**Fall Detection from video footage AI Research and Development Project**

**Introduction**

Falls, particularly among the elderly, represent a significant health concern due to the potential for severe injuries and fatalities. The medical consequences are often worsened by prolonged periods of immobility following a fall, making timely detection and intervention critical. The challenge in fall detection lies in accurately identifying falls amidst various daily activities. The difficulty increases by the variability in how falls occur and the need to differentiate them from other activities, such as sitting down or stumbling (Espinosa et al., 2019). Key considerations include detecting the presence of a person, tracking their movement, and accurately interpreting their actions.

There are two primary approaches for fall detection systems: sensor-based and vision based systems. Sensor-based systems utilise wearable devices, ambient sensors, and smart devices to monitor movements and detect falls based on acceleration and other metrics. However, these systems may suffer from limitations, such as false positives and negatives, leading to unnecessary alarms or missed incidents. In contrast, vision based systems employ cameras and image processing techniques, often enhanced by deep learning algorithms, to analyse visual data for fall detection.

The final prototype developed utilises a vision-based approach for fall detection. The system employs pose estimation models, specifically pre-trained MediaPipe and MoveNet models, to detect and track body movements in video recordings. These models extract the position of the figure in each frame as row numeric data. This data is then classified as either a fall or a non-fall using trained binary classification models, including Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM) networks. SVMs are effective in high-dimensional spaces due to their ability to find optimal hyperplanes that separate classes with the maximum margin. MLPs, with their multiple layers and activation functions, are well-suited for capturing complex patterns in data. LSTMs, designed to handle sequential data, excel in capturing temporal dependencies and long-range patterns due to their memory cells and gating mechanisms, which make them ideal for sequence prediction tasks.

Two benchmark fall datasets, Le2i and URFD (University of Rzeszow Fall Detection) (Charfi et al., 2013; Kwolek & Kepski, 2014), were used in the system's development. The Le2i Fall Detection Dataset was employed for training, offering a comprehensive collection of video recordings that simulate falls and non-fall activities under various conditions. The URFD dataset was utilised for testing. It includes video and sensor data collected during controlled experiments, where participants performed falls and other daily activities. The primary reason for using the URFD dataset was to leverage its sensor data registered per frame, providing a reliable ground truth for evaluation.

The integration of deep learning techniques has advanced the vision-based fall detection systems. Such systems can be categorised into auto-encoder based, CNN based, and LSTM based techniques. The input data for these systems can include RGB video, thermal images, and skeleton data (Alam et al., 2022). Skeleton data provides a powerful and efficient means of analysing human motion, making it a valuable tool for fall detection and other applications related to human activity recognition. It can be easily integrated into various machine learning models, including LSTM and CNN architectures, allowing these models to learn fall patterns based on joint movements. The use of skeleton data also facilitates real-time processing, enabling immediate detection and response to falls (Chua et al., 2015).

This project investigates the effectiveness of different models in fall detection by utilising pose estimation techniques and classification algorithms. Firstly, we employed MediaPipe and MoveNet for pose estimation. Secondly, we used SVM, MLP, and LSTM for classification, developing six predictive models in total. These models were trained with annotated data from the Le2i dataset and tested on fall sequences from the URFD dataset. Our findings demonstrate that the models effectively distinguished between falls and non-fall activities, with accuracy rates ranging from 78% to 84%.

**Personal Reflection**

The project successfully utilised pose estimation models, specifically MediaPipe and MoveNet, to extract skeletal data from video footage. One of our main findings is the advantage of Mediapipe in execution time. These models were effective in detecting poses and, when combined with machine learning classifiers such as SVM, MLP, and LSTM, enabled differentiation between falls and non-fall activities. We trained and tested the models using benchmark datasets widely used in the academic field Le2i and URFD. We followed the recent approaches of solving the Fall detection problem. Although it is not a complete solution it is a great starting step.

To enhance the model, I suggest combining the features extracted by MediaPipe and MoveNet into a single dataset. MediaPipe provides consistent and reliable data by only returning results when confident in a detected pose, while MoveNet offers continuous data regardless of confidence. Therefore, merging these features could leverage MediaPipe's reliability and MoveNet's speed, resulting in a more robust model. This approach would shift the focus from comparing pose models to integrating them for improved prediction accuracy. Processing data in batches of frames, such as every 20 frames, could help the model recognise fall patterns more effectively. This could assist the LSTM model in creating sequential predictions based on groups of frames, reducing false detections where activities resembling falls are incorrectly classified. Real-time detection could also improve the prototype, enabling practical, real-world applications. Real time detection also can improve this prototype, changing the perspective to analysing live stream, would put our work into real-world applications.

My background in video processing techniques was particularly useful, as I handled tasks related to video manipulation and processing, gaining extensive knowledge and skills in libraries such as OpenCV and HTML display. One of the key challenges I faced was learning how the pretrained pose models were working to ensure consistent key point extraction across models and balancing real-time processing with accuracy. We needed not only to know how to use the models but also to understand the inner workings of the models for effective handling volumes of video data. Understanding and adjusting pre-trained models were crucial for our objectives, though the time spent on specific models may not be applicable in other contexts.

The project highlighted the need for thorough data preprocessing and integrating diverse data sources. In future projects, I will prioritise the quality of detection and use a sequential approach to account for movement patterns, as deep learning models can learn from these patterns even if they are not immediately interpretable.

Ethical considerations, especially regarding privacy and data protection, are crucial when using video footage for fall detection. We used publicly available datasets to avoid these concerns but anonymising data and securing consent remain essential.

References

Alam, E., Sufian, A., Dutta, P., & Leo, M. (2022). Vision-based human fall detection systems using deep learning: A review. *Computers in Biology and Medicine, 146*,

105626. 10.1016/j.compbiomed.2022.105626

Charfi, I., Miteran, J., Dubois, J., Atri, M., & Tourki, R. (2013). Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. *Journal of Electronic Imaging, 22*(4), 041106.

Chua, J., Chang, Y. C., & Lim, W. K. (2015). A simple vision-based fall detection technique for indoor video surveillance. *Signal, Image and Video Processing, 9*, 623-633.

Espinosa, R., Ponce, H., Gutiérrez, S., Martínez-Villaseñor, L., Brieva, J., & Moya-Albor, E. (2019). A vision-based approach for fall detection using multiple cameras and convolutional neural networks: A case study using the UP-Fall detection dataset.

*Computers in Biology and Medicine, 115*, 103520.

10.1016/j.compbiomed.2019.103520

Kwolek, B., & 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.