Huang, P. Zhou, Y. Li, and F. Sun, "A lightweight attention-based CNN model for efficient gait recognition with wearable IMU sensors," *Sensors*, vol. 21, no. 8, p. 2866, Apr. 2021.
[40] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, p. 973.
[41] H. J. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S.-M. Makela, "Identifying people from gait pattern with accelerometers," *Proc. SPIE*, vol. 5779, pp. 7–14, Apr. 2005.
[42] D. Gafurov, K. Helkala, and T. Sondrol, "Biometric gait authentication using accelerometer sensor," *J. Comput.*, vol. 1, no. 7, pp. 51–59, Nov. 2006.
[43] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 491–502, Sep. 2007.
[44] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, "Identification of individual walking patterns using gait acceleration," in *Proc. 1st Int. Conf. Bioinf. Biomed. Eng.*, Jul. 2007, pp. 543–546.
[45] L. Rong, Z. Jianzhong, L. Ming, and H. Xiangfeng, "A wearable acceleration sensor system for gait recognition," in *Proc. 2nd IEEE Conf. Ind. Electron. Appl.*, May 2007, pp. 2654–2659.
[46] G. Trivino, A. Alvarez-Alvarez, and G. Bailador, "Application of the computational theory of perceptions to human gait pattern recognition," *Pattern Recognit.*, vol. 43, no. 7, pp. 2572–2581, Jul. 2010.
[47] S. Sprager and M. B. Juric, "Inertial sensor-based gait recognition: A review," *IEEE Sensors J.*, vol. 15, no. 9, pp. 22089–22127, Sep. 2015.
[48] L. Wu, J. Yang, M. Zhou, Y. Chen, and Q. Wang, "LVID: A multimodal biometrics authentication system on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1572–1585, 2020.
[49] A. H. Johnston and G. M. Weiss, "Smartwatch-based biometric gait recognition," in *Proc. IEEE 7th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2015, pp. 1–6.
[50] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–7.

[51] B. Sun, Y. Wang, and J. Banda, "Gait characteristic analysis and identification based on the iPhone's accelerometer and gyrometer," *Sensors*, vol. 14, no. 9, pp. 17037–17054, Sep. 2014.

[52] J. Le Moing and I. Stengel, "The smartphone as a gait recognition device impact of selected parameters on gait recognition," in *Proc. Int. Conf. Inf. Syst. Secur. Privacy (ICISSP)*, Feb. 2015, pp. 322–328.

[53] H. Abujrida, E. Agu, and K. Pahlavan, "Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data," in *Proc. IEEE Healthcare Innov. Point Care Technol. (HI-POCT)*, Nov. 2017, pp. 208–211.

[54] J. Juen, Q. Cheng, V. Prieto-Centurion, J. A. Krishnan, and B. Schatz, "Health monitors for chronic disease by gait analysis with mobile phones," *Telemed. e-Health*, vol. 20, no. 11, pp. 1035–1041, Nov. 2014.

[55] Y. Ren, Y. Chen, M. C. Chuah, and J. Yang, "User verification leveraging gait recognition for smartphone enabled mobile healthcare systems," *IEEE Trans. Mobile Comput.*, vol. 14, no. 9, pp. 1961–1974, Sep. 2015.

[56] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: From authentication to imitation," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3209–3221, Nov. 2017.

[57] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides, "Gait-ID on the move: Pace independent human identification using cell phone accelerometer dynamics," in *Proc. IEEE 5th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2012, pp. 8–15.

[58] P. Fernandez-Lopez, J. Sanchez-Casanova, P. Tirado-Martin, and J. Liu-Jimenez, "Optimizing resources on smartphone gait recognition," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 31–36.

[59] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "DeepSense: A unified deep learning framework for time-series mobile sensing data processing," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 351–360.

[60] V. N. Bobic, M. D. Djuric-Jovieic, S. M. Radovanovic, N. T. Dragaevic, V. S. Kostic, and M. B. Popovic, "Challenges of stride segmentation and their implementation for impaired gait," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2284–2287.

[61] L. Tran, T. Hoang, T. Nguyen, H. Kim, and D. Choi, "Multi-model long short-term memory network for gait recognition using window-based data segment," *IEEE Access*, vol. 9, pp. 23826–23839, 2021.

[62] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *Proc. 6th Int. Conf. Mobile Comput., Appl. Services*, 2014, pp. 197–205.

[63] L. Xu, Y. He, and J. He, "Valid inertial gait data recovery for gait recognition: A multi-mode adaptive orthogonal matching pursuit," *J. Phys., Conf. Ser.*, vol. 2026, no. 1, Sep. 2021, Art. no. 012038.

[64] L. Zhang, X. Wu, and D. Luo, "Recognizing human activities from raw accelerometer data using deep neural networks," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2015, pp. 865–870.

[65] E. Rastegari, S. Azizian, and H. Ali, "Machine learning and similarity network approaches to support automatic classification of Parkinson's diseases using accelerometer-based gait analysis," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 1–12.

[66] N. Ortiz, R. D. H. Beleño, R. J. Moreno, M. Mauledeoux, and O. F. A. Sānchez, "Survey of biometric pattern recognition via machine learning techniques," *Contemp. Eng. Sci.*, vol. 11, no. 34, pp. 1677–1694, Jan. 2018.

[67] R. Damaševičius, M. Vasiljevas, J. Šalkevičius, and M. Woźniak, "Human activity recognition in AAL environments using random projections," *Comput. Math. Methods Med.*, vol. 2016, pp. 1–17, May 2016.

[68] Y. Watanabe, "Influence of holding smart phone for acceleration-based gait authentication," in *Proc. 5th Int. Conf. Emerg. Secur. Technol.*, Sep. 2014, pp. 30–33.

[69] H. Chan, H. Zheng, H. Wang, R. Sterritt, and D. Newell, "Smart mobile phone based gait assessment of patients with low back pain," in *Proc. 9th Int. Conf. Natural Comput. (ICNC)*, Jul. 2013, pp. 1062–1066.

[70] G. Li, L. Huang, and H. Xu, "IWalk: Let your smartphone remember you," in *Proc. 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2017, pp. 414–418.

[71] C. Nickel, T. Wirtl, and C. Busch, "Authentication of smartphone users based on the way they walk using k-NN algorithm," in *Proc. 8th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Jul. 2012, pp. 16–20.

[72] S. Choi, I.-H. Youn, R. LeMay, S. Burns, and J.-H. Youn, "Biometric gait recognition based on wireless acceleration sensor using k-nearest neighbor classification," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2014, pp. 1091–1095.

[73] F. I. Pratama and A. Budianita, "Optimization of K-Nn classification in human gait recognition," in *Proc. 5th Int. Conf. Informat. Comput. (ICIC)*, Nov. 2020, pp. 1–5.

[74] K. Ren, Q. Wang, C. Wang, Z. Qin, and X. Lin, "The security of autonomous driving: Threats, defenses, and future directions," *Proc. IEEE*, vol. 108, no. 2, pp. 357–372, Feb. 2020.

[75] L. Zhao, Q. Wang, Q. Zou, Y. Zhang, and Y. Chen, "Privacy-preserving collaborative deep learning with unreliable participants," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1486–1500, Sep. 2019.

[76] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[77] J. Man and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[78] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio, "Full-body person recognition system," *Pattern Recognit.*, vol. 36, no. 9, pp. 1997–2006, 2003.

[79] K.-T. Nguyen, T.-L. Vo-Tran, D.-T. Dinh, and M.-T. Tran, "Gait recognition with multi-region size convolutional neural network for authentication with wearable sensors," in *Proc. Int. Conf. Future Data Secur. Eng.*, Cham, Switzerland: Springer, 2017, pp. 197–212.

[80] J. Hannink, T. Kautz, C. F. Pasluosta, K.-G. Gaßmann, J. Klucken, and B. M. Eskofier, "Sensor-based gait parameter extraction with deep convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 85–93, Jan. 2017.

[81] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *J. Mach. Learn. Res.*, vol. 7, pp. 1531–1565, Jul. 2006.

[82] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.

[83] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, "A comprehensive study on cross-view gait based human identification with deep CNNs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 209–226, Feb. 2016.

[84] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "On input/output architectures for convolutional neural network-based cross-view gait recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 9, pp. 2708–2719, Sep. 2017.

[85] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and A. Bouridane, "Gait recognition for person re-identification," *J. Supercomput.*, vol. 77, no. 4, pp. 3653–3672, Apr. 2021.

[86] S. Gul, M. I. Malik, G. M. Khan, and F. Shafait, "Multi-view gait recognition system using spatio-temporal features and deep learning," *Expert Syst. Appl.*, vol. 179, Oct. 2021, Art. no. 115057.

[87] X. Liu, M. Chen, T. Liang, C. Lou, H. Wang, and X. Liu, "A lightweight double-channel depthwise separable convolutional neural network for multimodal fusion gait recognition," *Math. Biosci. Eng.*, vol. 19, no. 2, pp. 1195–1212, 2021.

[88] C. Li, X. Min, S. Sun, W. Lin, and Z. Tang, "DeepGait: A learning deep convolutional representation for view-invariant gait recognition using joint Bayesian," *Appl. Sci.*, vol. 7, no. 3, p. 210, Feb. 2017.

[89] F. M. Castro, M. J. Marín-Jiménez, N. Guil, and N. P. D. L. Blanca, "Automatic learning of gait signatures for people identification," in *Proc. Int. Work-Conf. Artif. Neural Netw.*, Cham, Switzerland: Springer, 2017, pp. 257–270.

[90] S. Tong, Y. Fu, X. Yue, and H. Ling, "Multi-view gait recognition based on a spatial-temporal deep neural network," *IEEE Access*, vol. 6, pp. 57583–57596, 2018.

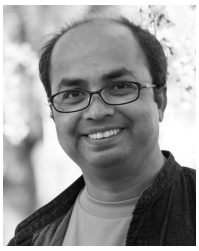
[91] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, "Long-term recurrent convolutional networks for visual recognition and description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2625–2634.

[92] S. Yu, H. Chen, E. B. G. Reyes, and N. Poh, "GaitGAN: Invariant gait feature extraction using generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 30–37.

[93] X. Chen, X. Luo, J. Weng, W. Luo, H. Li, and Q. Tian, "Multi-view gait image generation for cross-view gait recognition," *IEEE Trans. Image Process.*, vol. 30, pp. 3041–3055, 2021.

[94] 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 Recognit.*, vol. 47, no. 1, pp. 228–237, 2014.

[95] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[96] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 630–645.
[97] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, J. Fürnkranz and T. Joachims, Eds. Madison, WI, USA: Omnipress, Jun. 2010, pp. 807–814. [Online]. Available: https://icml.cc/Conferences/2010/papers/432.pdf
[98] D. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in *Proc. 4th Int. Conf. Learn. Represent., (ICLR)*, Y. Bengio and Y. LeCun, Eds. San Juan, Puerto Rico, May 2016, pp. 1–14.
[99] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. ICML*, vol. 30. Princeton, NJ, USA: Citeseer, 2013, p. 3.
[100] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[101] W. Dai, C. Dai, S. Qu, J. Li, and S. Das, "Very deep convolutional neural networks for raw waveforms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 421–425.
[102] A. Shah, E. Kadam, H. Shah, S. Shinde, and S. Shingade, "Deep residual networks with exponential linear unit," in *Proc. 3rd Int. Symp. Comput. Vis. Internet*, Sep. 2016, pp. 59–65.
[103] S. Allamy and A. L. Koerich, "1D CNN architectures for music genre classification," 2021, *arXiv:2105.07302*.
[104] W. Wang, C. Ren, X. He, H. Chen, and L. Qing, "Video super-resolution via residual learning," *IEEE Access*, vol. 6, pp. 23767–23777, 2018.
[105] F. Chollet. (2015). *Keras*. [Online]. Available: https://github.com/fchollet/keras
[106] S. Sheikholeslami, "Ablation programming for machine learning," M.Sc. thesis, School Elect. Eng. Comput. Sci., KTH Roy. Inst. Technol. Stockholm, Sweden, Jul. 2019, doi: 10.13140/RG.2.2.27959.06567.
[107] L. Du, "How much deep learning does neural style transfer really need? An ablation study," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 3150–3159.
[108] G. Chen, P. Chen, Y. Shi, C.-Y. Hsieh, B. Liao, and S. Zhang, "Rethinking the usage of batch normalization and dropout in the training of deep neural networks," 2019, *arXiv:1905.05928*.
[109] F. J. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, Jan. 2016.



MD. AL MEHEDI HASAN received the B.Sc., M.Sc., and Ph.D. degrees in computer science and engineering from the Department of Computer Science and Engineering, University of Rajshahi, Rajshahi, Bangladesh, in 2005, 2007, and 2017, respectively. He became a Lecturer, an Assistant Professor, an Associate Professor, and a Professor with the Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology (RUET), Rajshahi, in 2007, 2010, 2018, and 2019, respectively. He has (co)authored more than 100 publications published in widely cited journals and conferences. His research interests include bioinformatics, artificial intelligence, pattern recognition, medical image and signal processing, machine learning, computer vision, data mining, big data analysis, probabilistic and statistical inference, operating systems, computer networks, and security.

FUAD AL ABIR was born in Khalishpur, Khulna, Bangladesh. Currently, he is a final year student of computer science and engineering with the Rajshahi University of Engineering and Technology (RUET), Bangladesh. He joined the Machine Learning Research Group, RUET, in August 2019, under the supervision of Prof. Dr. Md. Al Mehedi Hasan and (co)authored a few articles published in reputed journals and conferences. He is very keen on various types of data (tabular, time series, sensors, and medical imaging) and their application in healthcare, generative art, and explainable machine learning.

MD. AL SIAM was born in Mymensingh, Bangladesh, in 1997. He is currently pursuing the B.Sc. degree in computer science and engineering with the Rajshahi University of Engineering and Technology (RUET), Rajshahi, Bangladesh. His research interests include computer vision, the Internet of Things, sensor data, pattern recognition, and deep learning. He is also interested in problem solving and competitive programming, and participated in various programming contests. He is currently working on human activity recognition using wearable sensors. He is an active member of the Machine Learning Research Group, RUET.

JUNGPIL SHIN (Senior Member, IEEE) received the B.Sc. degree in computer science and statistics and the M.Sc. degree in computer science from Pusan National University, South Korea, in 1990 and 1994, respectively, and the Ph.D. degree in computer science and communication engineering from Kyushu University, Japan, in 1999, under the scholarship from the Japanese Government (MEXT). He was an Associate Professor, a Senior Associate Professor, and a Professor with the School of Computer Science and Engineering, The University of Aizu, Japan, in 1999, 2004, and 2019, respectively. He has coauthored more than 250 publications published in widely cited journals and conferences. His research interests include pattern recognition; image processing; computer vision and machine learning; human–computer interaction; non-touch interface; human gesture recognition; automatic control; Parkinson's disease diagnosis; ADHD diagnosis; user authentication; machine intelligence; and handwriting analysis, recognition, and synthesis. He is a member of ACM, IEICE, IPSJ, KISS, and KIPS. He has served several conferences as the program chair and a program committee member for numerous international conferences. He serves as a reviewer for several IEEE and SCI major journals. He serves as an Editor for IEEE journals and *Sensors* (MDPI).