Video-Based Inpatient Fall Risk Assessment: A Case Study

Ziqing Wang[1,2], Mohammad Ali Armin[1], Simon Denman[3], Lars Petersson[1], David Ahmedt-Aristizabal[1,3]

Abstract—Inpatient falls are a serious safety issue in hospitals and healthcare facilities. Recent advances in video analytics for patient monitoring provide a non-intrusive avenue to reduce this risk through continuous activity monitoring. However, in-bed fall risk assessment systems have received less attention in the literature. The majority of prior studies have focused on fall event detection, and do not consider the circumstances that may indicate an imminent inpatient fall. Here, we propose a video-based system that can monitor the risk of a patient falling, and alert staff of unsafe behaviour to help prevent falls before they occur. We propose an approach that leverages recent advances in human localisation and skeleton pose estimation to extract spatial features from video frames recorded in a simulated environment. We demonstrate that body positions can be effectively recognised and provide useful evidence for fall risk assessment. This work highlights the benefits of video-based models for analysing behaviours of interest, and demonstrates how such a system could enable sufficient lead time for healthcare professionals to respond and address patient needs, which is necessary for the development of fall intervention programs.

I. INTRODUCTION

Falls in the ward, in particular those from the bed, are a persistent problem and are commonly associated with injuries such as soreness and bone fractures and often result in a prolonged hospital stay. In mental health hospitals and some psychogeriatric units, these events are of particular concern due to patient cognitive impairment, dizziness or vertigo [1], [2]. Such incidents are one of the main concerns for all staff involved in the care of patients, and can lead to anxiety or guilt, and potentially litigation. In most hospitals, medical staff follow well-defined protocols to prevent falls, however research into systems capable of generating immediate alerts to enable medical assists to prevent falls has received limited attention from researchers.

Considering the importance of patient behaviour monitoring, several in-clinic patient monitoring systems using computer vision and deep learning have been introduced to provide an objective assessment of a patient's behaviour. These vision-based systems have attracted great attention due to their non-invasive nature and have shown promising results in analysing patient-specific pose [3] (for example, sleeping pose [4], [5]), epileptic patients [6], breathing disorders [7] and infant motions [8]. Furthermore, camera-based fall detection [9]–[11] and fall prediction systems [12], [13], which detect when a fall occurs rather than seeking to predict a fall

[1] CSIRO, DATA61, Canberra, Australia. Corresponding author: `david.ahmedtaristizabal@data61.csiro.au`
[2] Australian National University, Canberra, Australia
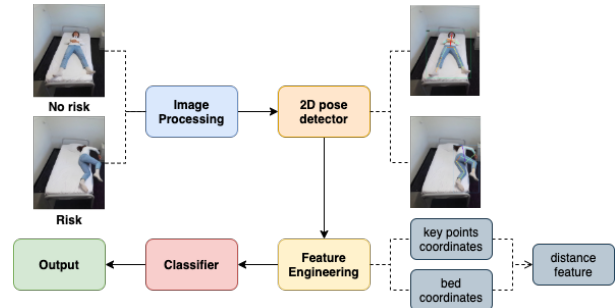[3] SAIVT Research lab, Queensland University of Technology, Brisbane, Australia.

Fig. 1. Overview of the proposed fall risk monitoring system. Given a single RGB image collected with a custom camera system placed in the ceiling, the system generates 2D human pose predictions. Next, the relative position of the human and the bed is computed to predict the risk of falling.

before it happens, have also received considerable attention recently and have achieved effective results using existing simulated datasets [14]–[16], or synthetic libraries [11], [17]. However, while human pose estimation has become the de-facto standard for inpatient analysis, its application to the prevention of falls from the bed remains limited.

To prevent falls, some systems detect the position in which the patient is lying with respect to the edge of the bed, or detect the patient's bed-exit behaviour. In these scenarios, the system monitors a key human pose or human motion to predict the risk of falling. These studies have used commercial pressure mat systems to detect the human off-bed position [18]; however, pressure pads have been shown to generate a high volume of false alarms leading to alarm fatigue [19]. On the other hand, camera-based systems seek to detect a sitting posture [20] or the bed exit action from a sequence of human images [21]. Although, many hospital patients fall as they get out of bed, there are risk factors for falling such as uncontrolled motions caused by agitations, restless sleep and abnormal dreams that lead to a patient trying to climb out of the bed for protection, or "jumping" from the bed. Patients attempting to perform these unassisted activities account for a large proportion of inpatient falls, and they are the focus of this paper.

In this paper, we explore the feasibility of adapting pose-based frameworks to identify patients' behaviour and assess the risk of falling from the bed.

Our main contributions are summarized as follows:

1) We introduce a flexible vision-based fall risk detection system capable of detecting actions in a novel simulated environment that may indicate an imminent inpatient fall.
2) We propose a robust but simple non-obtrusive monitoring system to capture relative body position information to assess the risk of falling from a bed.

Fig. 2. Human and bed localisation. A dynamic and fast instance segmentation approach is used to localise the region of interest.

Fig. 3. Selected samples of human actions collected in the simulated dataset. Top: Not at risk of falling. Bottom: At risk of falling.

II. MATERIALS AND METHODS

In this paper, we propose a non-invasive landmark-based approach to capture in-bed human pose, and predict whether the human in an image will fall from the bed they are currently occupying.

A. Fall prevention simulated dataset and pre-processing

Existing vision-based fall detection datasets do not cover inpatient fall events, or examples of patients at risk of falling. Considering this limitation, we designed and collected a simulated dataset with the actions of interest represented by two classes, *not at risk* and *at risk* of falling. To generate this data two main stages were covered: i) recruitment, experimental setup, and data collection, and ii) data annotation and pre-processing.

We collect data from participants in a simulated hospital environment. Volunteers lie down on the bed and simulate in-bed patient actions such as trying to climb out of the bed for protection, turning around, exiting the bed, and falling. All images and videos are collected with a custom recording device equipped with a Microsoft Azure Kinect camera, and are saved for a further prepossessing phase.

To estimate relative positions of the human and bed, we first define the region of interest as the location of the bed that contains the participant. This process helps to deal with different camera-bed viewing angles, and changes in the inclination angle of the bed which may impact the risk assessment. We perform object boundary detection using SOLO [22], which is a dynamic and fast state-of-the-art instance segmentation method. In our network, we use pre-trained weights, trained on the COCO dataset [23], to detect the human and the bed. Then, we crop and resize all images
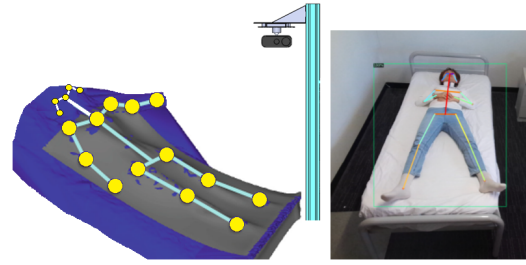

Fig. 4. Representation of the human key points detected in a selected image.

to a resolution of 1080×828 pixels as input to the system. Fig. 2 depicts selected examples of the participant and bed detection. Finally, each image is separated into two classes to conduct the proposed analysis: *at risk* and *not at risk* of falling from the bed, as illustrated in Fig. 3. We address the data imbalance in our dataset by adopting data augmentation techniques including oversampling and adding Gaussian noise to images to obtain the same number of samples for both classes. We argue that this pre-processing step does not impact generalization.

B. Fall risk assessment system

Fig. 1 shows the overall architecture of our framework which has three main modules: **i**) 2D key-point estimation, which takes an RGB image and produces body joint locations in 2D space; **ii**) A human-bed relative position estimation and feature engineering; and **iii**) a fall risk classifier which combines human pose and relative position features to accurately discriminate between *at risk* and *not at risk* cases. In the following, we describe in detail the individual components of our framework.

1) Human localisation and pose identification: Quantifying a person's posture and limb articulation is useful for understanding patient behaviour. Human pose estimation from static images has shown strong performance in detecting positions of interest for the analysis of seizure disorders [6] and bed-exit posture [20].

We aim to employ a robust 2D pose prediction technique to extract consistent poses from a hospital environment, where challenges such as self-occlusion and similarities between the background and foreground are present. We adopt the Mask-RCNN architecture [24] to predict 2D locations of body joints and their corresponding confidence scores, which is a lightweight, yet highly effective approach implemented in Detectron2 [25]. Here, a keypoint location is modeled as a one-hot mask where Mask-RCNN predicts k masks, one for each of K keypoints (17 key points coordinates in this study). We fine tune a pre-trained Mask-RCNN model trained on the COCO dataset [23], enabling the model to detect keypoints clearly in our dataset. When participants perform activities such as turning around, their body parts may overlap and some key points cannot be detected clearly. Based on an analysis of the pose estimation results, we define the most stable coordinates for a later feature engineering step. Fig. 4 illustrates the keypoint layout and a detected human pose.

2) Relative position determination and feature engineering: Human body postures during events such as the patient
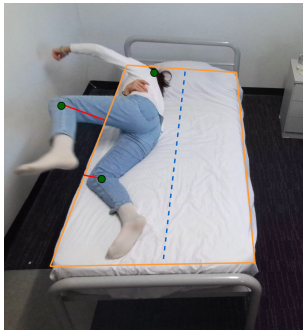
Fig. 5. Representation of the distance feature estimation. Head and knees detected by the pose estimation algorithm are marked as green dots, and all lines represent the contour of the detected bed.

TABLE I
MULTI-FOLD CROSS-VALIDATION PERFORMANCE (10-TIME AVERAGE)

Feature representation	Test Accuracy (%)	
	Light GBM	SVM
$Dis < Knee, Bed >$	94.44	91.11
$Dis < Knee, Bed > + Dis < Head, Bed >$	96.11	92.89
$17 Keypoints + Dis < Knee, Bed >$	96.67	93.89
$17 Keypoints + Dis < Knee, Bed > + Dis < Head, Bed >$	**97.22**	94.20

Fig. 6. Qualitative results of human pose estimation on patients *not at risk* and *at risk* of falling (trying to climb out of the bed for protection).

climbing out of the bed are complex and varied, so it is essential to define a criteria to classify whether a patient is at risk of falling from the bed or not. One criteria is to track and estimate how much of the knee is outside of the bed.

To calculate the distance between the knee and the bed, we need to determine which side of the bed a patient is most likely to fall. As shown in Fig. 5, the left line, middle line and right line of the bed define the bed position, and the head and knees are marked as green dots. When the head and the two knees are all on the left of the middle line, the human body is defined as being on the left side of the bed. When the head and two knees are all on the right of middle line, the human body is defined as being on the right side of the bed. If the human body is on the left side of the bed, we will calculate the distance between the two knees and the left line of the bed. If the human body is on the right side of the bed, we will calculate the distance between the two knees and the right line of the bed. Defining where the human body is located can also help to remove unnecessary features and decrease feature dimensionality.

The distance feature is the distance between the knees and the bed boundary. Based on the human body's location on the bed, we can determine which boundary should be used to calculate the distance between the knee and the bed. If the knee is outside of the bed, the distance value is negative and if the knee is inside the bed, the distance value is positive.

3) Fall risk classification: Each output feature related to the human pose coordinates and relative position between the

body and the bed (i.e. distance features) is fed to a classifier to learn probabilistic distributions with respect to the target class. We adopt the LightGBM classifier [26] as a fast, distributed, high performance implementation of gradient boosted trees for supervised classification with robustness to overfitting. The following setting is used in the experiment: boosting type (gradient boosting decision tree), boosting learning rate (0.1), number of boosted trees to fit (100), maximum tree leaves (31), maximum tree depth (no limit). We also use a traditional support vector machine (SVM) classifier with automatic Bayesian optimization: kernel (sigmoid), shrinking (true), cache size (200).

III. EVALUATION

A. Experimental setup

To evaluate and compare the most discriminative features from the landmark-based analysis, we adopt four feature sets: **i)** the distance between knees and the bed; **ii)** set (i) plus the distance between the head and the bed; **iii)** set (i) plus 17 body keypoint coordinates; and **iv)** set (iii) plus the distance between the head and the bed. These feature sets are listed in Table I.

As an ablation study, we investigate a region-based approach by training a ResNet50 [27] architecture to extract a spatial representation directly from images, and perform classification using a fully connected layer with a sigmoid activation function. This model was implemented in Keras [28] and trained by optimizing the categorical cross-entropy loss with the Adam optimizer [29].

Both models (landmark-based and region-based), are assessed through a 10-fold cross validation to ensure that the training and test data is disjoint. For each fold, the data sample of each class is randomly split into 90% for training and 10% for testing.

B. Experimental results and discussion

While general pose estimation frameworks are effective when subjects are located in uncluttered settings, they can be unreliable when applied to noisy environments such as patient monitoring rooms. This can cause even more confusion when multiple frames are fed into the framework, thus in our study we consider a static scene. From the pose estimation results on the dataset, we confirm that knee points are the most stable keypoints, so we choose to use the knee as a reference to classify *at risk* events and use the head location to identify the human position on the bed. Selected samples showing estimated joint locations are presented in Fig. 6.

The region-based approach achieved an average accuracy of 65.8% on the test set, whereas the landmark-based approach using LightGBM achieved 97.22% accuracy. The cross-validation performance for each set of features and the proposed classifiers are shown in Table I. Our landmark-based results indicate that larger feature sets can improve system performance. The best accuracy is obtained from the fourth feature set using 17 keypoint coordinates, the distance between the knee and the bed and the distance

between the head and the bed. This performance is likely due to the distance features indicating whether key body parts overlap the boundary of the bed, which indicates a high probability of falling from the bed. This relative position information is hard to capture without feature engineering and is the main reason that the region-based approach shows low performance.

In most hospitals, there are programs and policies for fall prevention, but there is limited research into systems capable of generating immediate alerts for medical assistance to prevent falls. A video-based alarm system for fall prevention similar to our proposed framework needs the ability to detect highly accurate relative human position in the bed. Such a system is able to issue an alert as early as possible once it detects a position where there is a high risk of falling. Further, it is envisioned that our approach is cost effective and low maintenance.

The majority of inpatient fall studies focus solely on fall risk factors but may not identify potential causal factors for falls (e.g. what triggered the fall) which is necessary for fall intervention programs. An interesting direction for future research is the creation of libraries of behaviours to identify these factors and patients at high risk of falling in the early stages of monitoring.

IV. CONCLUSIONS

In this paper, we introduce a vision based monitoring system that incorporates state-of-the-art computer vision techniques to assess the risk of falling from a bed. Considering the lack of datasets to assess fall risk and fall prevention, we introduce a simulated dataset that includes in-bed human actions such as trying to climb out of the bed for protection, turning around, and bed-exit events. Our results in this particular case study show a promising technology that can have a positive impact on monitoring inpatients at risk of falling. Our proposed system has a high accuracy, resulting in lower false alarm rates for medical staff and thus a reduction in the likelihood of alarm fatigue.

Ethics statement: The experimental procedures involving human subjects described in this paper were approved by the CSIRO Health and Medical Human Research Ethics Committee (CHMHREC).

REFERENCES

[1] D. Oliver, F. Healey, and T. P. Haines, "Preventing falls and fall-related injuries in hospitals," *Clinics in geriatric medicine*, vol. 26, no. 4, pp. 645–692, 2010.

[2] M. Vassallo, T. Azeem, M. Pirwani, J. C. Sharma, and S. C. Allen, "An epidemiological study of falls on integrated general medical wards." *International Journal of Clinical Practice*, vol. 54, no. 10, pp. 654–657, 2000.

[3] K. Chen, P. Gabriel, A. Alasfour, C. Gong, W. K. Doyle, O. Devinsky, D. Friedman, P. Dugan, L. Melloni, T. Thesen *et al.*, "Patient-specific pose estimation in clinical environments," *IEEE journal of translational engineering in health and medicine*, vol. 6, pp. 1–11, 2018.

[4] S. Liu and S. Ostadabbas, "Seeing under the cover: A physics guided learning approach for in-bed pose estimation," in *MICCAI*, 2019, pp. 236–245.

[5] S. Liu, Y. Yin, and S. Ostadabbas, "In-bed pose estimation: Deep learning with shallow dataset," *IEEE journal of translational engineering in health and medicine*, vol. 7, pp. 1–12, 2019.

[6] D. Ahmedt-Aristizabal, S. Denman, K. Nguyen, S. Sridharan, S. Dionisio, and C. Fookes, "Understanding patients' behavior: Vision-based analysis of seizure disorders," *IEEE journal of biomedical and health informatics*, vol. 23, no. 6, pp. 2583–2591, 2019.

[7] M. Martinez, D. Ahmedt-Aristizabal, T. Väth, C. Fookes, A. Benz, and R. Stiefelhagen, "A vision-based system for breathing disorder identification: A deep learning perspective," in *EMBC*, 2019, pp. 6529–6532.

[8] N. Hesse, S. Pujades, M. Black, M. Arens, U. Hofmann, and S. Schroeder, "Learning and tracking the 3d body shape of freely moving infants from rgb-d sequences," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[9] N. Lu, Y. Wu, L. Feng, and J. Song, "Deep learning for fall detection: Three-dimensional cnn combined with lstm on video kinematic data," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 314–323, 2018.

[10] Y. Chen, W. Li, L. Wang, J. Hu, and M. Ye, "Vision-based fall event detection in complex background using attention guided bi-directional lstm," *IEEE Access*, vol. 8, pp. 161 337–161 348, 2020.

[11] U. Asif, B. Mashford, S. Von Cavallar, S. Yohanandan, S. Roy, J. Tang, and S. Harrer, "Privacy preserving human fall detection using video data," in *Machine Learning for Health Workshop*, 2020, pp. 39–51.

[12] M. Hua, Y. Nan, and S. Lian, "Falls prediction based on body keypoints and seq2seq architecture," in *ICCV Workshops*, 2019.

[13] A. Masalha, N. Eichler, S. Raz, A. Toledano-Shubi, D. Niv, I. Shimshoni, and H. Hel-Or, "Predicting fall probability based on a validated balance scale," in *CVPR Workshops*, 2020, pp. 302–303.

[14] E. Auvinet, C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Multiple cameras fall dataset," *DIRO-Université de Montréal, Tech. Rep*, vol. 1350, 2010.

[15] I. Charfi, J. Miteran, J. Dubois, M. Atri, and R. Tourki, "Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and adaboost-based classification," *Journal of Electronic Imaging*, vol. 22, no. 4, p. 041106, 2013.

[16] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," *Computer methods and programs in biomedicine*, vol. 117, no. 3, pp. 489–501, 2014.

[17] U. Asif, S. Von Cavallar, J. Tang, and S. Harre, "SSHFD: Single shot human fall detection with occluded joints resilience," *arXiv preprint arXiv:2004.00797*, 2020.

[18] W. Viriyavit and V. Sornlertlamvanich, "Bed position classification by a neural network and bayesian network using noninvasive sensors for fall prevention," *Journal of Sensors*, vol. 2020, 2020.

[19] R. I. Shorr, A. M. Chandler, L. C. Mion, T. M. Waters, M. Liu, M. J. Daniels, L. A. Kessler, and S. T. Miller, "Effects of an intervention to increase bed alarm use to prevent falls in hospitalized patients: a cluster randomized trial," *Annals of internal medicine*, vol. 157, no. 10, pp. 692–699, 2012.

[20] M. Inoue, R. Taguchi, and T. Umezaki, "Bed-exit prediction applying neural network combining bed position detection and patient posture estimation," in *EMBC*, 2019, pp. 3208–3211.

[21] M. Inoue and R. Taguchi, "Bed exit action detection based on patient posture with long short-term memory," in *EMBC*, 2020, pp. 4390–4393.

[22] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "SOLOv2: Dynamic and fast instance segmentation," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[23] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*, 2014, pp. 740–755.

[24] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *ICCV*, 2017, pp. 2961–2969.

[25] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[26] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in neural information processing systems*, 2017, pp. 3146–3154.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[28] F. Chollet *et al.*, "Keras," 2017.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.