

Privacy-Preserving, Thermal Vision With Human in the Loop Fall Detection Alert System

Abdallah Naser ^{ORCID}, Ahmad Lotfi ^{ORCID}, *Senior Member, IEEE*, Maria Drolence Mwanje, and Junpei Zhong ^{ORCID}, *Senior Member, IEEE*

Abstract—To support the independent living of older adults in their own homes, it is essential to identify their abnormal behaviors before triggering an automated alert system. Existing normal vision sensing approaches to detect human falls in the activities of daily living (ADL) experienced acceptability issues due to outstanding privacy concerns when they are deployed in personal environments. Besides, false alerts (false-positive) fall detection has not been addressed thoroughly in systems that report abnormal human behaviors as emergency alerts to the information support. This article proposes a novel human-in-the-loop fall detection approach in the ADLs using a low-resolution thermal sensor array. The motivation for enabling a human interactive model, fall detection confirmation, is to influence resource efficiency by reducing false-positive alerts while keeping the false-negative fall predictions as low as possible. The proposed approach is based on the motion sequence classification of human movements using a recurrent neural network. The proposed approach is evaluated with comprehensive experiments using different learning techniques, users, and domestic environment conditions. This article shows a performance accuracy of 99.7% to detect human falls from various typical ADLs.

Index Terms—Activities of daily living (ADL) recognition, anomaly detection, fall detection, human-in-the-loop, machine learning, optical flow, privacy-preserving approach, thermal sensor array (TSA).

I. INTRODUCTION

THERE has been an increase in the ageing population over recent years. According to the World Health Organization, the older adults community aged 60+ years is expected to grow to 22% in 2050 from 12% in 2015 worldwide [1]. Due to the ageing effects, individuals experience a decline in mental and

physical ability. The decline in health is associated with several health conditions such as dementia, depression, diabetes, and increased risk of falls in older adults, leading to an increase in abnormal human behaviors that require regular medical attention. Globally, over 684 000 people suffer death due to falls each year [2]. Furthermore, 50% of falls occur among the older adults, who live alone or in nursing homes. Thus, there is a necessity to reduce the post effects of falls through an automatic fall detection system that can trigger a rapid and accurate response to the fall [3], [4], [5].

Fall detection approaches for older adults infuse methods based on wearable, vision, and ambient sensing technologies. The performance of wearable devices depends on the user's acceptability and ability to keep the device on and in the correct position. These limitations are critical for older adults with dementia, as the user may forget to carry these devices. On the other hand, vision-based approaches usually use a camera to capture images. Such methods solve the problem of fixed device placements in wearable-based strategies and can capture multiple events visibly and concurrently. The major drawback of vision-based approaches is an infringement on people's privacy and its dependence on the light intensity, direction etc. Besides, human falls can occur anytime, and therefore the ambient light dependency in cameras has a significant concern regarding the reliability of this approach in a potentially dark home environment. Finally, ambient-based devices utilize environmental data such as floor pressure, sound, and movement of objects. This approach provides better privacy. However, performance is not satisfied as surrounding objects may affect measurements, leading to high false positives [6]. Therefore, previous fall detection systems for older adults are a tradeoff between privacy and performance.

Recently, the thermal sensor array (TSA) has been utilized as an emerging sensing technology in human-centred applications to bridge the gap between privacy and performance in vision and wearable-based approaches [14], which is a key issue experienced in previous fall detection systems. However, most of the previous reported work that explored the use of TSA in human fall detection did not take into account the effect of false-positive fall detection scenarios on the consequent waste of emergency alerting and response resources, leading to real concerns about the deployability of such systems in a wide range of environments. This article proposes a human-in-the-loop fall detection system based on human motion analysis. In doing so, the false-positive human fall emergency alerts and reactions are

IEEE Transactions on
Human-Machine Systems

Print ISSN: 2168-2291

Online ISSN: 2168-2305

DOI: 10.1109/THMS.2022.3203021

kept to the minimum in a deployable TSA-based fall detection along with maintaining low false-negative predicted fall cases. In summary, the main contributions of this article include the following.

- 1) An accountable, privacy-preserving, contactless, and low-cost human-in-the-loop fall detection approach using a low-resolution thermal vision sensor.
- 2) The system maintains a high-performance fall detection and reduces human fall false-positive alerts reported to the information support by enabling a human-interaction interface.
- 3) This article explores the use and efficiency of optical flow features in activities of daily living (ADL) recognition and fall detection using TSA sensing technology.
- 4) Comprehensive experiments are conducted to validate the use of optical flow features with Long short-term memory (LSTM) and Bidirectional long-short term memory (Bi-LSTM) in the prediction stage.
- 5) A thorough evaluation of the conducted experiments versus the proposed pre-processing technique on the TSA's output.
- 6) The use of optical flow features with the TSA enables the proposed system to operate in a dark or thermal noisy environment. To the best of the authors' knowledge, the human-in-the-loop fall detection system with optical flow feature extraction on the TSA is proposed for the first time in this article.

The remaining parts of this article are organized as follows: in Section II, a summary of the related work is presented. Section III explains the proposed framework architecture. Comprehensive experiments and evaluations are presented and discussed in Sections IV, and Section V followed by pertinent conclusions drawn in Section VII.

II. RELATED WORK

The use of TSA sensors in human-centred applications is relatively low compared to wearable, vision-based, and ambient sensors [15]. Conventional wearable devices use an accelerator's sensor to measure the object acceleration [16], [17], [18]. The wearable approach is based on utilizing the change in motion, location, and posture of the monitored object. In general, approaches that employ wearable devices are cost-effective and easy to design and commission. On the other hand, these need to be worn by the user for accurate results, and they inconvenience users [6]. In addition, individuals are prone to forgetting to wear the device, or situations such as the need to take a shower force users to take off the devices, hence the inability to detect falls accurately, leading to low performance.

In contrast to wearable devices, vision-based methods include the use of ordinary cameras [19], [20], [21], which solves the fixed body device location in the signal acquisition stage. The vision-based approach can be subdivided based on operation, including change in shape, inactivity, head change analysis, posture, and Spatio-temporal [6]. For instance, in the inactivity technique, the user's period of inactivity on the floor contributes to detecting a fall. However, the main drawbacks are light

dependency and its violation of people's privacy. Finally, ambient sensing approach for fall detection, for example, pressure sensing [22], Passive-Infra-Red (PIR), and floor vibration [23]. Ambient sensing ensures user privacy, which is a critical issue in vision-based approaches and is more convenient than wearable-based approaches. However, its detection is affected by all variables within the environment, resulting in low performance. The use of TSA in fall detection has emerged in recent years to bridge the gap between performance and user privacy concerns [10], [24], [25], [26]. TSA has also been proposed in other applications including human distance estimation [27], physical distancing [28], [29], occupancy estimation [14], [30], [31], [32], and human activity recognition [33], [34]. A critical comparison of TSA settings, data-driven methods, and performance of the state-of-art TSA-based fall detection systems has been provided in Tables I and II.

Although TSA-based systems achieve excellent performance results, the issue of false-positive in human fall detection has not been addressed. This raises an important question about the consequences of deploying such systems on a larger scale, especially when dealing with many false-positive fall alerts. Hence, the need for the research is reported in this article.

III. HUMAN IN THE LOOP FALL DETECTION USING A LOW-RESOLUTION TSA

The proposed approach, depicted in Fig. 1, consists of two main stages: stage 1) human fall detection among ADL, and stage 2) human interaction interface. To detect human falls, the proposed system should be capable of distinguishing between normal and abnormal human activities. When a fall is detected, the proposed approach is accountable to the users to confirm the fall to the information support by means of an emergency alert, a notification to family members, etc. Human-based fall confirmation is provided by a mobile application interface model, which requires the user to confirm TSA-based fall detection to proceed to the alert phase. If the user does not confirm the fall, e.g., due to the consequences of the fall, the system will automatically consider this a human fall. Detailed descriptions of the functional phases of these two stages are provided below.

A. Sensor Placement and Acquisition

The signal acquisition in the proposed approach is based on a Far Infrared TSA sensor.¹ The used sensor attains object temperature readings as a 32×24 matrix. Considering the relatively low-resolution of the sensor, it is impossible to identify information regarding people in the sensor's field of view (FoV). Therefore, it has been claimed as a privacy-preserving sensing approach. The sensor can be accessed via the *I2C* interface, and its current consumption is less than 23 mA. This consumption makes it suitable for a battery-powered solution. Furthermore, TSA has a high-refresh-rate between 0.5 and 64 Hz, making it capable of detecting swift human movements. Besides, the sensor can be considered a user-friendly device, since it is not

¹The sensor details can be obtained from the Melexis website: <https://www.melexis.com/en/product/MLX90640/>

TABLE I
COMPARISON OF TSA TYPE, RESOLUTION, PLACEMENT, AND SENSOR PLACEMENT ADAPTABILITY PROPOSED FOR HUMAN FALL DETECTION SYSTEMS

Ref.	TSA Type and Resolution	TSA Placement	Adaptive Placement
[7]	Panasonic's Grid-EYE (8x8)	Ceiling	No
[8]	Melixis (16x4)	Wall	No
[9]	Panasonic's Grid-EYE (8x8)	Mini-robot	No
[10]	Panasonic's Grid-EYE (8x8)	Ceiling	No
[11]	Panasonic's Grid-EYE (8x8)	Ceiling	No
[12]	Melixis (16x4)	Ceiling, side by side	No
[13]	Melixis (16x4)	Ceiling, side by side	No
[11]	Panasonic's Grid-EYE (8x8)	Ceiling	No

TABLE II
COMPARISON OF THE PERFORMED ACTIVITIES, DATA-DRIVEN APPROACH, AND TSA-BASED FALL DETECTION STATE-OF-ART PERFORMANCE RESULTS

Ref.	Performed Activities	Learning Algorithm	Accuracy (%)	Accountability
[7]	Standing Sitting Lying	CNN	–	No
[8]	Sitting Bending Squatting Walking Standing	k-NN	93%	No
[9]	Standing Sitting Picking up	SVM	88,7% – 94,7%	No
[10]	Sitting Walking	k-NN	94.3% – 95.8%	No
[11]	Walking Jogging Squatting Lying down Staying still	Random Forest	–	No
[12]	Sitting Bending Squatting Walking Standing	Voting classifier	97.75%	No
[13]	Standing Sitting Lying	Logistic regression	99.94%	No
[11]	Staying seated Staying up walking standing up	3D-ConvNet	97.22%	No

worn by the user, thus offering users a high level of convenience. Also, TSA is light-independent, which means it can still operate successfully in dark environments. Although, it is sensitive to ambient temperature and thermal noises. Thus, it is crucial to ensure appropriate data processing techniques that are specifically suited to this sensing approach.

To determine the temperature of a specific region in the thermal image, this region has to illuminate at least one complete FIR (pixel). Otherwise, the pixel will represent a mixed temperature of the object and the adjacent background. For example, the human presence shown in Fig. 2 shows a variation in human acquired temperatures. One possible reason for this variance is dressing. However, the facial region of the human presence has also suffered from the temperature variance. This

is because some of the acquired face temperatures were not fully illuminated by a full pixel.

Another important TSA's consideration is the human presence in the sensor's FoV versus its placement. Hence, the FoV of the used sensor is $55^\circ \times 35^\circ$. Based on that, the size of the inspection area m at a distance d can be computed as follows:

$$m = 2 \times d \times \tan\left(\frac{\text{FoV}}{2}\right). \quad (1)$$

Placing the TSA in an elevated position on the wall would acquire the upper parts of the human presence in the sensor's FoV, but not the lower parts. Fig. 3(a) shows the effects of sensor position height on the acquired thermal human presence for a short participant, while Fig. 3(b) showing a taller participant. It

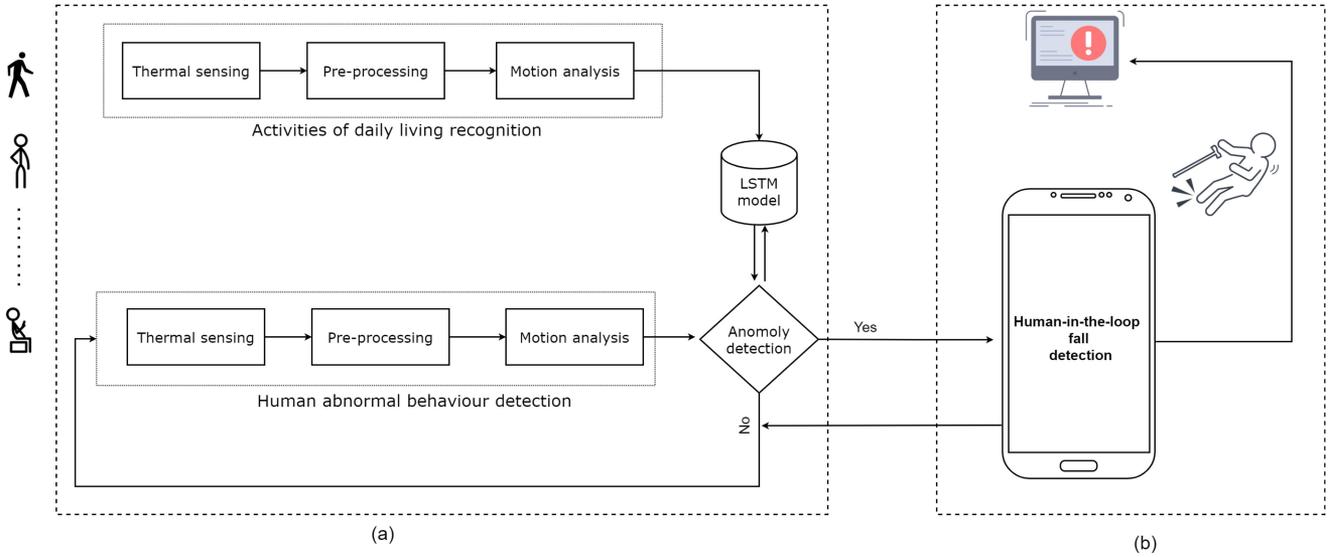


Fig. 1. Schematic diagram of the proposed accountable human-in-the-loop fall detection system based on optical flow feature extraction, where (a) the ADL recognition and human fall detection stage, (b) the human interactive interface to confirm the TSA-based human fall detection.

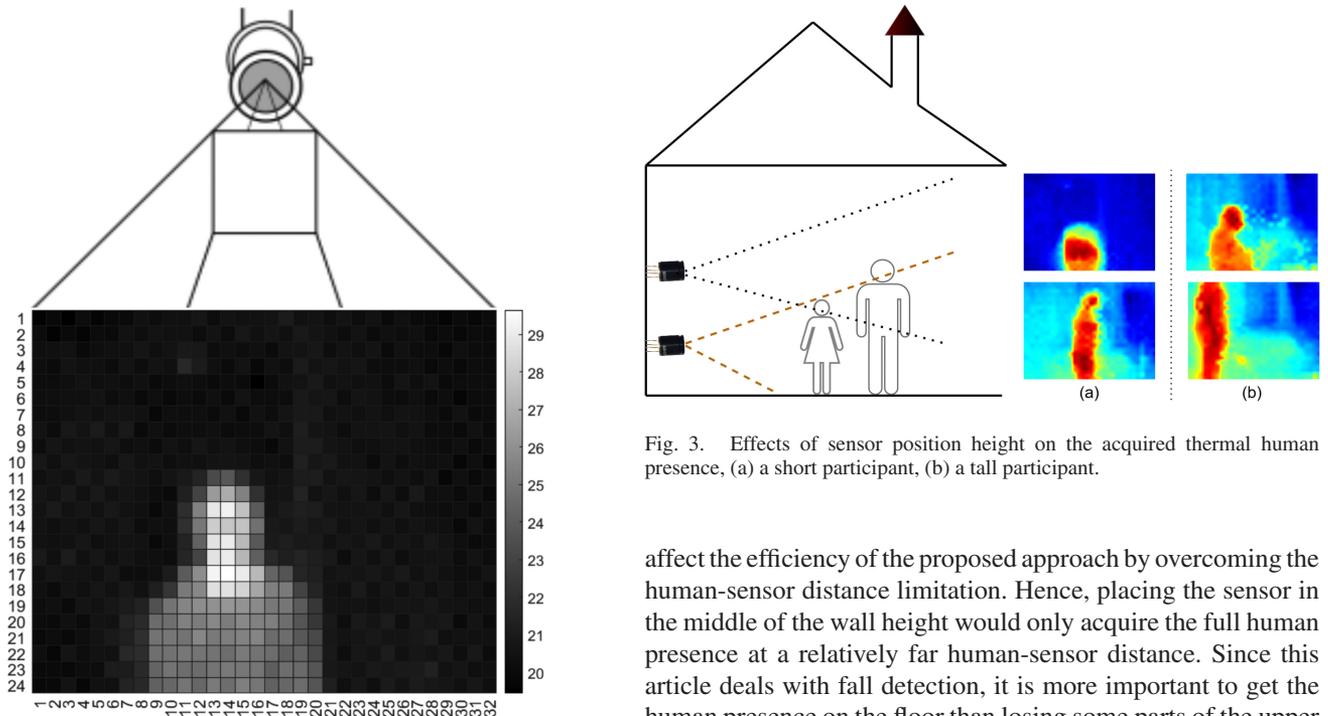


Fig. 2. Visualization of TSA spatial temperature matrix representing the presence of a human subject.

can be seen from these figures that significant human presence has not been acquired for the short participant and is fully visible from short height sensor placement. On the other hand, short height sensor placement has observed some missing upper parts for tall human participants. Since this article deals with the low-resolution thermal imaging for fall detection, we propose the sensor placement for human fall detection to be at a short height from the floor for a wider inspection area. This will positively

Fig. 3. Effects of sensor position height on the acquired thermal human presence, (a) a short participant, (b) a tall participant.

affect the efficiency of the proposed approach by overcoming the human-sensor distance limitation. Hence, placing the sensor in the middle of the wall height would only acquire the full human presence at a relatively far human-sensor distance. Since this article deals with fall detection, it is more important to get the human presence on the floor than losing some parts of the upper human presence for relatively tall users.

B. Preprocessing

When a person moves in an environment, it will generate background thermal noise from previous human motion positions. The generated noise creates biased motion estimation. Therefore, the acquired heat-map image is split into background and foreground classes to eliminate the background pixels using the Otsu threshold method [35]. Technically, the threshold k of separating each $2D$ temperature matrix x_t into foreground t_f and background t_b classes should minimise the intra-class

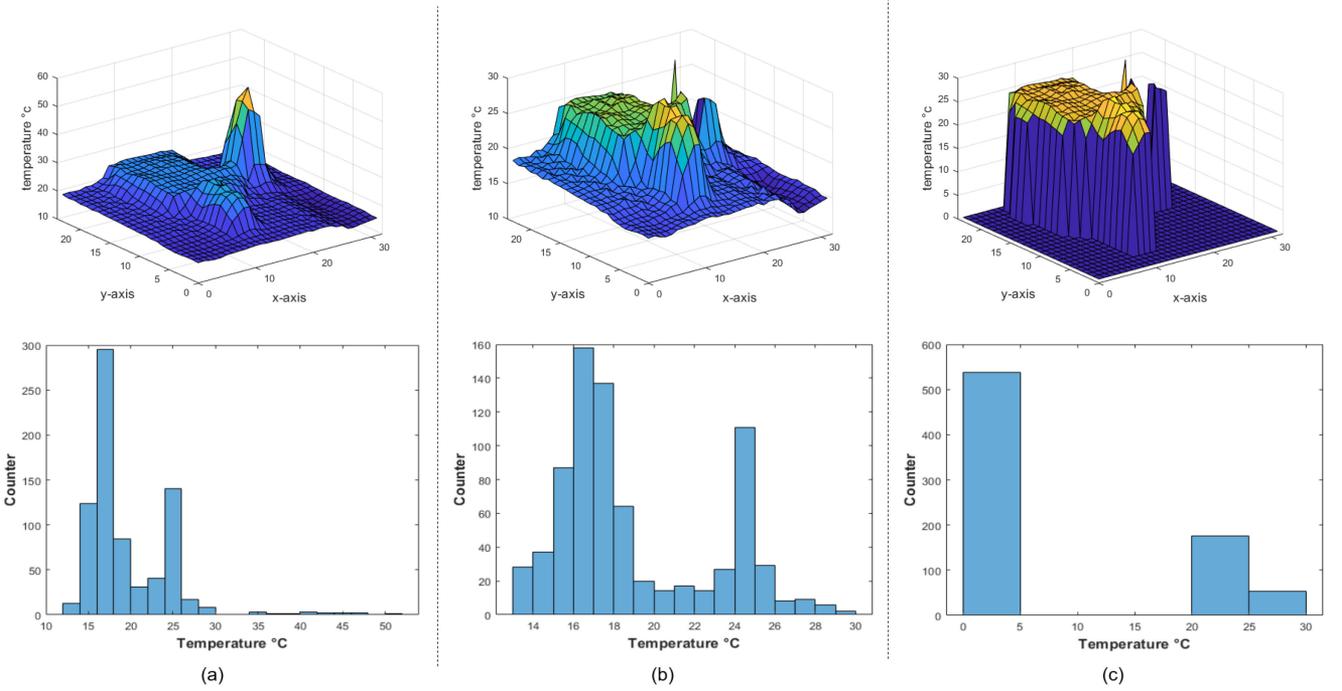


Fig. 4. Illustration of the spread of the temperature values throughout the pre-processing phase using (a) the original temperature surface plot and its corresponding histogram, (b) after application of a temperature filter, (c) the result of separating the human presence (foreground) and the ambient environment (background) into two classes.

variance $\sigma_w^2(k)$, defined as a weighted sum of variances of the two classes

$$\sigma_w^2(k) = \omega_{t_f}(k)\sigma_{t_f}^2(k) + \omega_{t_b}(k)\sigma_{t_b}^2(k) \quad (2)$$

where ω_{t_f} and ω_{t_b} are the probabilities of the foreground and background classes, respectively separated by k , while σ_{t_f} and σ_{t_b} are the intraclass variances of these two classes. The class probability $\omega_{t_f}, \omega_{t_b}(k)$ is found from the L bins of thermal image histogram

$$\omega_{t_f}(k) = \sum_{i=0}^{k-1} p(i) \quad (3)$$

$$\omega_{t_b}(k) = \sum_{i=k}^{L-1} p(i). \quad (4)$$

In a binary classification problem, minimizing the intraclass variance (data spread within a class) is equivalent to maximising the interclass variance (variation between classes). Hence

$$\begin{aligned} \sigma_b^2(k) &= \sigma^2 - \sigma_w^2(k) = \omega_{t_f}(\mu_{t_f} - \mu_K)^2 + \omega_{t_b}(\mu_{t_b} - \mu_K)^2 \\ &= \omega_0(k)\omega_1(k)[\mu_0(t) - \mu_1(k)]^2 \end{aligned} \quad (5)$$

which is exposed in terms of class probabilities ω and class means μ , where $\mu_{t_f}(k)$, $\mu_{t_b}(k)$, and μ_K represent

$$\mu_{t_f}(k) = \frac{\sum_{i=0}^{k-1} ip(i)}{\omega_{t_f}(k)} \quad (6)$$

$$\mu_{t_b}(k) = \frac{\sum_{i=k}^{L-1} ip(i)}{\omega_{t_b}(k)} \quad (7)$$

$$\mu_K = \sum_{i=0}^{L-1} ip(i). \quad (8)$$

Finding ω and μ iteratively, and k for each thermal scene, the algorithm would be capable of separating the acquired TSA's output into background and foreground (human presence) classes. However, this method assumes that the histogram of temperature values has a bi-modal distribution, which is not the case in real-world scenarios, where perhaps another warm object is in the sensor's FoV, e.g., a Hot kettle yields three temperature peaks as shown in the surface plot of the acquired thermal image and its corresponding histogram of Fig. 4(a). To overcome this limitation, a modification prior to applying this method is introduced by using a temperature filter to convert those temperatures higher than the TSA-based human empirical value of 33 °C to the minimum temperature in the acquired TSA output as follows:

$$x_t = \begin{cases} x_i & \text{for } x_i \leq 33 \\ \min(x_t) & \text{otherwise} \end{cases} \quad \text{where } x_i = \text{FIR}. \quad (9)$$

Fig. 4(b) illustrates the result of applying this modification filter on the original acquired TSA's output visualized in the surface plot and its corresponding histogram in Fig. 4(a), while Fig. 4(c) shows the result of the complete pre-processing phase.

C. Motion Feature Extraction

Two optical flow estimation methods are utilized to support motion analysis and model evaluation. The first method proposed by Horn and Schunck [36] is a sparse optical flow

estimation algorithm that estimates the global optical flow. The second algorithm is proposed by Farneback [37], a dense optical flow estimation algorithm that computes the local optical flow of the movements in the acquired scene. Given a particular object in motion at time t , with pixel points x and y . Displacement of the object by Δx and Δy over Δt forms a new image expressed

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (10)$$

considering Taylor series method of approximation

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t \quad (11)$$

hence

$$\frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t = 0 \quad (12)$$

once it is dividing by Δt

$$\frac{\partial I}{\partial x} \frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y} \frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t} \frac{\Delta t}{\Delta t} = 0 \quad (13)$$

results in

$$\frac{\partial I}{\partial x} u_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} = 0 \quad (14)$$

therefore

$$I_x u_x + I_y v_y + I_t = 0. \quad (15)$$

where u_x, v_y indicate the velocity of the frame in the horizontal and vertical directions, respectively, and the derivatives of x and y over time t are represented by I_x, I_y, I_t . Hence, (15) has two unknowns, u and v , which can be solved using various mathematical methods. For instance, in the Horn and Schunck (global optical flow estimation) algorithm [36], the velocity estimation between consecutive motion frames is based on two assumptions: brightness and smoothness. For consecutive pixels within two frames, shadows are neglected, and the direction of pixels is the same. Under the smoothness assumption, the derivative of u and v with reference to the x and y directions is calculated in (16). Therefore, the expression of this global optical flow estimation can be expressed as follows:

$$E = (I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \quad dx dy \quad (16)$$

where $(I_x u + I_y v + I_t)^2$ indicates brightness constancy along x, y , and t dimensions, while α indicates the weighing factor that determines the brightness and smoothness values.

$(\|\nabla u\|^2 + \|\nabla v\|^2) dx dy$ indicates derivatives of u and v with reference to the x and y directions. The smoothness assumption is that all pixels in a particular neighborhood of a thermal image are observed to move in the same direction.

In contrast to the global optical flow algorithm, the local estimation of the optical flow [37] does not aim to solve the optical flow equation expressed in (15). Instead, it considers quadratic polynomial expansion. For each pixel within an image frame, there is a polynomial approximation of the neighborhood. The velocity of a pixel upon displacement is determined by

minimizing the error function $e(X)$ of the neighborhood area expressed in the equation as follows:

$$e(X) = \sum_{\Delta D \in D} \omega(\Delta D) \|A_M(X + \Delta D) \Delta X - \Delta b(X + \Delta D)\|^2 \quad (17)$$

where D indicates the neighborhood area, ΔX signifies the pixel displacement, $\omega(\Delta D)$ indicates the Gaussian weighting function that calculates the degree of the neighborhood area. The higher the Gaussian weighting value, the closer the target pixel. This article constructs three pyramid levels to support the motion estimation between pixels with larger displacement in the local optical flow estimation. A high pyramid level indicates high pixel displacement. A 0.5 pyramid scale is specified at each level, which defines the down-sampling rate with three iterations and a neighborhood size of 5.

Both optical flow estimations return the magnitude and orientation of the movement in a sequence of frames. Since the proposed approach aims to analyse the velocity and direction of objects in motion, no motion is observed for all background static objects. It implies that the magnitude and orientation vectors are low for static objects in the background. Therefore, each activity obtains the highest k magnitude and its corresponding orientation during feature extraction. The elimination of unnecessary background optical flow vectors in the preprocessing phase makes the algorithm more robust because the potential ambient thermal noise is filtered. Fig. 5 shows a sample of applying a global optical flow estimation of a sample of fall activity on the preprocessed TSA's output. The length of the arrows shown in Fig. 5(a)–(d) represents the magnitude of the velocity of human motion. Besides, Fig. 5(a) and (b) shows the human falling toward the ground, while Fig. 5(c) shows that the human body is about to completely touch the ground, and Fig. 5(d) shows a slowdown in human movement and the gradual transition to the state of inactivity.

D. Human ADL Recognition and Fall Detection

To detect an abnormal behavior (human fall in this study), the proposed system should be capable of recognising associated normal ADL activities, e.g., walking, standing, sitting etc. Therefore, two recurrent neural network (RNN) architectures are implemented separately, specifically, the LSTM [38] and Bi-LSTM [39] with adaptive moment estimation (ADAM) [40] and stochastic gradient descent with momentum (SGDM) [41] optimization algorithms during the training phase to perform ADL recognition including the fall detection. Hence, these two networks can perform a sequence frame classification of the extracted motion vectors from the previous step, which was the justification behind this classification approach in this work.

Each network architecture consists of input, hidden, and output layers. Unlike the LSTMs, the Bi-LSTM network is designed to support both future and past sequence training. Technically, A Bi-LSTM supports two LSTM layers, including forward and backward, arranged parallel to each other. Therefore, its operation is similar to the LSTM, except that the motion sequence of human movement is passed in both directions during the training

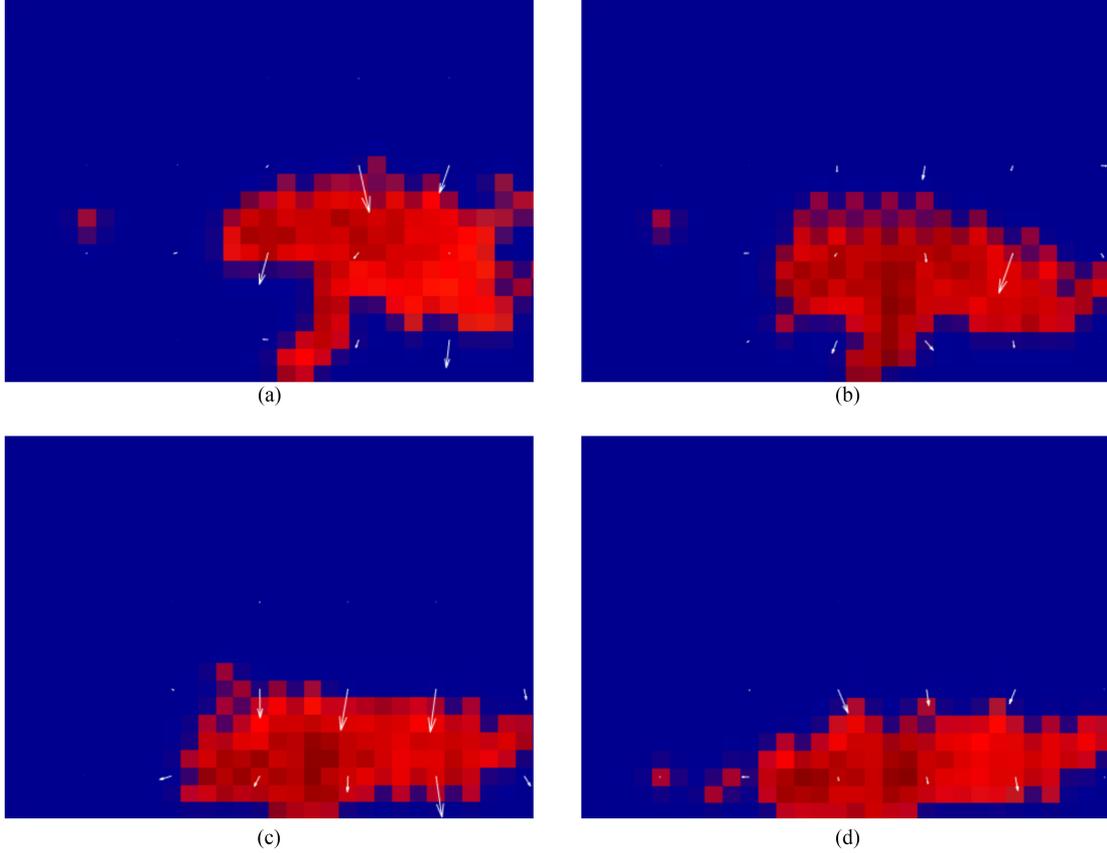


Fig. 5. Illustrative result of optical flow features on preprocessed TSA output for a subset of the human fall motion sequence (a) the fall of a human, (b) the acquired following human fall sequence, (c) the human body is about to completely touch the ground, (d) human motion speed decreases significantly indicating a state of human inactivity.

stage, described as follows:

$$\vec{h}_t = \text{LSTM}(x_t, \vec{h}_{t-1}; \vec{W}) \quad (18)$$

$$\overleftarrow{h}_t = \text{LSTM}(x_t, \overleftarrow{h}_{t-1}; \overleftarrow{W}) \quad (19)$$

$$y_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (20)$$

where y_t specifies the output, \vec{h}_t , \overleftarrow{h}_t signify the forward and backward LSTM layers, \vec{W} and \overleftarrow{W} signify the weights at the forward and backward LSTM layers. In addition to the above layers, each network architecture includes various training options. These include: 100 max epoch corresponds to the total training period, mini-batch size of 27 that indicates the number of iterations per epoch. The initial weights for the network are initialised using Glorot Initialiser [42].

Once a fall is detected, a human-interaction interface is triggered to ensure that the proposed approach is held accountable by human confirmation requirements before the fall is reported to the information support. In this article, a mobile interface is used as a human-interaction interface that asks the user in the form of an alert to cancel the fall during time t as shown in Fig. 6. Suppose the fall is not confirmed within the specified time. In that case, the system will automatically report the fall to the information support in the mean of emergency alert or notification to family members, etc. This human-in-the-loop approach

for fall detection would enable to overcome false-positive fall detection alerts, for example, to the emergency services while keeping the false-negative fall detection as low as possible.

IV. EXPERIMENTS

To evaluate the performance of the proposed methodology, comprehensive experimental work has been performed with different use classification networks, optimizers, and optical flow algorithms. The data collection stage consists of eight participants (six males and two females). The participants acted fall incidents during three ADL, including walking, sitting from a stand position, and standing from a sit position. Each participant carried out a particular activity five times. A total of 1282 image frames were captured as part of the dataset used for the experiments, which consisted of 226 falling forward image frames, 321 sitting, 288 standing, and 447 walking frames. During the data collection, the sensor was placed at 0.5 m from the ground and aligned with the wall. A detailed description of the experimental results is reported as follows.

A. Experiments 1

The first experiment involves the identification of falls in the presence of all other ADL. The dataset is divided into 70% for training and 30% for testing. Classification of local optical

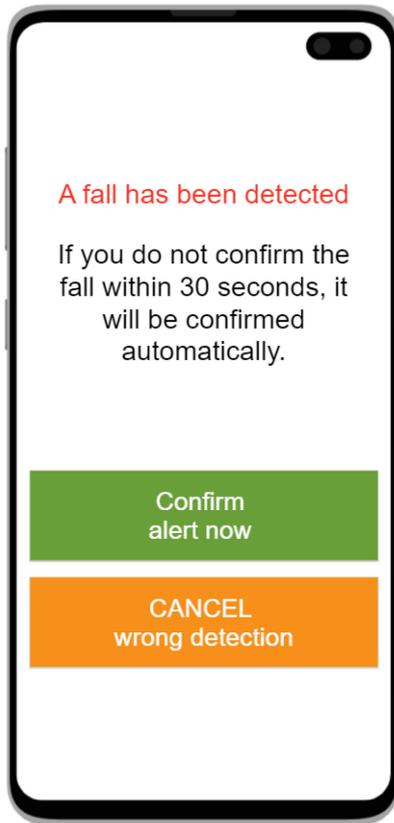


Fig. 6. Mobile-human interaction interface to confirm the detected fall prior to reporting it to the information support to minimize the false-positive alerts while keeping low false-negative cases.

flow extracted features using LSTM with ADAM optimizer yields 91.6% compared to SGDM optimizer at 91.8%. On the other hand, the classification of Horn and Schunck optical flow extracted features using LSTM with ADAM optimizer achieves performance at 99.4% compared to SGDM at 98.3%. Using Bi-LSTM network to classify Local optical flow extracted features with ADAM optimizer displays performance at 94.0% compared to SGDM at 84.9%. Finally, classification of Global optical flow extracted features using ADAM optimizer yields performance at 99.7% compared to SGDM at 86.3%. A summary of the results is shown in Fig. 7(a).

It can be observed from Fig. 7(a) that the ADAM optimizer gives better performance than the SGDM for all sequence classification algorithms. Not to mention, the Global motion extracted vectors are observed to provide better motion features for classification. The ability of the Global algorithm to process selected pixels within an image during motion estimations is an attribute of its outstanding performance.

B. Experiments 2

To further evaluate the performance of the developed model, the second set of experiments is performed. It includes two sets of activities, namely, falling forward and walking. Similar

to the first use case scenario, the performance of the preprocessing techniques is evaluated for each experiment. A dataset of 673 thermal frames was used, which consisted of 226 fall frames and 447 walk frames. Different optical flow methods are deployed and classified using LSTM and Bi-LSTM with different optimizers ADAM and SGDM. Classification of local optical flow extracted features using Bi-LSTM with ADAM optimizer yields 96.5% compared to SGDM at 86.0%. On the other hand, the classification of Global optical flow extracted features using Bi-LSTM with ADAM optimizer achieves performance at 35.1% compared to SGDM at 94.4%. Using LSTM network to classify local optical flow extracted features with ADAM optimizer displays performance at 67.5%, compared to SGDM at 90.7%. Finally, classification of Global optical flow extracted features using ADAM optimizer yields 86.1% compared to SGDM at 65.2%. A summary of the results is shown in Fig. 7(b).

C. Experiments 3

This experiment aims at identifying fall frames from a human sitting position. A dataset of 547 thermal frames was used, which consisted of 226 321 fall and sitting position frames. Similar to the previous scenarios, the performance of various parameters, including preprocessing techniques, classification algorithms, optimizers, and optical flow methods, is performed. Classification of local optical flow extracted features using Bi-LSTM with ADAM optimizer yields 85.7% compared to SGDM at 91.0%. On the other hand, the classification of Global optical flow extracted features using Bi-LSTM with ADAM optimizer achieves performance at 61.0% compared to SGDM at 99.9%. Using LSTM network to classify Local optical flow extracted features with ADAM optimizer displays performance at 42.7%, compared to SGDM at 83.3%. Finally, classification of Global optical flow extracted features using ADAM optimizer yields 99.6% compared to SGDM at 55.8%. A summary of the results is shown in Fig. 7(c).

D. Experiments 4

Finally, the performance of the proposed approach is evaluated using fall and stand frames. A dataset of 514 thermal frames was used, which consisted of 226 288 human fall and standing frames, respectively. Classification of local optical flow extracted features using Bi-LSTM with ADAM optimizer yields 91.1% compared to SGDM at 72.3%. On the other hand, the classification of Horn and Schunck optical flow extracted features using Bi-LSTM with ADAM optimizer achieves performance at 99.8% compared to SGDM at 94.5%. Using LSTM network to classify local optical flow extracted features with ADAM optimizer displays performance at 58.3%, compared to SGDM at 88.5%. Finally, classification of Global optical flow extracted features using LSTM with ADAM optimizer yields 95.3% compared to SGDM at 98.6%.

The results shown in Fig. 7(d) indicate Global motion extracted features with ADAM optimizer as the best performer during the classification of fall and stand activities.

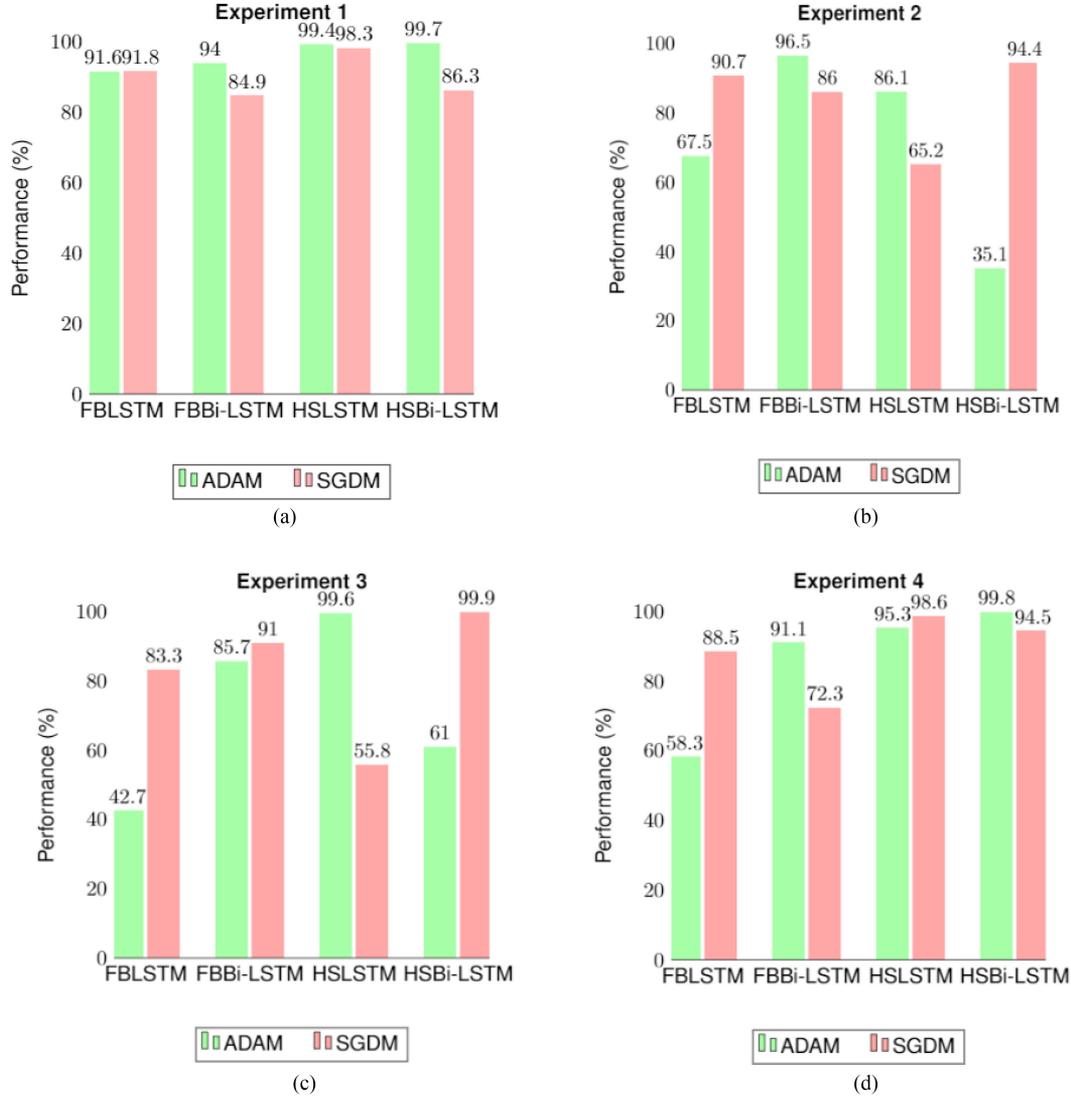


Fig. 7. Summary of experimental results for (a) detecting human falls among all other ADLs, (b) fall and walk, (c) fall and sitting, (d) fall and stand where FBLSTM indicates the classification of Farneback (local) optical flow extracted features with LSTM, FBBI-LSTM indicates the classification of Farneback optical flow extracted features with Bi-LSTM, HSLSTM indicates the classification of Horn and Schunck (global) optical flow extracted features with LSTM and HSBi-LSTM indicates the classification of Horn and Schunck optical flow extracted features with Bi-LSTM.

TABLE III
COMPARISON OF EVALUATIVE EXPERIMENTS PRIOR TO APPLYING THE PREPROCESSING TECHNIQUES USING LSTM FOR MOTION SEQUENCE CLASSIFICATION

Optical flow algorithm	optimiser	Evaluation 1	Evaluation 2	Evaluation 3	Evaluation 4
Farneback	SGDM	82.3%	57.8%	75.5%	59.6%
Franeback	ADAM	96.2%	99.9%	98.3%	87.7%
Horn-Schunck	SGDM	95.7%	98.7%	99.9%	45.2%
Horn-Schunck	ADAM	97.8%	98.5%	42.7%	45.2%

V. EXPERIMENTAL EVALUATION AND ANALYSIS

This section Presents an evaluation of the conducted experiments versus the proposed approach to validate the importance of applying the proposed preprocessing techniques on the TSA's output for human motion-based applications. Therefore,

throughout these evaluative experiments, the suggested preprocessing techniques are not applied to compare them with the results described in Section IV. Hence, the same dataset settings were used to repeat each evaluative experiment.

Tables III and IV show the results of evaluating the proposed approach prior applying the preprocessing techniques on conducted experiments described in Section IV using LSTM

TABLE IV
COMPARISON OF EVALUATIVE EXPERIMENTS PRIOR TO APPLYING THE PREPROCESSING TECHNIQUES USING BI-LSTM FOR MOTION SEQUENCE CLASSIFICATION

Optical flow algorithm	optimiser	Evaluation 5	Evaluation 6	Evaluation 7	Evaluation 8
Farneback	SGDM	92.3%	80.9%	75.2%	57.2%
Franeback	ADAM	91.6%	86.9%	69.7%	45.2%
Horn-Schunck	SGDM	91.4%	98.1%	99.7%	98.6%
Horn-Schunck	ADAM	84.5%	37.9%	51.0%	97.3%

and Bi-LSTM, respectively. It can be observed from these tables, the application of the preprocessing techniques offers an increase in performance of Global motion extracted features with LSTM and ADAM optimizer from 97.8% to 99.4%. In addition, the performance of Global motion extracted features with Bi-LSTM and ADAM solver increased from 84.5% to 99.7%. Second, the performance of local motion extracted features with Bi-LSTM and ADAM increased from 86.9% to 96.5% after applying the preprocessing techniques. Third, the classification of Global extracted features with LSTM and ADAM indicate increased performance from 42.7% to 99.6%. Finally, the sequence motion classification of Global motion extracted features with LSTM and ADAM optimizer indicates an increase in performance from 45.2% to 95.3%. In addition, the classification of Global motion extracted vectors with Bi-LSTM and ADAM optimizer indicates increased performance from 97.3% to 99.8%. It can be concluded from these empirical evaluations that using the proposed preprocessing technique on TSA outputs for human motion based application is necessary to achieve a high-performance and robust system against ambient thermal noise or noises induced by human movements. Besides, LSTM performs better than Bi-LSTM with ADAM optimizer in the conducted experiments.

VI. DISCUSSION

A deployable domestic human behaviour monitoring solution that can meet the urgency of economic and societal requirements of older adults should be accepted by the older adults themselves and the care service providers. That is, human-centred systems for domestic monitoring should meet the following stakeholders' acceptability factors.

- 1) Impacts, the proposed system should contain applicable solutions that have real economic and/or social impacts.
- 2) Privacy-preservation, the system should maintain the privacy of its users.
- 3) Reliability, the system should be reliable to perform its tasks in realistic domestic environments that may contain more than one occupant.
- 4) Convenience, the system should operate autonomously without interfering with normal human activities at a reasonable installation cost.
- 5) Accountability, systems should be accountable to the users.

Fall detection is abnormal human activity that can occur from standing, sitting or even unpredicted activity. This article has addressed a valid issue concerning the users' accountability to the system's decision. On the other hand, this

article continues our research on TSA-based human behavior monitoring to achieve the mentioned acceptability factors. In particular, the use of low-resolution TSA has been claimed as a privacy-preserving approach based on the assumption that no identifiable information can be extracted from the raw sensor outputs. However, our previous research work [43] found a linear relationship between low- and high-resolution thermal imaging. Thus, low-resolution signals can be mapped to high-resolution signals to extract human-identifiable information. Replacing the TSA's signal output with the proposed motion features enhances the privacy-preserving quality of TSA-based human monitoring schemes.

Unlike conventional cameras, TSA is not sensitive to light, but it is sensitive to ambient temperature. This paper ensures the system's reliability by proposing appropriate preprocessing techniques to enable the proposed approach to operate in a noisy thermal environment. The proposed preprocessing methods have been thoroughly evaluated to empirically validate their effectiveness through running the recognition experiments with and without applying the suggested preprocessing methods. Besides, the results of the used series-based learning algorithms show the reliability of the proposed solution in detecting human falls among various ADLs. The performance variance in empirical evaluations shown in Tables III and IV assesses the robustness of the proposed preprocessing techniques. The difference is due to the degree of thermal noise induced by human movements. The low difference between the results with or without preprocessing indicates less significant thermal noise, possibly due to slow human motion.

Although this article suggests placing the sensor at a short height from the floor for a wider inspection area, this placement may have its drawbacks, such as FoV getting cluttered with home furniture. In this line, our previous work [14], [27] has already contributed to resolving issues related to sensor placement flexibility, operating distance, and multi-occupancy environments. Furthermore, we have also explored the ability of the proposed motion analysis to identify the motion of the same human subject obtained from multiple sensors' placements in a multioccupancy environment [43].

The rationale behind suggesting the smartphone as a human interaction modality to confirm the detected falls is due to the fact that most people have smartphones and would therefore be more apt to use the technology they already have rather than adding an extra cost and effort with unfamiliar modality. Besides, the proposed modality could be a feature of an existing mobile healthcare tracking application, e.g., the U.K. National Health Service (NHS) mobile app. However, some older adults may not have smartphones to confirm the detection of human falls that



Fig. 8. Visual representation of the experimental results of human fall detection among all other ADLs in the form of a confusion matrix, (a) using Bi-LSTM with local optical flow with ADAM optimizer, (b) Bi-LSTM with the local optical flow with SGDM optimizer, (c) LSTM with Global optical flow estimation with Adam optimizer, and (d) LSTM with Global optical flow estimation with SGDM optimizer.

require urgent responses. In this case, there are no restrictions to switching the confirmation from a mobile notification to an automated landline call to confirm or cancel the detected fall.

Comprehensive experiments with robustness analysis were conducted to evaluate the performance of the proposed approach. Fig. 8 shows the confusion matrices of the obtained results on detecting the human falls among ADL activities performed with different human subjects. In particular, Fig. 8(a) and (b) use a Bi-LSTM network with ADAM and SGDM optimizers, respectively. While Fig. 8(c) and (d) show the results using the LSTM network with ADAM and SGDM. It can be observed from these figures that the false positive is higher when the Bi-LSTM network is used. Although the false-positive results in this experiment are zero, this does not mean that the network will always provide zero false positives while using the LSTM network since human falls can occur after unexpected human activity. Moreover, human fall data is collected from pretending human falls, since it is very difficult to obtain actual human fall incidents. Thus, this article moves a step forward to propose a human-in-the-loop approach to confirm human fall before it is reported to a centralised information support system to enhance system reliability and acceptability by considering the human accountability factor.

This article has significant social and economic impacts on various sectors, including the older adults' community. Therefore, it is important to conduct a pilot testing of the proposed approach under real-time operating conditions through the actual deployment of the proposed approach in older adults' homes for a period of time to evaluate the feasibility, time, cost, risk, and performance of the proposed approach before the widespread deployment.

VII. CONCLUSION

TSA bridges the gap between privacy and performance in human-centred applications, including fall detection and ADL recognition. This article proposes a motion-based approach for human-in-the-loop fall detection using a low-resolution TSA output. Including a human interactive model to confirm the fall detection results in a very positive impact on reducing the false positive fall detection cases while keeping the false-negative predictions as low as possible. By doing so, the potential deployability and reliability of the system are enhanced as fewer false fall alerts are reported to emergency or information support, resulting in a significant efficiency utilization of the resources.

Comprehensive experiments and evaluations were conducted to validate the performance of the proposed approach using local and global optical flow methods, different motion sequence classification algorithms and optimizers. The achieved results indicate that, in general, global optical flow estimation has better performance results with the low-resolution TSA output. The use of optical flow with this type of sensor overcomes the challenges of such sensors operating in a noisy thermal environment as ambient thermal noises have a shallow motion velocity compared to human motions. Besides, the proposed preprocessing techniques applied to low-resolution thermal images achieve better system performance, making the approach more robust against thermal noises induced by human movement. Future work can be conducted to utilise the proposed approach to sleep monitoring and abnormal human behavior detection.

REFERENCES

- [1] WHO, "Who global report on falls prevention in older age," World Health Organization, Tech. Rep., 2008. [Online]. Available: http://www.who.int/ageing/publications/Falls_prevention7March.pdf
- [2] World Health Organization, "Step safely: Strategies for preventing and managing falls across the life course," 2021. [Online]. Available: <https://apps.who.int/iris/handle/10665/340962>
- [3] F. Riquelme et al., "Ehomeniors dataset: An infrared thermal sensor dataset for automatic fall detection research," *Sensors*, vol. 19, no. 20, 2019, Art. no. 4565.
- [4] A. Shahzad and K. Kim, "Faldroid: An automated smart-phone-based fall detection system using multiple kernel learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 35–44, Jan. 2019.
- [5] H. Sadreazami, M. Bolic, and S. Rajan, "Contactless fall detection using time-frequency analysis and convolutional neural networks," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6842–6851, Oct. 2021.
- [6] M. Mubashir, L. Shao, and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, pp. 144–152, 2013.
- [7] J. Adolf et al., "Deep neural network based body posture recognitions and fall detection from low resolution infrared array sensor," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2018, pp. 2394–2399.
- [8] W.-H. Chen and H.-P. Ma, "A fall detection system based on infrared array sensors with tracking capability for the elderly at home," in *Proc. IEEE 17th Int. Conf. E-Health Netw., Appl. Serv.*, 2015, pp. 428–434.
- [9] Z. Chen and Y. Wang, "Infrared-ultrasonic sensor fusion for support vector machine-based fall detection," *J. Intell. Mater. Syst. Struct.*, vol. 29, no. 9, pp. 2027–2039, 2018.
- [10] S. Mashiyama, J. Hong, and T. Ohtsuki, "A fall detection system using low resolution infrared array sensor," in *Proc. IEEE 25th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun.*, 2014, pp. 2109–2113.
- [11] Z. Liu et al., "Fall detection and personnel tracking system using infrared array sensors," *IEEE Sensors J.*, vol. 20, no. 16, pp. 9558–9566, Aug. 2020.
- [12] Y. Ogawa and K. Naito, "Fall detection scheme based on temperature distribution with IR array sensor," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2020, pp. 1–5.
- [13] S. Shelke and B. Aksanli, "Static and dynamic activity detection with ambient sensors in smart spaces," *Sensors*, vol. 19, no. 4, 2019, Art. no. 804.
- [14] A. Naser, A. Lotfi, and J. Zhong, "Adaptive thermal sensor array placement for human segmentation and occupancy estimation," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1993–2002, Jan. 2021.
- [15] A. Naser et al., "Heat-map based occupancy estimation using adaptive boosting," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2020, pp. 1–7.
- [16] Y. Nizam, M. N. H. Mohd, and M. M. A. Jamil, "A study on human fall detection systems: Daily activity classification and sensing techniques," *Int. J. Integr. Eng.*, vol. 8, no. 1, pp. 35–43, 2016.
- [17] S. B. Khojasteh et al., "Improving fall detection using an on-wrist wearable accelerometer," *Sensors*, vol. 18, no. 5, 2018, Art. no. 1350.
- [18] F. Hussain et al., "Activity-aware fall detection and recognition based on wearable sensors," *IEEE Sensors J.*, vol. 19, no. 12, pp. 4528–4536, Jun. 2019.
- [19] A. Lotfi et al., "Supporting independent living for older adults; employing a visual based fall detection through analysing the motion and shape of the human body," *IEEE Access*, vol. 6, pp. 70272–70282, 2018.
- [20] F. Lezzar, D. Benmerzoug, and I. Kitouni, "Camera-based fall detection system for the elderly with occlusion recognition," *Appl. Med. Informat.*, vol. 42, no. 3, pp. 169–179, 2020.
- [21] N. F. Ribeiro and C. P. Santos, "Two fall-related and kinematic data-based approaches for an instrumented conventional cane," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 5, pp. 554–563, Oct. 2021.
- [22] S.-B. Jeon et al., "Self-powered fall detection system using pressure sensing triboelectric nanogenerators," *Nano Energy*, vol. 41, pp. 139–147, 2017.
- [23] A. Yazar et al., "Fall detection using single-tree complex wavelet transform," *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 1945–1952, 2013.
- [24] A. Hayashida, V. Moshnyaga, and K. Hashimoto, "The use of thermal ir array sensor for indoor fall detection," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2017, pp. 594–599.
- [25] C. Taramasco et al., "A novel monitoring system for fall detection in older people," *IEEE Access*, vol. 6, pp. 43563–43574, 2018.
- [26] S. Tateno et al., "Privacy-preserved fall detection method with three-dimensional convolutional neural network using low-resolution infrared array sensor," *Sensors*, vol. 20, no. 20, 2020, Art. no. 5957.
- [27] A. Naser, A. Lotfi, and J. Zhong, "Towards human distance estimation using a thermal sensor array," *Neural Comput. Appl.*, pp. 1–11, 2021.
- [28] A. Naser, A. Lotfi, and J. Zhong, "A novel privacy-preserving approach for physical distancing measurement using thermal sensor array," in *Proc. 14th Pervasive Technol. Related Assistive Environ. Conf.*, 2021, pp. 81–85.
- [29] S. Savazzi et al., "Processing of body-induced thermal signatures for physical distancing and temperature screening," *IEEE Sensors J.*, vol. 21, no. 13, pp. 14168–14179, Jul. 2021.
- [30] S. Arvidsson et al., "Sensor fusion and convolutional neural networks for indoor occupancy prediction using multiple low-cost low-resolution heat sensor data," *Sensors*, vol. 21, no. 4, 2021, Art. no. 1036.
- [31] Y. Yuan et al., "Occupancy estimation in buildings based on infrared array sensors detection," *IEEE Sensors J.*, vol. 20, no. 2, pp. 1043–1053, Jan. 2020.
- [32] M. Cokbas, P. Ishwar, and J. Konrad, "Low-resolution overhead thermal triwire for occupancy estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 88–89.
- [33] A. Naser et al., "Human activity of daily living recognition in presence of an animal pet using thermal sensor array," in *Proc. 13th ACM Int. Conf. Pervasive Technol. Related Assistive Environ.*, 2020, pp. 1–6.
- [34] C. Yin et al., "Device-free human activity recognition with low-resolution infrared array sensor using long short-term memory neural network," *Sensors*, vol. 21, no. 10, 2021, Art. no. 3551.
- [35] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
- [36] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 13, pp. 185–203, 1981.
- [37] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Proc. Scand. Conf. Image Anal.*, 2003, pp. 363–370.
- [38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [39] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
- [40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [41] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.
- [42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Stat., Workshop Conf. Proc.*, 2010, pp. 249–256.
- [43] A. Naser, A. Lotfi, and J. Zhong, "Calibration of low-resolution thermal imaging for human monitoring applications," *IEEE Sens. Lett.*, vol. 6, no. 3, pp. 1–4, Mar. 2022, Art. no. 7000904.