An Improved Fall Detection System that uses Convolutional Neural Networks and Optical Flows

Denise Beh Sin Jie[1], Lee Zheng Yao Daniel[2], Mah Shian Yew Brendan[3], Poh Yu Jie[4], Teo Ming Jun[5]

National University of Singapore

A0207982W[1], A0223259E[3], A0216349B[3], A0216055M[4], A0216305R[5]

Abstract

With the aging population problem that Singapore is facing, projections show that the elderly population may be as big as a quarter of the population. Such a phenomenon puts a heavy burden on the Singapore healthcare system as well as the other parts of the community. More thus needs to be done to support the elderly in Singapore.

With the prevalence of security video cameras in everyday life, detection through video feed has been made very possible. As the consequences of an elderly falling down can be long lasting and devastating, this study seeks to present a machine learning model that utilizes Convolutional Neural Networks for a vision based fall detection model. This model will also seek to alert caregivers promptly after the model detects a fall.

Introduction

Singapore is currently met with an unprecedented demographic shift towards an aging society, with the proportion of her population aged 65 years and above being projected to increase notably to 25% by 2030 (Soh et al. 2021). Yet with the growing prevalence of singlehood and smaller families in the country, it is no longer feasible to rely on children as caregivers (Heng 2021), leaving many elderly today vulnerable as they lack the support and care they need.

Thus we see an increase in dependency on alternate modes of care for the elderly. The nation's eldercare ecosystem has adapted over the years to address this demand — expanding the network of care institutions, improving standards of nursing homes and increasing capacity of purpose-built homes (Lin 2021). In addition, we see an unfortunate increase in the number of single elderly households, with an estimated 83,000 elderly persons to be living alone by 2030 (Ng and Fadhilah 2018).

It is reported that one-third of community-dwelling elderly aged 65 years and above will suffer at least one fall in a year (Ang et al. 2020). In a situation of a fall, their frailty coupled with other possible disabilities may make it difficult for them to seek timely assistance. While some falls result in minor consequences, left unattended, more serious cases may lead to irreversible consequences such as trauma, debilitating injuries and even death (Al-Aama 2011).

Despite the exponential demand for elderly care services, there is a clear shortage of healthcare professionals in Singapore (HRM Asia 2018). This is exacerbated by the recent Covid-19 pandemic which witnessed a surge in resignation from caregivers and healthcare workers. There is a pressing need to find new ways to lighten the load of nurses and other caregivers (Teo, 2021). Towards this, our group see an opportunity where the incorporation of automation in monitoring processes of the elderly may be beneficial. Hence we aim to develop an accurate fall detection model, which when given a video stream as input, can identify potential fall events and notify the relevant parties for assistance as soon as possible. This could then be utilized in various settings such as nursing homes, hospitals or even in homes.

Current Methods and Their Limitations

Several fall detection models have been put forth by the research community and commercial entities, and they can be largely categorized into three streams: ambience devices, wearable sensors, and camera or vision-based approaches.

The ambience device approach commonly sees sensors being utilized and installed at locations near the person at risk, such as the beds or floors of their homes (Tanwar et al. 2022). While this approach does provide a more convenient and non-intrusive means of fall detection, they also produce significantly higher false positive rates (Wang, Ellul and Azzopardi 2020). Its cost may accumulate due to the multiple installations around the homes of the elderly.

Singapore is also not unfamiliar with employing wearable devices to complement its campaigns, evident from its past efforts in Covid-19 tracing using TraceTogether Tokens (Asher 2020). The relatively good performance and affordability of wearable devices make them a very probable

approach (Wang, Ellul and Azzopardi 2020). However, it introduces an overhead to its users as it usually requires charging and even the conscious effort to remember to put it on daily. Though these can be taken care of by caregivers, it may not be well received or utilized by elderly persons who live alone.

On the other hand, vision-based approaches often eliminate the overhead and cost issues of previous methods, and have been shown to produce some significant results in the detection of falls (Chhetri et al. 2020). They still face the challenges of privacy concerns and complex environment conditions like dynamic lighting (Gutiérrez, Rodríguez and Martin 2021), but nevertheless, considering the benefits and the already prevalent usage of security cameras in Singapore, there were clear motivations for further research in this area.

Aims and Potential Applications

Thus, our group aims to develop a vision-based fall detection monitoring system which takes in video stream input from security cameras, and will alert relevant parties for assistance as soon as a potential fall is detected.

We see this being applicable in settings like nursing homes, elderly care centres, hospitals and homes. Already, Singapore has plans to build seven more nursing homes in estates such as Tampines, Punggol by 2023 as an effort to beef up the nation's eldercare support system (Ng 2021). Such a monitoring system can value add to such projects especially with the thinning of caregivers and healthcare personnel. In addition, Singapore will also be launching 200 assisted living HDB flats for seniors in Queenstown this year (Lin 2022). These flats offer senior-friendly designs and certain subscriptions to elderly care services, with their main target being seniors aged 65 years and above who desire to age independently but receive care if needed. The developed monitoring system would then be highly applicable in such a project as it ensures round-the-clock monitoring and access to timely assistance for the residents without being intrusive and with minimal additional cost to implement.

Vision-Based Fall Detection with Convolutional Neural Networks (CNN) by Núñez et al. (2017)

While privacy concerns have led to the rising popularity of depth cameras (Wang, Ellul and Azzopardi 2020), we chose to focus on related work using RGB inputs as this will facilitate future integration with existing security systems.

Our group looks to improve on the vision-based fall detection model presented by Núñez et al. (2017) which uses a CNN trained on optical flow images, and also pioneers the use of transfer learning from the action recognition domain to fall detection. Figure 1 shows the system architecture of the model. Stacks of optical flow images generated from RGB images of video streams are fed into a VGG16 CNN for feature extraction before being classified by a fully connected neural network (FC-NN) classifier.

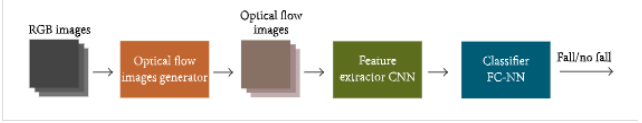

Figure 1: Full pipeline of Núñez et al.'s vision-based model

Their classifier model was trained using 3 datasets — UR Falls Dataset (URFD), Multiple Cameras Fall Dataset (Multicam) and Fall Detection Dataset (FDD).

In order to make the system more scenario-independent and also to model video motions, optical flow images were utilized as input to the neural networks for training. Optical flow is the perceived changes in motion of elements between consecutive frames of a video, modelled by the relative alterations of distance and angles between the camera and the recorded object. The use of optical flow as a means of foreground segmentation thus helps reduce the potential effects of environmental factors. The images are also fed to the feature extractor CNN in stacks, so as to represent the whole pattern of motion of a fall (Figure 2).

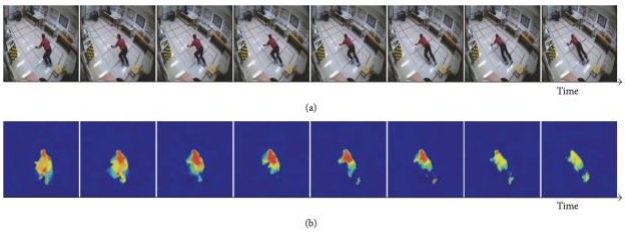

Figure 2: Sample of fall data (a) and their resultant optical flow horizontal displacement images (b), from Núñez et al.

Currently, a limitation of public datasets would be the low number of fall samples. Núñez et al. thus utilized transfer learning, taking advantage of the ability of CNNs to be sequentially trained on multiple datasets. They trained the CNN on the ImageNet dataset comprising 14 million images and 1,000 classes, and further trained it on the UCF101 dataset comprising 13,320 videos and 101 human actions. Fed to the network during training were the stacks of $R^{224x224x2L}$ optical flow images of the datasets. For each input stack, features extracted were saved in arrays $F \in R$ of size 4096 and these were used to fine-tune the last 2 fully connected. Figure 3 summarizes this training phase.

A problem they faced during classification was the poor ability in classifying the "fall" class. To counter this, the loss function was modified to enhance the importance of the fall class. Though this creates a model bias towards the fall

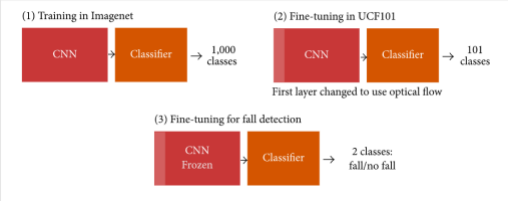

Figure 3: Núñez et al.'s (2017) use of transfer learning

class. Though this creates a model bias towards the positive classification of falls, this is acceptable in view of its use in an elderly fall detection model, where a false negative would be much more detrimental than a false positive.

They chose the binary cross-entropy loss function (1), where p is the prediction and t is the ground truth, and added class weights to the loss function (2).

$$loss(p,t) = -(t \cdot log(p) + (1-t) \cdot log(1-p)) \quad (1)$$

$$loss(p,t) = -(w_1 \cdot t \cdot log(p) + w_0 \cdot (1-t) \cdot log(1-p)) \quad (2)$$

w_0, the class weight for the "fall" class, was adjusted greater than 1.0 to increase the penalty on a wrong "fall" classification. The modified loss function was then minimized using backward propagation.

Methods

Our group built upon Núñez et al.'s (2017) vision-based fall detection architecture, feeding a CNN with optical flows followed by the three-step training process (Figure 3). We also focused our training and evaluation using the URFD dataset (Kwolek and Kepski 2014).

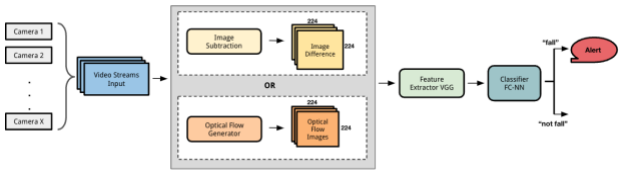

Figure 4: Full pipeline of our fall detection monitoring system built upon Núñez et al.'s (2017) model, including a comparison of an image subtraction method against their optical flow approach

In addition, due to the great computational cost of generating optical flows from images, we also decided to compare against an image subtraction approach as another means of detecting changes between frames. Our final model is then embedded in a monitoring system that is able to take in video streams from multiple cameras, sending an alert when a fall is detected. Our full code implementation can be found here.

Data Augmentation

The URFD dataset is made up of 70 videos - 30 falls and 40 activities of daily living (ADL) sequences. The small size of the dataset means that we run the risk of either underfitting our model where the model is unable to capture the true pattern in the dataset, or overfitting our model where the model has better predictive performance on the training data than unobserved test data, lacking the ability to generalize to real world situations.

Furthermore, the videos in the URFD dataset were filmed in static and stable lighting conditions. Specifically, the dataset did not sufficiently include videos that were of darker environments. This is problematic, since night time falls may go unnoticed due to poor classification. Other factors such as the view from which the videos were recorded were all similar throughout the dataset too. The lack of diversity and limited size of the dataset prompted us to implement data augmentation to increase the variation of images fed to the CNN, allowing our model to be more robust.

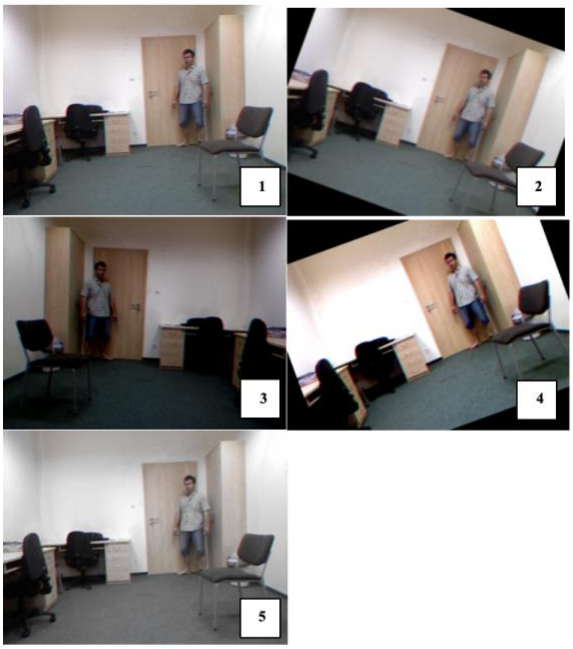

Figure 5: Image 1 shows the original frame; Image 2 depicts right rotation of 20° with box blur; Image 3 depicts horizontal flipping with decreased brightness; Image 4 depicts left right rotation of 20° with increased contrast; Image 5 depicts negative saturation

We utilized Python's openCV library (code), augmenting the images by adjusting the brightness, rotation and saturation of the images. Specifically, we made 4 changes to each original frame (Figure 5).

We ensured that the augmentation was not too far off from realistic scenarios so that our model would still be able to function properly in real life scenarios.

Through our augmentation of the images, we managed to expand the dataset tremendously by 5 folds. This may help the model to avoid overfitting and may allow it to become more versatile in different environments, which is especially necessary given the multiple possible settings we envision our system to be implemented in.

Adjustment of Training Epochs

Núñez et al. (2017) had used 3000 training epochs to train their classifier model. However, upon inspection, we see that changes in model accuracy and model loss are minimal way before the 3000th epoch (Figure 6).

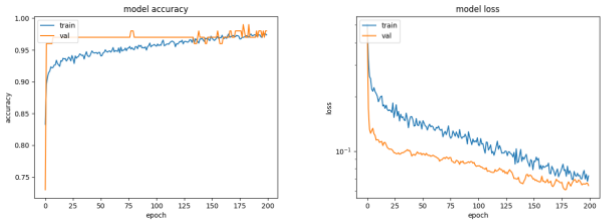

Figure 6: Model accuracy against number of training epochs (left), model loss against number of training epochs (right)

Too many training epochs would not only cause our model to overfit the training data, but would also unnecessarily take up a greater amount of time and computational power to train. From Figure 6, we see that the accuracy of the model does not change much beyond just 10 epochs. Thus considering this and the model loss, our group reduced the number of training epochs to 75, a number which we were relatively confident would not compromise on the performance of the model.

Adjustment of Learning Rate

As the number of training epochs was decreased drastically, we wanted to find out if the learning rate could be adjusted and increased accordingly. A learning rate too small may prolong the training process due to minute weight updates in each epoch, and may also cause the process to get stuck.

Since Núñez et al. (2017) had initially chosen a learning rate of 0.0001, we decided to compare the model performance against increased learning rates of 0.001 and 0.01, using a validation set comprising 20% of the whole dataset.

From Figure 7, we observe that increasing the learning rate to 0.001 saw an improvement in the model accuracy on both the training and validation set, with the accuracy plateauing at a higher accuracy than when the learning rate was at 0.0001. Thus, the model was able to learn faster with its performance not compromised, but enhanced.

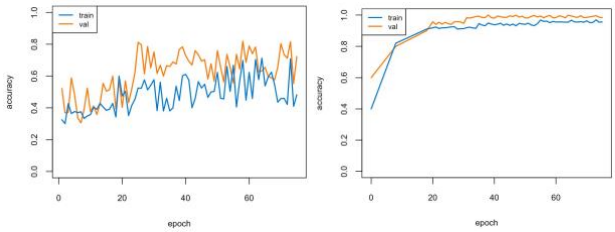

Figure 7: Model accuracy against number of training epochs for learning rate = 0.01 (left), model accuracy against number of training epochs for learning rate = 0.001 (right), both y-axes scaled from 0.0 to 1.0 for comparison

On the other hand, further increase of the learning rate to 0.01 caused the accuracy and precision of the model on the validation set to decrease significantly (Figure 7). Furthermore, after each epoch, the accuracy of the model had great fluctuations and did not seem to converge. This could be a result of setting the learning rate too high, which can often lead to unstable training processes and divergence from or overshooting of the optical minima. Thus, or group decided to raise the learning rate to 0.001.

Fall Alert System

For fall cases in the elderly, prompt responses are necessary and often can determine the seriousness of their injuries. As such, caregivers should be prompted swiftly about any potential falls so that they may attend to the elderly as soon as possible. Thus we developed a full monitoring software (code) that integrates and employs the fall detection model in the monitoring process, sending alerts when a potential fall is identified.

Our software contains two main modules: Camera module and Monitor module. It leverages on Python's multiprocessing library to monitor for falls over multiple video input streams from their respective security cameras. Each camera is launched as an instance of a Camera process, and a single instance of the monitoring process will receive persistent updates on all cameras' health status via inter-process communication. The software is designed for scalability, as each instance of the Camera process runs our fall detection algorithm independently. If a fall is detected by a specific camera, this will trigger a warning message to be sent from that Camera instance to the central Monitoring process module. This outputs an alert message to the central terminal to notify the necessary personnel of a fall at the suspected location. A screenshot of the potential fall location will also be outputted for inspection of false positives by the personnel.

Our monitoring system can be easily deployed to monitor large areas housing surveillance cameras. Ideally, it can be

used by elderly care facilities to centralize the supervision of the elderlies, where a single healthcare professional will be needed to standby for alerts.

Image Subtraction

In Núñez et al.'s method, optical flow stacks were fed through a VGG16 CNN network for the extraction of features before feeding the features into the FC-NN classifier. While optical flows have proven to be an effective and robust approach to foreground segmentation, especially in cases with dynamic backgrounds (Gutiérrez, Rodríguez and Martin), generating optical flows during image pre-processing requires a tremendous amount of computation time. It also requires light in the video to be static. In reality, dynamic lighting conditions may lead to unexpected displacement fields from optical flow, resulting in low classification accuracy. Thus, we decided to employ an alternative method that uses image subtraction input instead of optical flows.

Image subtraction will compare between two consecutive frames and identify the difference through the difference in pixels (Paranjape, 2000). Thus we decided to use image subtraction within the model as it would be computationally quicker as compared to relying on optical flow. This would help the model to be quicker in detecting falls in the real world, ensuring that alerts will reach caregivers faster to help the elderly get the prompt aid they require.

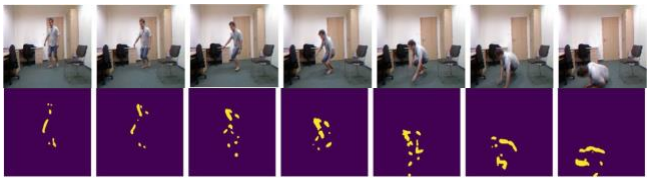

Figure 8: Comparison of images before image subtraction (top) and after image subtraction (bottom)

In the model (code), we first converted the image into a grayscale image, and then used Gaussian blur with a 9x9 kernel to reduce the noise in the image. Following that we subtracted the frames against the previous frame. After which, we put the resultant image through a binary threshold function, which assigns a binary value to the image based on whether it's above or below the threshold value.

The image is then dilated with a 3x3 kernel to fill in any holes that may occur from the previous processes. This process is iteratively done for all frames in a sliding window. This creates a set of images that will be fed to the model afterwards for training. A comparison of the pre-processed images and images post-processing is shown in Figure 8.

Results and Model Evaluation

Our group used the accuracy, recall, precision and F1-score of the two models as our performance measures. As seen from the results above in Figure 9, our model which used image subtraction had performed sub-optimally while our other proposed model with data augmentation, hyperparameter tuning using optical flows fared slightly better as compared to the model by Núñez et al. (2017), especially on recall. Thus data augmentation and tuning of the model's epoch and learning rate did provide improvements.

	Accuracy	Precision	Recall	F1
Núñez et al. 's (2017) method	99.35%	99.71%	94.25%	96.8%
Our proposed model (data augmentation, hyperparameter tuning)	99.95%	99.80%	98.23%	99.2%
Our proposed model (image subtraction)	93.16%	96.11%	94.48%	96.67%

Staying true to our group's initial conjecture, generating the image subtraction dataset did prove to be faster and less computationally intensive. However, our model had underperformed in terms of accuracy and precision as compared to the Núñez et al. (2017).

On deeper analysis, the optical flow dataset shows the apparent motion of the falling event based on the visual scene that is being captured, capturing explicitly both motions in the x and y direction. These bidirectional motions were then fed into the CNN. On the other hand, image subtraction only captures the subtractions between two consecutive frames and is not able to identify relative motions in explicit directions as well as optical flows. However, the lower computational cost of image subtraction will prove beneficial if the fall detection system is required to run in real time.

Limitations and Future Work

Our model utilizes CNN networks with optical flow on the data we have obtained. This would mean that the video feed from the cameras would have to undergo heavy processing before they can be fed to the model. This is compounded by a computationally expensive CNN model, which would add on to the total predicting time. As such, the model would not be able to predict the fall in a real time scenario, but the alert would actually only be sent some time after the fall has happened.

Another limitation would be ethical concerns. The model heavily utilizes CCTV video feed to detect the fall. We foresee this may bring up ethical concerns regarding privacy as well as security regarding the video feeds that were collected. We seek to avoid this problem by not

incorporating facial recognition capabilities within the model. Similarly, the video feed data will not be stored, the model will only use the real time video feed provided by the cameras. That way, there will be minimal data being stored and at risk of being stolen.

Next, the training data we utilized only contained images of a single person. This means that the model may be overfitted for images or video feeds that only contain a single moving person. However, this problem might only be overtly serious in the context of nursing homes, where a large portion of the people present are actually the elderly. This would entail that even with other people present, they may not be in the capacity to aid the fallen elderly. Our team hopes to overcome this shortcoming by training the model with video feeds containing 2 or more moving people. We will also employ object separation methods to properly distinguish between the moving people.

Team Contributions

Members	Contributions
Denise Beh Sin Jie (A0207982W)	Data augmentation, training of VGG16, optical flow generation, model hyperparameter tuning, fall alert system, report
Lee Zheng Yao Daniel (A0223259E)	Initial research lead, image subtraction
Mah Shian Yew Brendan (A0216349B)	Initial research, training of VGG16, report
Poh Yu Jie (A0216055M)	Data augmentation, image subtraction
Teo Ming Jun (A0216305R)	Initial research, training of VGG16, report

References

[Al-Aama, 2011] Al-Aama T. 2011. Falls in the elderly: spectrum and prevention. *Canadian family physician Medecin de famille canadien*, *57*(7), 771–776.

[Ang, Low and How 2020] Ang, G. C.; Low, S. L.; and How, C. H. 2020. Approach to falls among the elderly in the community. *Singapore Medical Journal*, *61*(3), 116–121. https://doi.org/10.11622/smedj.2020029

[Asher, 2020] Asher, S. (2020, July 4). *Tracetogether: Singapore turns to wearable contact-tracing covid tech*. BBC News. Retrieved from https://www.bbc.com/news/technology-53146360

[Beddiar, Oussalah and Nini, 2022] Beddiar, D. R.; Oussalah, M.; and Nini, B. (2022). Fall detection using body geometry and human pose estimation in video sequences. *Journal of Visual Communication and Image Representation*, *82*, 103407. https://doi.org/10.1016/j.jvcir.2021.103407

[Chhetri et al, 2020] Chhetri, S.; Alsadoon, A.; Al-Dala'in, T.; Prasad, P. W.; Rashid, T. A.; and Maag, A. 2020.. Deep Learning for vision-based Fall Detection System: Enhanced Optical Dynamic Flow. *Computational Intelligence*, *37*(1), 578–595. https://doi.org/10.1111/coin.12428

[Deng et al., 2009] J. Deng; W. Dong; and R. Socher. 2009 "ImageNet: a large-scale hierarchical image database," in Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255, Miami, Fla, USA, June 2009.

[Gutiérrez, Rodríguez and Martin 2021]Gutiérrez, Jesús ; Rodríguez, Víctor ; and Martin, Sergio. 2021 Comprehensive Review of Vision-Based Fall Detection Systems. Sensors, vol. 21, issue 3, p. 947

[Heng, 2021] Heng, J. 2021. For better aged care: The gaps in Singapore's nursing home market and alternative models of care for the elderly. https://www.businesstimes.com.sg/brunch/for-better-aged-care-the-gaps-in-singapores-nursing-home-market-and-alternative-models-of. Accessed 2022-03-30

[HRM Asia, 2018] HRM Asia Newsroom. (2018, October 24). *How one company tackled chronic talent shortages*. HRM Asia. Retrieved from https://hrmasia.com/tackling-recruitment-challenges-within-the-eldercare-sector/

[Kushwaha et al., 2020] Kushwaha, A.; Khare, A.; Prakash, O.; and Khare, M. (2020). Dense optical flow based background subtraction technique for object segmentation in moving camera environment. *IET Image Processing*, *14*(14), 3393–3404. https://doi.org/10.1049/iet-ipr.2019.0960

[Kwolek and Kepski 2014] Kwolek, B., and Kepski, M. 2014. Human fall detection on embedded platform using depth maps and wireless accelerometer, Computer Methods and Programs in Biomedicine. 117(3):489–501.

[Lin, 2021] Lin, C. 2021 . *In focus: Amid a rapidly ageing population, what are the missing pieces in Singapore's residential eldercare puzzle?* https://www.channelnewsasia.com/singapore/focus-amid-rapidly-ageing-population-what-are-missing-pieces-singapores-residential-eldercare-puzzle-2308441. 2022-04-13

[Ng 2021] Ng,W. 2021. Seven more nursing homes planned in next two years in estates such as Tampines, Punggol. https://www.straitstimes.com/singapore/health/seven-more-nursing-homes-planned-in-next-two-years-in-estates-such-as-tampines . Accessed 2022-04-14

[Paranjape, 2000] R. Paranjape. 2000. Fundamental Enhancement Techniques. 10.1016/B978-012373904-9.50008-8

[Tanwar et al., 2022] Tanwar, R.; Nandal, N.; Zamani, M.; and Manaf, A. A. 2022. Pathway of Trends and Technologies in Fall Detection: A Systematic Review. *Healthcare (Basel, Switzerland)*, *10*(1), 172. https://doi.org/10.3390/healthcare10010172

[Teo 2021] Teo, J. 2021.. *Hospitals in s'pore find ways to Lighten Nurses' covid-19 load to manage staff shortage*. https://www.straitstimes.com/singapore/health/hospitals-in-spore-find-ways-to-lighten-nurses-load-to-manage-staff-shortage. Accessed 15 April 2022

[Wang, Ellul and Azzopardi, 2020] Wang, X.; Ellul, J.; and Azzopardi, G. 2020. Elderly fall detection systems: A literature survey. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.00071