Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

Enhancing Elderly Safety: A Study on Vision-Based Monitoring for Abnormal Behaviour Detection

Nian Chi Tay[1], Tee Connie[2], Thian Song Ong[3]
[1,2,3] Multimedia University, Malaysia

Abstract:- This study addresses the crucial need for innovative approaches in supporting the well-being of the rapidly aging global population by focusing on Abnormal Behaviour Detection (ABD) in Activities of Daily Living (ADL) for the elderly. Emphasizing the inadequacies of traditional monitoring methods, our research employs computer vision to accurately identify abnormal behaviours such as changes in activity patterns and falls. Central to this study is the evaluation of state-of-the-art models that takes into the account of temporal information, including Convolutional Long Short-Term Memory (ConvLSTM), Long-term Recurrent Convolutional Network (LRCN), and Convolutional Neural Network - Gated Recurrent Unit (CNN-GRU). The key contributions of this study include the generation of synthetic abnormal behaviour data using the limited publicly available dataset as well as the analysis of the best model, ConvLSTM by enhancing the transparency and understanding of its predictive mechanisms. This research not only advances the academic understanding of elderly care in vision-based approach but also offers practical insights for enhancing safety and well-being in various living environments for older adults.

Keywords: *Abnormal Behaviour Detection, Activity of Daily Living, ConvLSTM, LRCN, CNN-GRU*

1. Introduction

The rapid aging of the global population underscores the urgent need for innovative technologies that support independent living and enhance the well-being of older adults. In this context, ABD in ADL has emerged as a promising research area with significant potential to revolutionize care for this population [1].

Traditional methods of monitoring, such as manual observation, are often inefficient and prone to overlooking subtle changes in individual behaviour, leading to delayed interventions and potentially negative health outcomes. ABD addresses these limitations by continuously monitoring individuals' ADL patterns using sensors, wearable devices, and computer vision. By automatically identifying deviations from established norms, referred to as "abnormal behaviours" [2], such as changes in activity duration or sequence, unusual intensity, repetitive actions, or falls, ABD offers several compelling benefits.

The benefits of ABD systems extend beyond assisted living facilities, offering valuable contributions in diverse settings such as smart homes and hospitals. In smart homes, integration of ABD systems allows for remote monitoring of individuals, further enhancing their safety and security [1]. In hospitals and other healthcare settings, continuous monitoring of patients' ADL patterns can contribute to early detection of complications and improve patient safety [3].

This paper specifically focuses on the application of ABD in ADL for elderly individuals, with a particular interest in the detection of two specific abnormal behaviours: (a) falling and (b) dangerous act like leaving the stove on but proceed with doing other activities. Falls are a major cause of injury and disability among older adults, and early detection through ABD systems can enable timely intervention and potentially prevent serious injuries. Dangerous acts involve leaving the stove on but proceed in carrying out other activities such as laying down, watching TV and reading book, which are extremely dangerous and thus classified as abnormal behaviours.

By investigating these specific abnormal behaviours without having the fuss of wearing sensors of different natures, this paper aims to contribute to improved understanding of ADL patterns in elderly by using vision-based input. The contributions of this paper include: (1) the innovative synthesis of normal behaviours to construct scenarios of abnormal behaviour, providing a more nuanced understanding of abnormal behaviour within the context of ADL and (2) the implementation of a comprehensive comparative study examining the efficacy of spatio-temporal models such as ConvLSTM, LRCN, and CNN-GRU in behaviour analysis.

2. Related Work

To address the growing need for efficient and reliable methods for monitoring and caring for elderly, researchers have explored various approaches. For sensor-based input, The authors [4] used Graph Convolutional Networks to detect cognitive impairment indicators. While the approach effectively utilizes advanced neural network architectures, it may face challenges in handling diverse and complex real-world data. Recurrent Neural Networks (RNN) were employed for activity recognition and abnormal behaviour detection by [5]. This method excels in sequential data processing, but its complexity can lead to difficulties in training and interpretation. CNN was utilized in [6] for detecting abnormal behaviour in dementia sufferers. CNNs are powerful in feature extraction from large datasets, though they require substantial computational resources and may not generalize well across different types of data. [7] applied Random Forests and Markov Chains for modeling normal behaviour and detecting deviations in smart home environments. This integrated approach offers a comprehensive analysis of behaviour patterns, but it may struggle with the privacy concerns and the variability of smart home environments . The study [8] involved using transfer learning and recursive auto-encoders for abnormal behaviour detection in dementia sufferers. This method leverages pre-trained models for efficiency, but it may face limitations in adapting to specific conditions of different dementia cases.

For vision-based input, the researchers in [9] and [10] processed dynamic images extracted from extended video footage for early detection of human falls. Authors from [11] described a system using a wireless camera to record daily activities of elderly individuals, focusing on identifying abnormal behaviours such as slips and falls. In study [12], data was gathered using a wireless camera network, where sensors were categorized as either functional or regular based on their placement in the home. The study in [13] transformed image sequences into motion history images for improved efficiency in human fall detection models. Depth images, commonly used in human action recognition, were applied in a machine learning model in study [14]. Additionally, [15] introduced a five-point inverted pendulum model based on human key points to detect fall behaviours. One notable study by [16] investigated the feasibility of using lifelog data and ML to detect abnormal behaviours in dementia patients residing in nursing homes. EDOF condenses video data into a single image, making fall detection faster and more accurate. This approach, while effectively addressing the challenges of dynamic lighting and reducing processing time, may require significant computational resources and could be limited by the availability and diversity of training data.

Both sensor-based and vision-based inputs have their respective strengths and weaknesses; however, since the target group of this study is elderly, we have chosen to focus on vision-based input as the primary method for our study to avoid the necessity of wearing sensors of various types.

3. Research Methodology

This paper studies ABD of elderly in ADL with three approaches: (a) convLSTM, (b) LRCN and (c) CNN-GRU. The overall framework of ABD of elderly in ADL is depicted in Fig. 1.

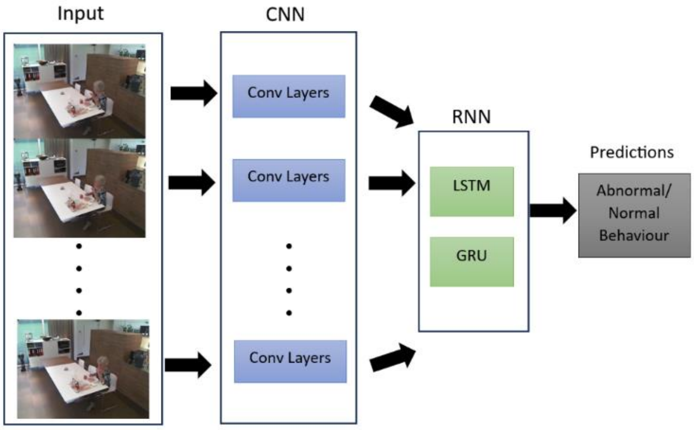

Figure 1 Framework of ABD of Elderly in ADL

The model framework depicted integrates sequential video frames as input through a series of convolutional (Conv) layers within a CNN to extract spatial features. These features are then fed into a Recurrent Neural Network (RNN) comprising LSTM or GRU layers to analyze temporal dependencies. The LSTM layer is designed to remember longer-term sequences, while the GRU layer focuses on the short-term, enabling the model to effectively understand both immediate and extended patterns in the data. The output from the RNN is then used to make final predictions regarding the nature of the behaviour exhibited in the video, distinguishing between normal and abnormal behaviour.

A. ConvLSTM

The ConvLSTM [17] addresses the main limitation of the Fully Connected LSTM (FC-LSTM), which traditionally unfolds inputs into 1D vectors, resulting in a loss of spatial information. The ConvLSTM model, in contrast, maintains the spatial integrity of the data by processing the inputs (X_1... X_t), cell outputs (C_1... C_t), hidden states (H_1... H_t), and gates (i_t, f_t, o_t) as 3D tensors, with the last two dimensions representing spatial dimensions (rows and columns). This design depicted in Fig. 2 allows the ConvLSTM to predict the future state of a cell in a spatial grid based on the inputs and past states of its neighboring cells. The process is facilitated by a convolution operator used in both state-to-state and input-to-state transitions.

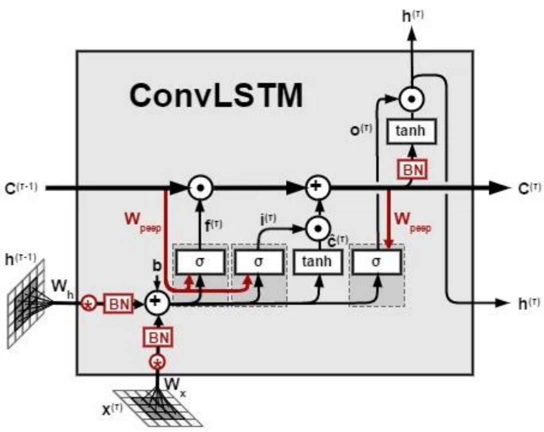

Figure 2 Structure of ConvLSTM [18]

The key equations of ConvLSTM, where '∗' denotes the convolution operator and '∘' represents the Hadamard product, are as follows:

$$
\begin{aligned}
i_t &= \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i) \\
f_t &= \sigma\big(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f\big) \\
\mathcal{C}_t &= f_t \circ \mathcal{C}_{t-1} + i_t \circ \tanh\,(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c) \\
o_t &= \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o) \\
\mathcal{H}_t &= o_t \circ \tanh\,(\mathcal{C}_t)
\end{aligned}
$$

B. LRCN

The LRCN model combines CNNs for feature extraction with LSTM networks for sequence prediction, creating an end-to-end trainable system for visual recognition and description tasks. This method allows for efficient processing of sequential visual data, such as videos, and can generate variable-length outputs like natural language descriptions. The integration of CNN and LSTM enables the LRCN model to learn both spatial features and temporal dynamics effectively, making it suitable for complex tasks like activity recognition, image captioning, and video description [19]. Fig. 3 depicts the architecture of LRCN model for activity recognition.

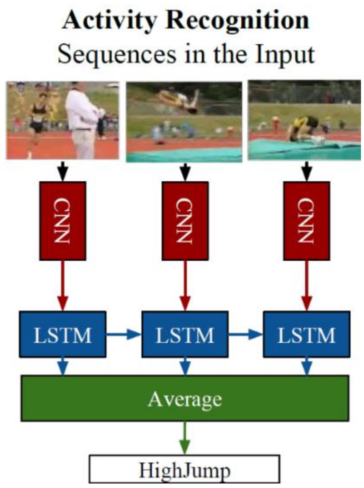

Figure 3 Structure of LRCN Model [19]

C. CNN-GRU

The authors [20] presented a novel CNN-GRU hybrid architecture, named ICGNet, designed for efficient human activity recognition (HAR) using data from wearable sensors. It utilizes a CNN block inspired by the Inception module for spatial feature extraction and a GRU block for capturing temporal dependencies in time-series data. This end-to-end model processes raw sensor data without manual feature engineering, aiming for high accuracy and low computational requirements, suitable for real-time embedded applications.

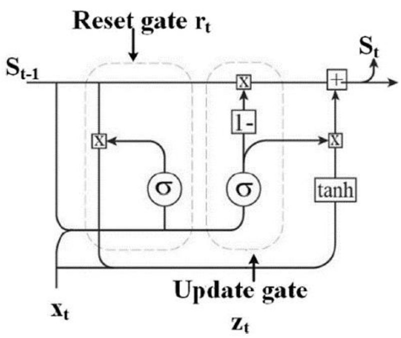

Figure 4 Structure of a GRU Cell [20]

GRUs are adept at addressing the challenges of exploding and vanishing gradients that are common in classic RNNs. The RNN section of the ICGNet architecture integrates two layered GRUs in succession. These GRU cells from Fig. 4 employ update and reset gates—mechanisms that control the flow and discard of information from one state to the next. The update gate in (1) is pivotal for carrying forward relevant data, while the reset gate decides what previous information to eliminate, enabling the GRU to retain critical past data for effectively capturing the temporal dynamics of sequential datasets.

$$z_t = \sigma(W_z x_t + U_z S_{t-1})$$
$$r_t = \sigma(W_r X_t + U_r S_{t-1}) \quad (1)$$

$$z_t = \sigma(W_z x_t + U_z S_{t-1})$$
$$r_t = \sigma(W_r X_t + U_r S_{t-1}) \quad (2)$$

In the GRU unit, x_t represents the current input and s_{t-1} the prior output. The update (z_t) and reset (r_t) gates regulate the information flow, with s_t denoting the current output and $\tilde{s}_t$ from (2), the potential output. Weight matrices W_z ,W_r,W_s,U_z, and U_r influence these outputs. The update gate z_t dictates the timing for refreshing s_t, while the reset gate r_t influences the calculation of $\tilde{s}_t$ by assessing the importance of s_{t-1} for the upcoming state.

4. Discussion and Analysis

A. Datasets Description

This study utilizes the Toyota Smarthome [21] dataset as well as CAUCAFall [22] dataset. Toyota Smarthome dataset features common daily activities of 18 senior subjects aged between 60 and 80. The CAUCAFall database features varied conditions such as occlusions, different lighting (natural, artificial, night), participant clothing diversity, background movement, various floor textures, different fall angles, and varying camera distances in uncontrolled home environments. Both datasets were reshaped into 64 x 64 pixels and train with a max sequence length of 20 to decrease the computational cost. The frames are selected evenly from sequence of images instead of random selection. Fig. 5-6 depict the visualization of activities with its labels.

The details of the activities are shown in Table I.

Table I Details of Activities used in Study

Class	Activities	Dataset
Normal Behaviour	Read book	Toyota Smarthome
	Take Pills	
	Cook	
	Walk	
	Make Coffee	
Abnormal Behaviour	Use stove -> Use telephone	
	Use stove -> Watch TV	
	Use stove -> Laydown	
	Fall	CAUCAFall

Normal activities that took place in living room and kitchen such as read book, take pills, cook, walk, and make coffee are selected. Since Toyota Smarthome dataset only consists of normal ADL and there is no public vision-based abnormal behaviour dataset of ADL available, therefore we came up with a way by merging some of the activities to create a sequence of abnormal activity. For instance, the video of using telephone is combined after the video of using stove to create a scene that the user does not turn off the stove and proceed to carry out other activities such as watch TV and laydown on sofa, which are considered as dangerous acts.

B. Experiments

The experiments were carried out in three parts, mainly (a) ConvLSTM, (b) LRCN and (c) CNN-GRU. The train test split of all the experiments is in 75/25 ratio. The experiments were conducted using Python 3.5 Keras on Google Colab with T4 GPU system using the code adapted from [23]. The Adam optimizer [24] is used for its effectiveness in handling varying gradients and adjusting learning rates, which is beneficial for complex neural network training. It combines the best aspects of other optimizers for consistent performance. A batch size of 4 is selected to balance computational efficiency and learning stability. Smaller

Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 6 (2023)

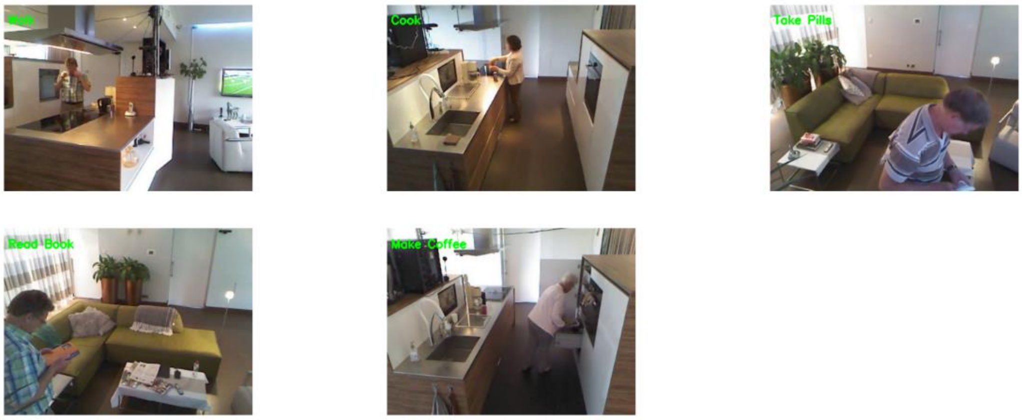

Figure 5 Normal activities selected

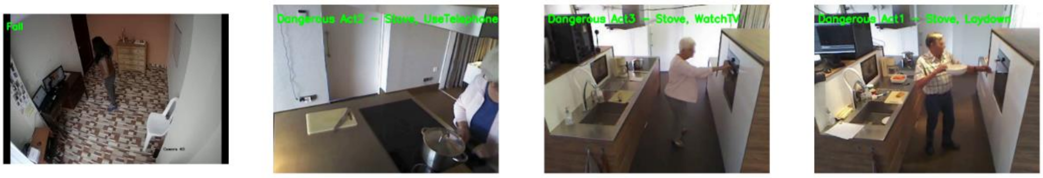

Figure 6 Abnormal activities selected

batch sizes allow more frequent updates, aiding quicker convergence, but can be volatile. A batch size of 4 offers an efficient update frequency while maintaining adequate training stability, especially when computing resources are limited.

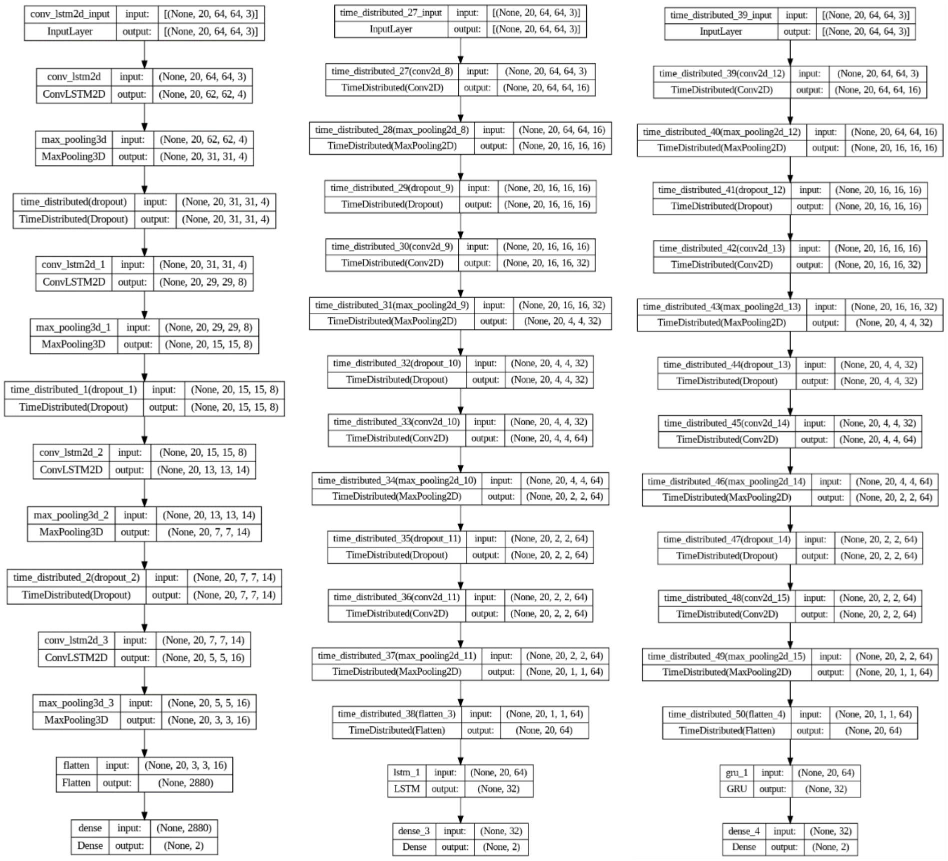

Figure 7 Structures of ConvLSTM (left), LRCN (mid) and CNN-GRU (right) deployed in this study.

ConvLSTM - To build the model, Keras *ConvLSTM2D* recurrent layers are employed. These layers not only process the data but also require specifications for the number of filters and kernel size, which are essential for performing the convolutional operations.

LRCN – The model merges CNN and LSTM layers within a unified model. The convolutional layers serve to extract spatial features from individual frames, and these features are then inputted into LSTM layer(s) at each timestep for modeling temporal sequences. Through this integrated process, the network is trained end-to-end to directly learn spatiotemporal features, yielding a more robust model. Additionally, the *TimeDistributed* wrapper layer will be utilized, enabling the consistent application of a singular layer to each frame of the video independently.

CNN-GRU – The model is similar to LRCN but the RNN is replaced by GRU.

Fig. 7 displays the structures of all three models deployed in this study.

5. Results and Discussion

The classification reports are displayed in Fig. 8-10 below. It is noted that class 0 represents abnormal behaviour and class 1 represents normal behaviour.

	precision	recall	f1-score	support
0	0.99	0.95	0.97	118
1	0.98	1.00	0.99	372
accuracy			0.99	490
macro avg	0.99	0.97	0.98	490
weighted avg	0.99	0.99	0.99	490

Figure 8 Classification Reports of ConvLSTM

	precision	recall	f1-score	support
0	0.97	0.84	0.90	118
1	0.95	0.99	0.97	372
accuracy			0.96	490
macro avg	0.96	0.92	0.94	490
weighted avg	0.96	0.96	0.95	490

Figure 9 Classification Reports of LRCN

	precision	recall	f1-score	support
0	0.96	0.66	0.78	118
1	0.90	0.99	0.94	372
accuracy			0.91	490
macro avg	0.93	0.83	0.86	490
weighted avg	0.92	0.91	0.91	490

Figure 10 Classification Reports of CNN-GRU

The ConvLSTM model is the standout performer with an exceptional precision of 0.99 for the abnormal behaviour class and an impressive recall of 0.95, coupled with a nearly perfect overall accuracy of 0.99. This model also exhibits a remarkable ability to identify the normal behaviour class, with a precision of 0.98 and recall of 1.00. In contrast, the CNN-GRU model, despite a high overall accuracy of 0.91, lags in recall for the abnormal behaviour class at 0.66, indicating a potential issue with false negatives. The LRCN model, while achieving a solid accuracy of 0.96, shows a lower precision for the normal behaviour class at 0.95 compared to ConvLSTM. These figures underline the ConvLSTM's superior capacity for correct classification, making it the most reliable, whereas the CNN-GRU, with its lower recall for the abnormal behaviour class, might be considered the least robust of the three.

Fig. 11-13 show the total loss vs total validation loss, and total accuracy vs total validation accuracy plots over epoch for three models.

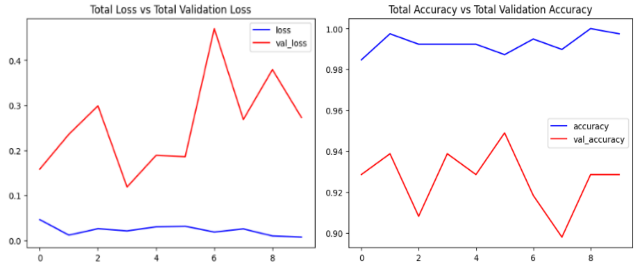

Figure 11 Training and Validation Loss Metrices (left) and Training and Validation Accuracy Metrices (right) of ConvLSTM

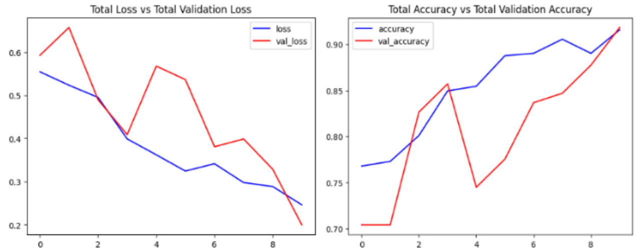

Figure 12 Training and Validation Loss Metrices (left) and Training and Validation Accuracy Metrices (right) of LRCN

Figure 13 Training and Validation Loss Metrices (left) and Training and Validation Accuracy Metrices (right) of CNN-GRU

All three models are improving in performance over epochs. The CNN-GRU model might be overfitting as indicated by the divergence between training and validation loss. The ConvLSTM model shows a promising balance between learning and generalization. The LRCN model exhibits a unique pattern where validation loss falls below training loss, which could indicate a particularly effective generalization. Fig. 14-16 show the confusion matrices for all the three models.

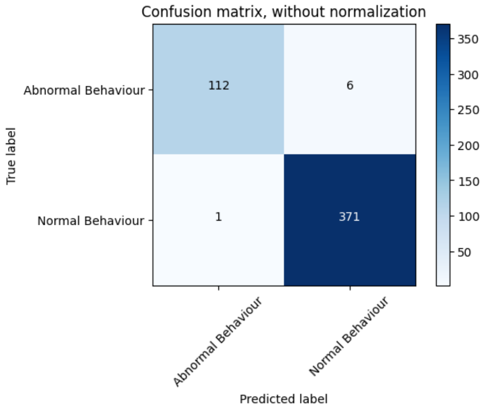

Figure 14 Confusion Matrix of ConvLSTM

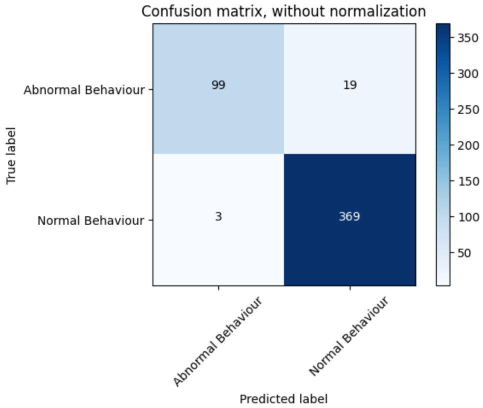

Figure 15 Confusion Matrix of LRCN

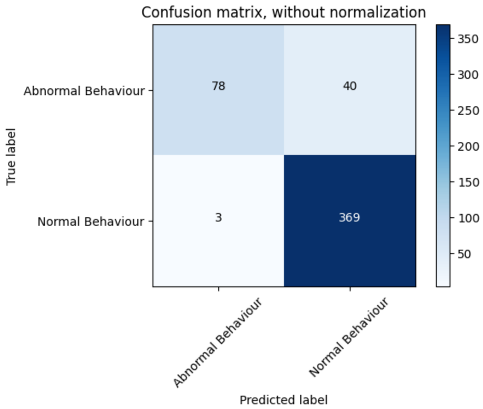

Figure 16 Confusion Matrix of CNN-GRU

ConvLSTM model emerges as the most accurate, exhibiting superior precision with the highest number of correctly identified abnormal behaviours (112 true positives) and the best normal behaviour recognition (371 true negatives), alongside the fewest misclassifications (1 false positive and 6 false negatives). In contrast, the CNN-GRU model, underperforms relative to the others, with a significant number of abnormal behaviours being misclassified as normal (40 false negatives), despite a low false positive rate (3). The LRCN model represents a middle ground, with fewer false negatives (19) than the first model but more than the second, indicating a higher sensitivity to abnormal behaviour than the CNN-GRU model but falling short of the ConvLSTM 's high accuracy.

To enhance our comprehension of the model, we delved into the performance metrics for specific actions like cooking, walking, and falling. This analysis will pinpoint areas needing refinement and help guide targeted improvements in the model's predictive accuracy.

The confusion matrix in Fig. 17 indicates that the model is proficient at correctly identifying activities like 'Walking', with a high number of true positives (78). However, it struggles with activities involving the stove ('Dangerous Act1' and 'Dangerous Act3'), often confusing them with each other and with 'UseTelephone'. The 'Fall' category is well-recognized with few misclassifications. Activities with lower instances, such as 'Make Coffee' and 'Read Book', see some confusion with each other and with 'Cook', suggesting a need to improve the model's ability to differentiate between similar activities.

The confusion matrix suggests that activities involving the kitchen counter, such as 'Use Stove, Laydown' and 'Make Coffee', are often conflated, potentially due to the shared environment. Similarly, 'Use Stove, WatchTV' is occasionally misidentified as 'Read Book', which may be attributed to the commonality of the living room setting for both activities. These observations underline the necessity for enhanced discriminative features within the model, especially in distinguishing activities occurring in identical settings, to improve the precision of ABD in ADL.

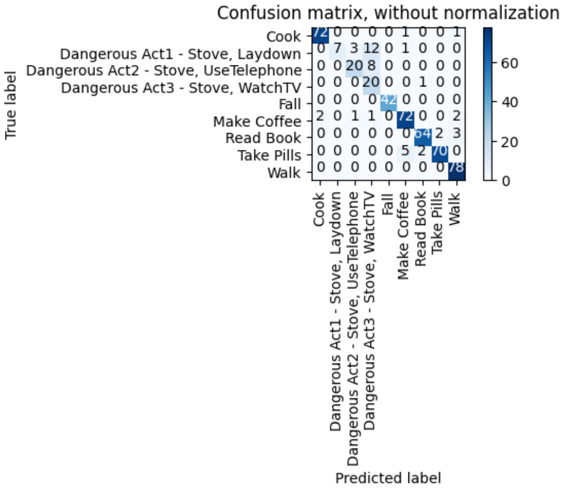

Figure 17 Confusion Matrix of All the Activities

6. Conclusion and Further Research

In this study, we have made significant strides in enhancing elderly safety in ADL, with a particular emphasis on detecting abnormal behaviours such as falls and dangerous acts like leaving the stove on but proceed with doing other activities. Among the models analyzed, ConvLSTM emerged as the most effective over LRCN and CNN-GRU, demonstrating superior performance in ABD. This finding is particularly relevant in settings like smart homes and healthcare facilities, where accurate and timely detection of abnormal behaviours is crucial for effective elderly care. The limited availability of datasets especially vision-based data for studying abnormal behaviour in elderly ADL poses a significant limitation to the research field. This scarcity of data can impact the model's ability to accurately learn and predict abnormal behaviours. Addressing this challenge, we mitigated this limitation by generating synthetic datasets through combination of normal behaviour videos.

Another critical area of research involves exploring the impact of similar settings on the misclassification of behaviours as abnormal or normal. The intricacies of different actions within similar environments can contribute to misclassifications, affecting the accuracy and reliability of behaviour detection models. Future studies should delve deeper into this aspect, examining how variations in actions within similar settings can be more accurately distinguished. This will involve not only refining existing models but also developing new techniques and algorithms that are more sensitive to subtle differences in behaviour patterns.

Refrences

[1] N. C. Tay, T. Connie, T. S. Ong, A. B. J. Teoh, and P. S. Teh, "A Review of Abnormal Behavior Detection in Activities of Daily Living," *IEEE Access*, vol. 11, pp. 5069–5088, 2023, doi: 10.1109/ACCESS.2023.3234974.

[2] Y. Wang *et al.*, "A Survey on Ambient Sensor-Based Abnormal Behaviour Detection for Elderly People in Healthcare," *Electronics*, vol. 12, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/electronics12071539.

[3] O. D. Lara and M. A. Labrador, "A Survey on Human Activity Recognition using Wearable Sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013, doi: 10.1109/SURV.2012.110112.00192.

[4] D. Arifoglu, H. N. Charif, and A. Bouchachia, "Detecting indicators of cognitive impairment via Graph Convolutional Networks," *Engineering Applications of Artificial Intelligence*, vol. 89, p. 103401, Mar. 2020, doi: 10.1016/j.engappai.2019.103401.

[5] D. Arifoglu and A. Bouchachia, "Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks," *Procedia Computer Science*, vol. 110, pp. 86–93, 2017, doi: 10.1016/j.procs.2017.06.121.

[6] D. Arifoglu and A. Bouchachia, "Detection of abnormal behaviour for dementia sufferers using Convolutional Neural Networks," *Artificial Intelligence in Medicine*, vol. 94, pp. 88–95, Mar. 2019, doi: 10.1016/j.artmed.2019.01.005.

[7] J. Lundström, E. Järpe, and A. Verikas, "Detecting and exploring deviating behaviour of smart home residents," *Expert Systems with Applications*, vol. 55, pp. 429–440, Aug. 2016, doi: 10.1016/j.eswa.2016.02.030.

[8] D. Arifoglu and A. Bouchachia, "Abnormal Behaviour Detection for Dementia Sufferers via Transfer Learning and Recursive Auto-Encoders," in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Mar. 2019, pp. 529–534. doi: 10.1109/PERCOMW.2019.8730744.

[9] Y. Fan, G. Wen, D. Li, S. Qiu, and M. D. Levine, "Early event detection based on dynamic images of surveillance videos," *Journal of Visual Communication and Image Representation*, vol. 51, pp. 70–75, Feb. 2018, doi: 10.1016/j.jvcir.2018.01.002.

[10] Q. Feng, C. Gao, L. Wang, Y. Zhao, T. Song, and Q. Li, "Spatio-temporal fall event detection in complex scenes using attention guided LSTM," *Pattern Recognition Letters*, vol. 130, pp. 242–249, Feb. 2020, doi: 10.1016/j.patrec.2018.08.031.

[11] F. Cardile, G. Iannizzotto, and F. L. Rosa, "A vision-based system for elderly patients monitoring," in *3rd International Conference on Human System Interaction*, May 2010, pp. 195–202. doi: 10.1109/HSI.2010.5514566.

[12] G. Sun and Z. Wang, "Fall detection algorithm for the elderly based on human posture estimation," in *2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, Apr. 2020, pp. 172–176. doi: 10.1109/IPEC49694.2020.9114962.

[13] V. A. Nguyen, T. H. Le, and T. T. Nguyen, "Single camera based fall detection using motion and human shape features," in *Proceedings of the Seventh Symposium on Information and Communication Technology*, in SoICT '16. New York, NY, USA: Association for Computing Machinery, Dec. 2016, pp. 339–344. doi: 10.1145/3011077.3011103.

[14] X. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, "Depth-Based Human Fall Detection via Shape Features and Improved Extreme Learning Machine," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1915–1922, Nov. 2014, doi: 10.1109/JBHI.2014.2304357.

[15] J. Zhang, C. Wu, and Y. Wang, "Human Fall Detection Based on Body Posture Spatio-Temporal Evolution," *Sensors (Basel)*, vol. 20, no. 3, p. 946, Feb. 2020, doi: 10.3390/s20030946.

[16] S. Chhetri, A. Alsadoon, T. A. D. in, P. W. C. Prasad, T. A. Rashid, and A. Maag, "Deep Learning for Vision-Based Fall Detection System: Enhanced Optical Dynamic Flow," *Computational Intelligence*, vol. 37, no. 1, pp. 578–595, Feb. 2021, doi: 10.1111/coin.12428.

[17] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. Woo, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting." arXiv, Jun. 12, 2015. doi: 10.48550/arXiv.1506.04214.

[18] A. Xavier, "An introduction to ConvLSTM," Neuronio. Accessed: Dec. 16, 2023. [Online]. Available: https://medium.com/neuronio/an-introduction-to-convlstm-55c9025563a7

[19] J. Donahue *et al.*, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description." arXiv, May 31, 2016. doi: 10.48550/arXiv.1411.4389.

[20] N. Dua, S. N. Singh, V. B. Semwal, and S. K. Challa, "Inception inspired CNN-GRU hybrid network for human activity recognition," *Multimed Tools Appl*, vol. 82, no. 4, pp. 5369–5403, Feb. 2023, doi: 10.1007/s11042-021-11885-x.

[21] "Toyota Smarthome: Real-World Activities of Daily Living." Accessed: Dec. 16, 2023. [Online]. Available: https://project.inria.fr/toyotasmarthome/

[22] J. C. Eraso, E. Muñoz, M. Muñoz, and J. Pinto, "Dataset CAUCAFall," vol. 4, Jul. 2022, doi: 10.17632/7w7fccy7ky.4.

[23] T. Anwar, R. Naeem, and M. Anjum, "Human Activity Recognition using TensorFlow (CNN + LSTM) | Bleed AI." Accessed: Dec. 17, 2023. [Online]. Available: https://bleedaiacademy.com/human-activity-recognition-using-tensorflow-cnn-lstm/

[24] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. doi: 10.48550/arXiv.1412.6980.