Fall risk monitoring scheme based on human posture estimation using Transfer learning

M.Sc. Research Project
Data Analytics

Chitra Raghavi Balasubramanian
Student ID: x18183409

School of Computing
National College of Ireland

Supervisor: Christian Horn

National College of Ireland Project Submission Sheet School of Computing

Student Name:	Chitra Raghavi Balasubramanian
Student ID:	x18183409
Programme:	Data Analytics
Year:	2020
Module:	M.Sc. Research Project
Supervisor:	Christian Horn
Submission Due Date:	17/08/2020
Project Title:	Fall risk monitoring scheme based on human posture estimation using Transfer learning
Word Count:	7822
Page Count:	21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Chitra Raghavi Balasubramanian
Date:	17/08/2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	Q
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	Q
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	Q

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Fall risk monitoring scheme based on human posture estimation using Transfer learning

Abstract

For the past few years, it has been witnessed a raise in fall detection-related research projects. Therefore, this paper presents the fall detection system using a vision-based approach. For conducting this experiment, publicly available dataset has been taken that contains a record of Falls and Activities of daily life (ADL). This project mainly deals with the classification of fall and not-fall images. Before training the model, human motion was tracked in the consecutive frames using optical flow algorithm. Further, these optical flow images are fed into Convolutional neural network (CNN) to extract the features from the images for detecting the fall events. VGG-16, ResNet-50 and DenseNet-205 were implemented based on transfer learning approach with the help of pre-trained "ImageNet" data. A 10-fold cross-validation was performed on the training set and applied it on the testing set to improve the accuracy of the model. Finally, the comparison has been made between three CNN architectures such as VGG-16, ResNet-50 and DenseNet-201, and found that ResNet-50 performed better in classifying the falls and not fall labels correctly. The performance of the developed model is evaluated using Accuracy and Confusion matrix.

1. Introduction

Globally, falls are the most important public health issues that leads to fatal and non-fatal injuries. As reported by Health service executive (HSE), presently 1 million people of Ireland's population is advancing years (aged 60 or older) and this will be expected to increase to 18% in a period of next 25 years. The number of elderly populations over 85 years or above will rise from 74,000 to 356,000 by 2041 CARDI (2015). According to the newsletter from Irish health, the economic cost of falls is estimated to be about 500 million euro each year. Also, HSE 2019 report have forecasted that the fall related injuries will cost 1 billion euro in 2020 and it may rise to 2 billion euro in the next 10 years (i.e., 2030) (Connell et al., 2020). Overall, falls are the major challenge for all age groups that causes serious injuries and the treatment for fall related injuries cost more for the economy. Due to these repercussion, Ireland's health service executive has inaugurated a project called AFFINITY [2018-2023] which aims to prevent falls and associated injuries in Ireland's ageing population in order to reduce the risk of falls health service (2019)

To address these problems, fall detection system has come into existence. In the initial stage, falls were identified using wearable devices, but when the device is far away from the users. It might be difficult for them to press the emergency button Ren,L. and Peng,Y. (2019) .Due to this reason, Vision based approaches have come into existence. But these system also facing trouble with accuracy and reliability. There are various techniques proposed in the related work that describes the machine learning algorithms used in recognizing the human activities for detecting the falls [Section 2]. In the consecutive section, we will review about the motivation, research question and research objectives for the proposed fall detection model.

1.1 Motivation

There are two significant approaches including fall detection and fall prevention to overcome the problem of falls in elderly people. The fall detection system has considered to be a more extensively researched problem and have gained a lot of concentration among the researchers. Similarly, this research project was chosen where the author can contribute to some extent for detecting the fall events using computer vision techniques. The first fall detection article published by brown studied only the large acceleration impact (Doulamis et al., 2018). For improving the accuracy of fall detector, other researchers have started to apply complex machine learning algorithm by integrating two or more sensors. A comprehensive analysis was conducted by (Singh et al., 2020) on fall detection technologies and the primary goal of their study is to suggest the detailed review on sensor and vision-based technology employed for detecting falls. It also suggests the choice of selecting appropriate classification algorithm and feature extraction techniques based on the accuracy and sensitivity of the previous research that has performed on fall detection. To enhance the flexibleness of fall detection systems, better detection algorithm and better machine learning algorithm is important. A review on available fall detection systems has been shown in figure 1.

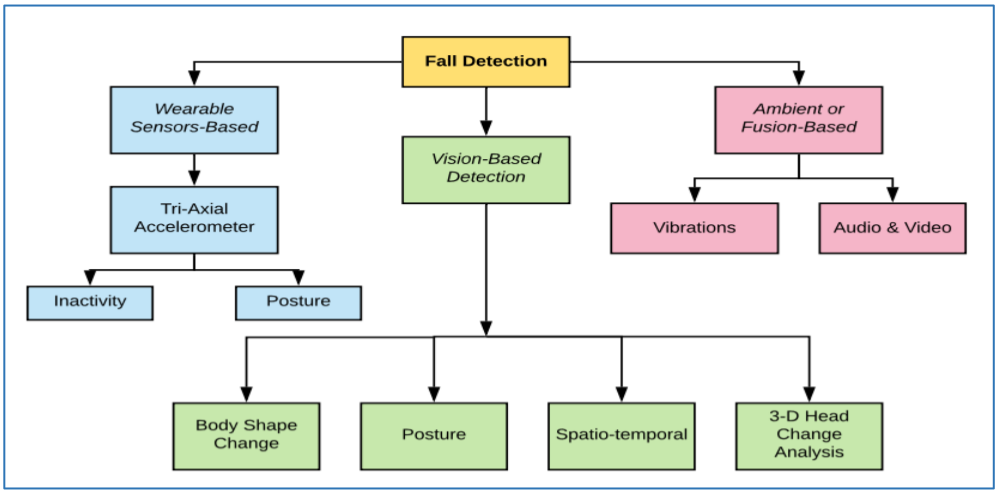

Figure 1: A Review of Fall Monitoring System

1.2 Research Question

This technical report tries to provide a solution to the following research question:

RQ: How accurately and efficiently transfer learning and deep learning algorithm classify fall events from optical flow images?

1.3 Research Objective

Convolutional Neural Networks (CNN) have achieved beyond state-of-the-art results by utilizing traditional computer vision techniques. The accuracy of the models is highly based on feature extraction technique (Doulamis et al., 2018). This paper presents an appropriate model in the domain of computer vision that uses CNN architecture for extracting generic features. Firstly, the RGB images of fall datasets were converted into optical flow images, then these images were fed into Convolutional neural network (CNN). Transfer learning-based approach will be implemented for reusing the weights of pre-trained classification layers. The architectures of CNN model such as VGG, ResNet and DenseNet were used for applying transfer learning. Finally, the comparison is made between the developed architectures of CNN (VGG-16, ResNet-50 and DenseNet-201). This performance will effectively detect and classify the fall activities, so economic burden causes due to fall can be reduces and death rate increasing following the fall accident can be minimised to great extent. (Table:1) provides the overview of the research objectives that are mentioned in this section.

Objectives	Description	Evaluation Metrics
Objective 1	Literature Review in Fall Detection Techniques and Identified Gaps	-
Objective 2	Image Pre-processing with Optical Flow Images Generator and Image Augmentation	-
Objective 3	Convolutional Neural Network Architecture developed for implemented models (VGG16, ResNet50 and DenseNet)	-
Objective 3	Implementation and Evaluation of VGG-16 trained on ImageNet to UR Fall Detection Dataset	Accuracy and Confusion Matrix
Objective 4	Implementation and Evaluation of ResNet-50 trained on ImageNet to UR Fall Detection Dataset	Accuracy and Confusion Matrix
Objective 5	Implementation and Evaluation of DenseNet-201 trained on ImageNet to UR Fall Detection Dataset	Accuracy and Confusion Matrix
Objective 6	Compare the results of developed models	-

Table 1: Research objectives for detection of fall events using Transfer learning

2. Literature Review on Fall Detection Techniques and Identified Gaps

2.1 Introduction

In the past decades, many machine learning and deep learning techniques have been proposed for fall detection system to identify whether a person undergoes fall or not. Based on the different types of video cameras and audio sensors falls can be detected. Some of these are based on depth sensors such as Microsoft Kinect, while others are based on wearable devices such as accelerometers. Different approaches and methodologies have applied to analyse human activity, extract the features, and classify the extracted features to detect the fall. Followings are the review on how the fall detection techniques and approaches have been used in the previous studies.

2.2 Review Based on Input Data Types for Detecting Falls

In this section, the review has been conducted based on different input types such as sensor-based, image and video-based data for detecting a fall event.

(Ogawa et al., 2020) presented a fall detection systems (FDS) based on temperature change in a non-wearable form, with the help of IR array sensor that analyse temperature distributions using machine learning algorithms for more accurate and quicker detection of a fall event. It is shown that the temperature of a human will increase when they change their position from standing to lying. Four types of actions are observed from the activities of daily life including fall, walking, lying and none. They have identified fall event based on the body temperature and areas occupied by a human. (Giuffrida et al., 2019) proposed FDS using wearable sensors namely tri-axial accelerometers and gyroscope for automatically requesting the necessary assistance. The human activities from the sensor signals are observed and processed for classifying falls and activities of daily life based on the time-intervals and slicing operation. (Sadreazami et al., 2019) have detected fall using Ultra-Wideband Radar by receiving radar time series signals. Deep residual networks are adopted to extract various features such as time, frequency, and time-frequency, particularly the multi variate features from the radar signal are employed for classifying the fall activities. The result from the Deep residual network is compared using classification metrics such as Gaussian support vector machine (GSVM), K-nearest neighbors (KNN), multi-layer perception (MLP) and dynamic time warping (DTW). (Kido et al., 2019) has presented the risk of fall in toilet rooms by implementing thermal imaging sensors. They have monitored the motions of a person in a toilet room, a person fall has been identified by comparing the normal activity patterns. Discriminated analysis and cross analysis were performed on the extracted patterns from the sensors and the accuracy was identified by creating the discriminant equation.

(Sase et al., 2018) studied the computer-vision techniques by analysing the depth-based video files for supporting older people to overcome the risk of falls. They have used two types of dataset namely URFD and SDU fall in which young volunteers have participated in the fall activities. At first, the frames from the depth video is extracted and the region of interest (ROI) is detected for each frame. Based on the threshold value the system has identified whether the fall activity has occurred or not. (Sheieh et al., 2009) proposed an elderly fall detection system using Multi-camera video dataset which includes falling positions such as lying and bending down on the ground. The multi-image streams are processed by exploring four steps at the same time including image fetch, image processing, human edge sequence generation and pattern recognition to reduce the time-delay.

(Abraham et al., 2019) presented emergency handling system for detecting falls. Once the fall event has occurred, the live recording of the condition is captured and sent to the nearest police station with the exact location of the users. The experiment is conducted on Sisfall dataset that contains 38 participants including young people, adults and elderly persons. In (silva et al., 2020), the author has used UMA fall dataset which consists of 19 experiments of Activities of Daily Life (ADL) and falls. The classification algorithm such as Random Forest and Decision Tree are used after extracting the features from the image frames, random forest achieved highest accuracy rate and learning rate. In (Hagui et al., 2015) the author has done experiment on fall detection by observing the video files using Coupled Hidden Markov Model (CHMM) for silhouette extraction, classification, and modelling. They have used Le2i databases and

YouTube video clips which contains 20 Activities of Daily life (ADL) and 30 fall activities. The performance of the CHMM is evaluated using F1-score and ROC curve.

From this section, it has been shown that sensor-based approaches provide acceleration measures such as horizontal, vertical and acceleration (Giuffrida et al., 2019). However, they also used the computer vision techniques to determine the prediction of sensor-based approach. Therefore, the vision-based detection is more convenient and effective for identifying the fall activities because it mainly focuses on frames of the images.

2.3 An investigation on Image Processing Techniques for identifying the Region of Interest (ROI)

(Thummala et al., 2020) have implemented fall detection using background subtraction method for detecting an object in an image. The separation of the background from foreground has been done by using Gaussian Mixture Model to track humans. Motion history image and Shape deformation are employed for extracting the feature from the video sequence. Shape deformation is used to overcome the problem of changing object while it is moving. Similarly, the movement pattern of an object is described by replicating a continuous sequence of motion. Motion history image have achieved an accuracy of 72.5% for identifying the abnormal activity and 68.9% for detecting a person fall from a chair and 52.3% for a person's normal activity. Overall, they have achieved 95% of accuracy. (Bhavya et al., 2016) has implemented a novel approach for fall detection using motion vector and accumulated map image. The foreground extraction has done by employing background subtraction method with bounding box and morphology filtering for reducing the noise occurred in the images. Likewise, combined local-global approach was applied to estimate the motion vector. With the motion vectors, they have found some difficulties in distinguishing lying down and falling down. Due to this reason, the accumulated image map is obtained from the foreground object. At last, the fall event was classified by applying K-nearest neighbor (KNN). (Wang et al., 2016) has proposed deep convolutional neural network for deep action recognition. In their work, they have used stacked optical flow field and stacked wrapped optical flow field.

(Zerrouki et al., 2018) proposed a vision-based approach for human action recognition based on discrepancy in body shape. Instead of using geometric shape, they have used pixel representing body's silhouette for performing image segmentation. While implementing the background subtraction method for extracting the features, the pixel was combating some noises. To reduce this problem a filtering phase is added using morphological operators such as dilation, holes fitting and erosion. The action in the video sequence is recognized using the ratios computed for different activities: standing, bending, sitting, and squatting. Consequently, the classification algorithm such as KNN, Naïve Bayes, SVM and AdaBoost is used for detecting the fall activities and the performance of the model is evaluated by kappa coefficient and recognition rate. It has been shown that Adaboost has achieved higher accuracy of 93% when compared to other models.

(Lofti et al., 2018) studied a novel computer vision technique for performing fall detection based on spot of human head, projection histogram, motion information and variations in shape orientation. Background subtraction is employed for detecting the variation between the present image and the background image. The motion of the human is identified by employing tMHI (timed Motion History Image) . The changes in human shape is analysed by fitting an ellipse around the human body to get an information about the body posture. Then the extracted features from an image or a frame is passed as an input to the Multilayer Perceptron Neural network (MLP) to do the classification step and MLP has achieved 99% of accuracy. (Feng at al., 2017) presented the home surveillance fall detection based on RGB images using UR fall detection dataset. Instead of using background subtraction for getting the foreground, the system was implemented using R-CNN to detect the bounding boxes for Motion history image (MHI) and Histogram of oriented gradient (HOG). R-CNN is used along with pretrained VGG16 model to obtain a high precision rate. After detecting the MHI, the extracted features from MHI are sent to SVM classifier to perform classification task. The motion is not continuous in all the frame. Due to this reason they have set a threshold value for classifying fall and not-fall activity. Their model has achieved 96% of precision and 98% of recall.

(Hsieh et al., 2017) proposed fall detection system using Convolutional neural network with optical boundary region. Boundary based models are relatively fast and time expensive. To overcome this problem, CNN is used. They have obtained two learning phases for recognizing the object in an image frame namely Feature feedback mechanism scheme and Rule based motion detection. The optical flow feedback system and sharing mechanism is used for identifying the object while moving across the frame. Likewise, the rule base motion detection contains two steps, first step builds the region of point interest by employing fast region-based CNN. In second step, the information retrieved from faster CNN was used as the spatial information of 3DCNN object detection. Thus, the CNN model using optical flow image as a feature extraction technique have achieved 82.7% of accuracy. (Ge et al., 2017) implemented fall detection system based on CNN. The spatial and temporal streams are used with CNN for capturing the motion and appearance feature. The features from the image frames were extracted by using CNN and fine-tuning is employed through backpropagation for updating the weights of the pre-trained network. They have used 16 layers, 13 convolutional layers and last 3 are fully connected dense layers. Their research has used SVM classifier for classifying fall and non-fall events, the performance of this model is evaluated with accuracy, true positive rate, and false positive rate.

2.4 A Review on Transfer Learning Techniques

A network needs a greater number of training data to learn the features effectively. (Sadreazami et al., 2019) research study implemented transfer learning approach for detecting the occurrence of fall events. An ImageNet is a pre-trained approach used for employing transfer learning. It uses an ultra-wideband radar signals for obtaining time-frequency representations. To fine-tune the VGG network, time-frequency representations are used. On the other hand, classification algorithms such as K-nearest neighbors (KNN), Gaussian support vector machine (GSVM) and linear support vector machine (LSVM) were also executed. Finally, they have evaluated the performance of these models with accuracy, precision, and sensitivity. The evaluation metrics are performed with 3-cross fold validation. The transfer learning technique with VGG16 have achieved the overall accuracy of 95%, precision of 96% and sensitivity of 96%.

(Yhdego et al., 2019) proposed Musculoskeletal modelling with deep convolutional neural network (DCNN) and Transfer Learning for fall detection. They have conducted this experiment on UR-fall detection dataset. The accelerometer data has been taken for their proposed approach; median filtering is used to eliminate the high-frequency noise. To implement transfer learning, the 1D signal data is transformed to RGB images that helps to attains the feature vectors from the pre-trained model. The Alexnet architecture is employed for classifying the images and continuous wavelet transform (CWT) is employed for processing the accelerometer time-series data. They have trained the CNN on Huge annotated dataset of images that has been used for extracting the deep features. Since ImageNet dataset is different from RGB images of the fall detection dataset, thus the last three layers of Alexnet was fine-tuned for classifying the fall datasets based on RGB image. Instead of using SoftMax, they have used support vector machine for classifying the fall and non-fall events. Results are evaluated using accuracy, precision, sensitivity, and specificity. With transfer learning, the model has achieved 92% of accuracy.

(Camerio et al., 2019) processed video files using deep convolutional network. Three-steps of image processing techniques are employed; Optical flow is used for detecting relative motion among nearby pixel by performing Lucas-Kanadc and Gunner Franeback. The Pose estimation gives the key-points of human anatomical positioning and it generates a frame that can be used on a feed-forward network. The multi-stream model is employed for getting information such as temporal, spatial, and spatial temporal information, this model contains Vgg-16, Hidden Markov Model and K- nearest neighbor. On the top of that, VGG16 and AlexNet were trained on ImageNet and UCF101 respectively for improving the accuracy. The evaluation of result is conducted based on 5-cross fold validation.

2.5 Summary

Based on the literature review analysis, different approaches were explored in accordance with the fall detection system. Even though many classification algorithms have applied on detecting fall events, Convolutional neural network have achieved state-of-art results in the field of image classification and outperformed the human being’s performance. Therefore, this technical report implements Convolutional Neural Network (CNN) for classifying the fall activities using optical flow image. Transfer learning is implemented using ImageNet data to pretrain the model as mentioned in (Sadreazami et al., 2019) (Yhdego et al., 2019). This project extending the use of optical flow algorithm based on dense flow algorithm as mentioned in (Wang et al., 2016) and it also extend the use of VGG16, Resnet50 and DenseNet-201 based on transfer learning for fall recognition data. This chapter provides the overview of existing system available for detecting falls as mentioned in the introduction. The next portion delivers details about the proposed methodology.

3. Scientific Methodology Approach

3.1 Introduction

The most popular methodology in data mining is the Cross Industry Standard Process. The first phase is to understand the business that generates clarity to the business problem and so delivering a business focus for the project. The necessity for this understanding is to analyse the data and try new methods. Steps involved in CRISP-DM methodology are Business understanding, Data understanding , Data preparation, Modeling, Evaluation and Deployment (Trnka. A., 2010). The stages of the complete CRISP-DM approach related to this project are shown in (Figure 2) and the above-mentioned steps are briefly discussed in the following.

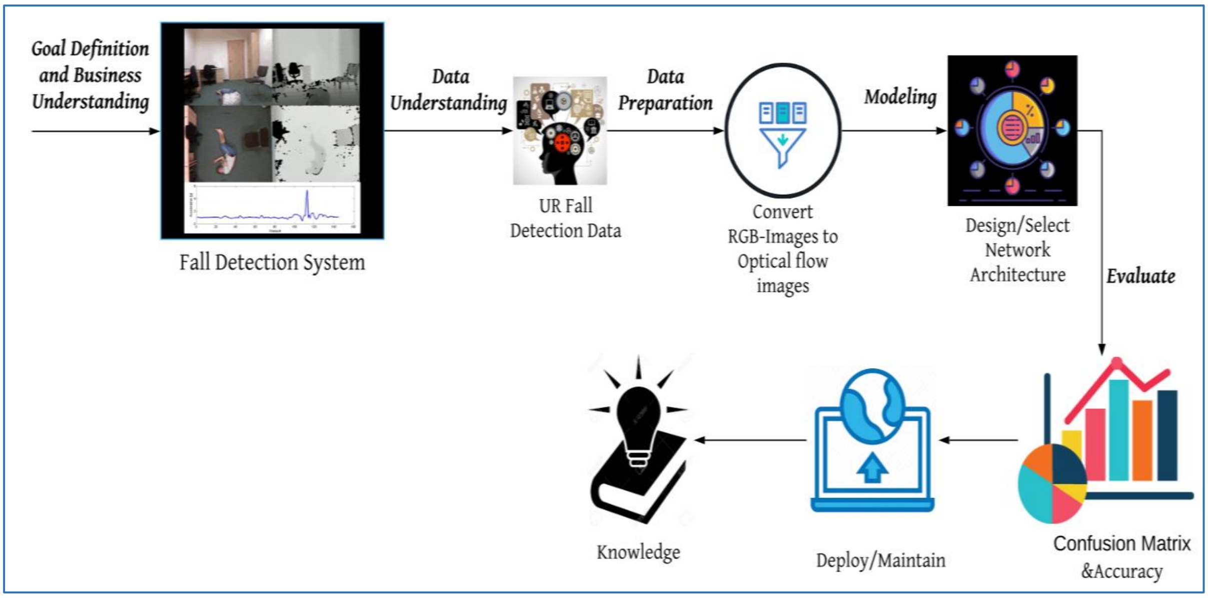

Figure 2: Fall Detection Methodology Approach used

3.2 Business Understanding

In this research project, the main objective is detecting whether a person will undergo a fall or not from Activities of Daily Life (ADL) and Falls. Earlier responses to fall activities is important because that might decrease the serious consequences such as health-based risks and injuries. The proposed system would be helping to classify person's normal activity and fall activity that in turn helps the health care services to prevent the occurrences of fall.

3.3 Data Understanding and Data Fetching

The quality of the data must be assessed to check whether it will support the goal defined in the business understanding stage. The data understanding phase begins from obtaining the data and gaining knowledge about it. The data collected for this research is UR fall detection and it has been taken from publicly accessible portal that was offered by Interdisciplinary Centre for Computational Modelling from University of Rzeszow. The dataset contains 70 sequences (30 falls + 40 activities of daily living), each video stream is stored in separate zip archive in form of png image sequence. It contains depth and RGB images for camera 0 and camera 1. This project acquires the RGB images from camera 0 for performing the fall classification problem. Here, all the images of a video are compressed in a folder. The details about total number of frames of each dataset have shown in Table 2.

Falls	30
Not-falls	40
Total Frames	9,838
Number of Fall frames	900
Number of No-fall frames	8,938

Table 2: Total number of frames

3.4 Data Preparation

The data preparation phase shown in (Figure 3) is essentially important because the machines should understand the image data. Therefore, the RGB images are converted to optical flow images for a good understanding of human gestures. The optical flow algorithm has chosen due to the changes in background lightning condition. The motion of consecutive image frame is represented using optical flow algorithm in order to avoid unnecessary information from the images and to overcome the impact of background image features. By removing the influence of static images, only motion of the human is taken into account for identifying the fall easily (He et al., 2017). Before training deep neural network, image augmentation techniques such as random rotation, zoom range, horizontal and vertical flip were created to build a good image classifier to achieve better results and to enhance the performance of neural networks.

3.5 Modeling

This stage of the project implements the modeling techniques used for classifying images of fall and not-fall. Once the optical flow images are generated, the appropriate tool for automatic feature extractors is to use Convolutional neural network (CNN). Therefore, CNN is implemented for extracting the generic features from the images. This paper was selected to use three architectural design for action recognition with CNN based fall classifier namely VGG16, ResNet50 and DenseNet201 (Amjoud and Amrouch, 2020). The main reason for selecting these architectures was due to its better results achieved in other related fields. For improving the CNN model, transfer learning is applied on developed architecture which helps the model in recognizing the image more precisely based on the pre-trained weights as discussed in (Sadreazami et al., 2019) (Yhdego et al., 2019).

3.6 Evaluation

The evaluation metrics is used to present how good our approach is in predicting falls. The evaluation of the result for binary classification problems is generally derived from the confusion matrix. This project uses confusion matrix and accuracy score to evaluate the performance of VGG16, ResNet50 and DenseNet201 (Yhdego et al., 2019) (Ge et al., 2017).

3.7 Conclusion

The CRISP-DM methodology was chosen to present the methodology part of this project and revised based on my requirements. The sourced dataset was obtained from the Interdisciplinary Centre for Computational Modelling from University of Rzeszow and the image processing techniques was applied based on optical flow extraction. Convolutional neural network (CNN) and transfer learning was selected for further implementation with the appropriate deep learning architecture such as VGG-16, ResNet-50 and DenseNet-201 for classifying the fall events.

4. Design Specification

The following architectural diagram will provide the tools and techniques that were used to develop this project are shown in (Figure 3). These steps involve pre-processing the given image, tracking the motion of consecutive image, applying transfer learning for the developed architecture and performance evaluation for the applied models.

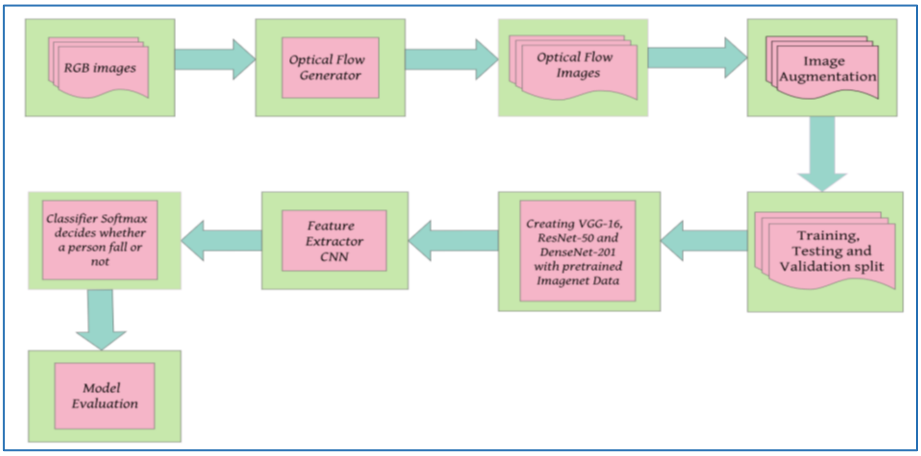

Figure 3: The design architecture of proposed work

5. Implementation

This section provides the overall implementation of the system including the detailed review on image processing techniques applied for tracking the motion in the consecutive image frames. Also, how the architectural models are implemented using transfer learning to classify fall events accurately from optical flow images. Finally, the comparison has been made among the developed models.

5.1 Pre-processing the Image

There are two stages of data pre-processing steps were applied for this project namely Optical flow images generator and Image augmentation. The dataset of vision-based fall detection system was pre-processed independently. The detailed review on these techniques are mentioned in the following.

5.1.1 Creation of Optical Flow Images Generator

Optical flow is one of the most suitable algorithms to model quick events like falls. The human gestures class are difficult to identify with a single frame and a static scene. The extraction of motion requires to be developed for a nice understanding of the gestures class. Therefore, the optical flow algorithm is used for representing the motion of human gestures as displacement vector fields between two consecutive image frames. The motion pattern throughout the stacked frame can be represented by stacking 2L optical flow images.

TVL-1 optical flow algorithm proposed by (Zach et al., 2007) was chosen due to its better performance when background lightning condition differs. Like the Farneback method, TVL-1 also uses two frames to estimate the optical flow and it is more robust against noise than the approach proposed by (Horn and schunck., 2003). To create stacks $O \in R^{224x224x2*L}$, the horizontal and vertical component of the vector field (d_t^x and d_t^y) are taken separately and stack them together. The optical flow images are computed using dense optical flow extraction tools for videos provided by (Wang et al., 2016) based on OpenCV. A schematic diagram of RGB images converted to optical flow images for detecting falls is given in (figure 4).

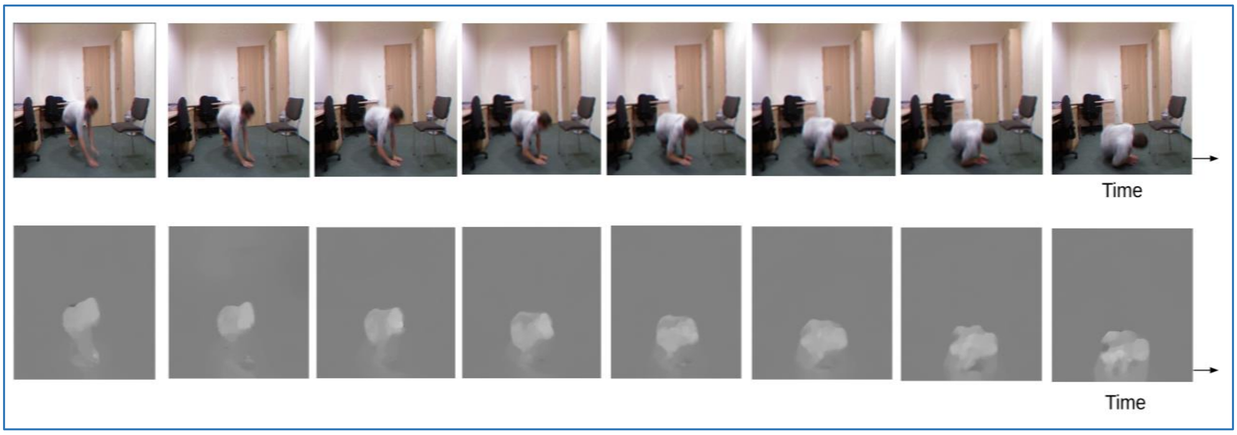

Figure 4: Sample of sequential images of a fall from UR Fall Detection Data and corresponding optical flow images

5.1.2 Image Augmentation

Before training the data, optical flow images are expanded with data augmentation techniques, to overcome the problem of translation invariance and overfitting. This step not only increases the dataset but also provides the opportunity for CNN to learn from a greater variation in data. The process of image augmentation is defined as creating duplicates of the original images by rotating, flipping, and zooming. Image augmentation techniques used in this project was shown in (table 3). ImageDataGenerator class is used for performing Data augmentation are; the batch size of 16; the images are zoomed in with the range of 2; and the images were horizontally and vertically flipped with an image rotation of 90 degrees by rotating it right and left Khoshgoftaar,T and Shorten, C. (2019)

Augmentation parameters	Values used
Batch Size	16
Zoom Range	2
Rotation Range	90
Horizontal Flip	True
Vertical Flip	True

Table 3: Image Augmentation Values

5.2 Convolutional Neural Network Architecture

The architectural model of Convolutional neural network (CNN) consists of convolution layer, activation layer, pooling layer, and fully connected layer. These operations in CNN helps to extract the image features and compresses the volume of data. CNN is useful for achieving supervised learning problems like image classification by training the model through gradient descent and back propagation algorithms. The description of operational procedure of CNN used for all the models (VGG16, ResNet50 and DenseNet-201) are defined in the following:

1. The CNN model was built with 4 layers. A **Sequential**() model is used to initialize the neural network by passing a list of layers using the add() method which then, collectively constitute a model.

2. The first layer is **GlobalAveragePooling2D()** for spatial data that sums out the spatial information. it also reduces the total number of parameters in the model that helps to minimize overfitting.
3. The second layer is **Dropouts(),** 50% of dropout regularization is used to improve the performance of neural network and to overcome the co-adaptation of feature detector. This layer doubles the number of iterations essential to converge. However, training time for each epoch is fewer.
4. The third layer is **BatchNormalization()** used to decrease the training time and to achieve better performance by allowing each layer of a network to learn more independently by itself.
5. The last layer is **Dense()** with 2 neurons, the activation function for this layer is set to softmax layer to compute the probabilities corresponding to the two output layers, which matches our number of presumed classes in our fall detection classification problem.
6. The compile() method of the sequential model class is accomplished to configure the learning process. Three arguments are required for compilation: a loss function, an optimizer, and a list of metrics.
7. The compilation used for this classification problem are the binary_crossentropy loss function, the Adam optimizer, and the accuracy metric.

5.3 Experiment 1: Implementation, Evaluation and Results of Transfer Learning with VGG16

5.3.1 Implementation

The VGG-16 architecture is a simple and widely used convolutional neural network architecture for ImageNet. In this project, the VGG16 net was trained in the ImageNet dataset (Deng et al., 2009). From the ImageNet datasets, the network learns generic features for object recognition such as corners and textures. Initially, these pretrained weights facilitates the network to learn more optical flow-oriented features for classifying the fall and not-fall events. The default input shape of 224 X 224 pixels with 3 input channels are employed. Once the VGG 16 architecture is created and trained on ImageNet then it was used in CNN with four trainable layers (refer Section 5.3). The trainable layers of CNN are freeze for preventing their weights being updated during training. Moreover, the ReduceLROnPlateau is used for reducing the learning rate when a metrics has stopped improving. The stochastic gradient descent optimizer is employed to minimise the error during training. A learning rate was chosen to be 0.0001 and a momentum of 0.9 is used. VGG 16 was implemented based on 10-fold cross validation in order to improve the accuracy of training and validation set.

5.3.2 Evaluation and Results

This experiment is conducted by splitting the data into training set and testing set with a ratio of 60:40, Where 60% of data is randomly selected for training and 40% is for testing. The optical flow images of fall detection data are divided into training and validation set. The validation with training data is implemented, where the number of Epoch is 10. The evaluation metrics such as confusion metrics, accuracy, and loss are used for presenting the results. The performance of the model for training and validation data is shown by plotting Accuracy and Loss graph per epoch. In this experiment, the accuracy is calculated for the testing set. The VGG16 net has achieved the testing accuracy of 86% for classifying fall and not-fall events.

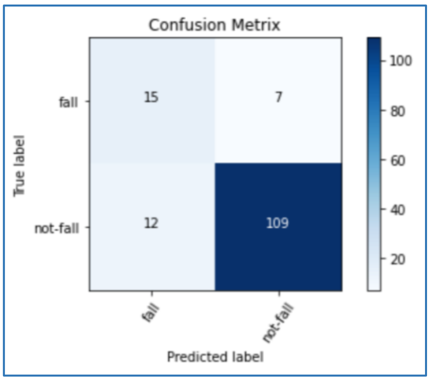

Figure 5: Confusion Matrix for VGG-16

The confusion matrix shown in (Figure 5) provides information about how well the model is capable of predicting the fall event. Out off 22 fall events, 15 fall events were predicted correctly, and 7 events were predicted wrongly as not-fall. Even though, when the data was imbalanced, VGG -16 performed better in terms of achieving the fall events. The accuracy and loss graph are shown in (Figure 6)

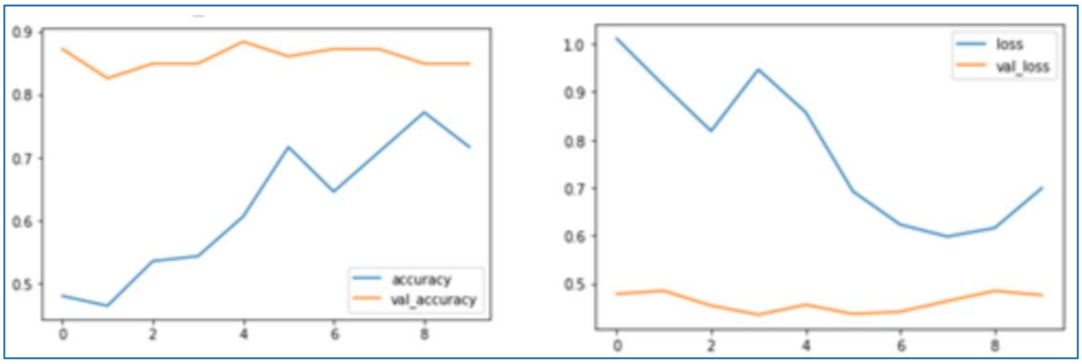

Figure 6: Accuracy and Loss graph of VGG-16 while training

5.4 Experiment 2: Implementation, Evaluation and Results of Transfer Learning with Resnet-50

5.4.1 Implementation

The ResNet-50 architecture pre-trained CNN architecture are employed for classifying optical flow images into 2 classes namely fall and not-fall. In this project, ResNet-50 architecture have been trained on a subset of ImageNet database (Deng et al., 2009). As mentioned in VGG16 (refer section 5.4.1), the network learns the generic features from the ImageNet for identifying the fall events using optical flow images. The image is resized with Input shape of 224X224 pixel and 3 input channels. After initializing ResNet-50 with the pre-trained weights, the CNN layers are passed to extract the features from the images (Refer Section 5.2). Then, the layers of CNN were frozen from preventing the weights being updated during training process. The stochastic gradient descent (SGD) with momentum of 0.9 and learning rate of 0.001 was employed. The 10-fold cross validation is applied for improving the accuracy of the training set.

5.4.2 Evaluation and Results

The ResNet-50 model was trained to classify optical flow images of fall and not-fall events. This model uses 60% of data for training and 40% of data for testing. For classifying the fall events, the performance of the model is mainly evaluated with the accuracy score of testing set and with the confusion matrix. The testing set of ResNet-50 has achieved 94% of accuracy after 10-fold cross validation is applied on the training set and the validation set.

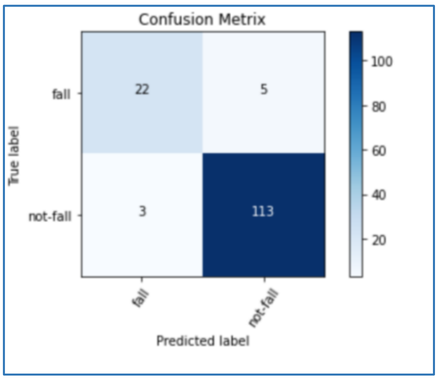

Figure 7: Confusion Matrix for ResNet-50

From figure 7, the confusion matix of ResNet model shows that the model performed well in terms of predicting both the fall and the not-fall events. With the imbalanced data, the ResNet-50 model has correctly classified 22 fall labels and misclassified 5 event as not-fall. Therefore, the ResNet model was good in understanding and extracting the features from the images and performing well in detecting and classifying the fall events. It also helps in reducing the false alarm that occurs in the fall detection system. The accuracy and loss graph of training and validation set was shown in figure 8. This graph was plotted before applying 10-cross fold validation in the training and validation set.

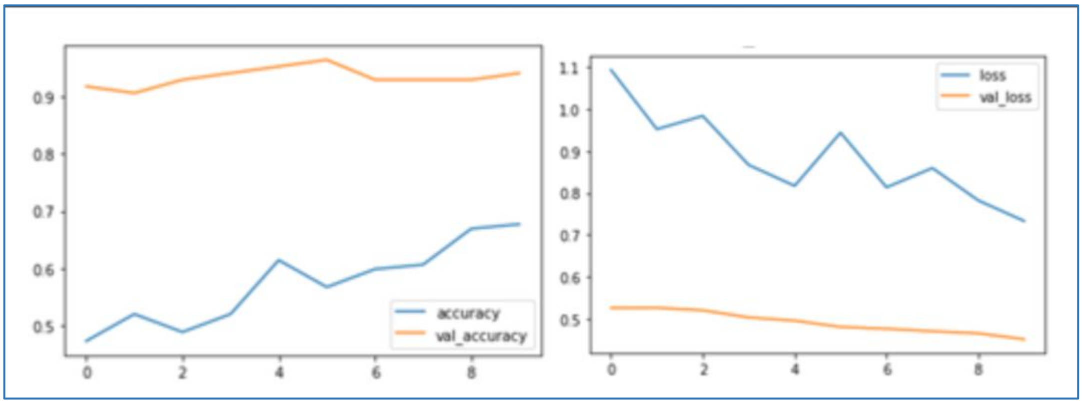

Figure 8: Accuracy and Loss graph of ResNet-50 while training

5.5 Experiment 2: Implementation, Evaluation and Results of Transfer Learning with DenseNet-201

5.5.1 Implementation

The Dense Convolutional Network (Dense-201) is trained on ImageNet for learning and extracting the features for detecting and classifying the falls and the not-falls images. This model is implemented similar to ResNet-50 and VGG16 (Refer Section 5.2.1 and 5.3.1). The stochastic gradient descent (SGD) with momentum is used for optimizing the parameters during the training process. The learning rate of the model is chosen as 0.0001 and momentum of 0.9 is used.

5.5.2 Evaluation and Results

The DenseNet-201 pretrained CNN architecture are implemented to classify optical flow images of fall events into two classes. The performance of DenseNet-201 is evaluated based on accuracy and confusion matrix. Similar to the other models, the dataset for training and testing set were divided into 60:40 ratio. The testing accuracy of DenseNet have achieved the accuracy of 79 % . The 10-fold cross validation is applied on the training and the validation set which helped in improving the accuracy of the model.

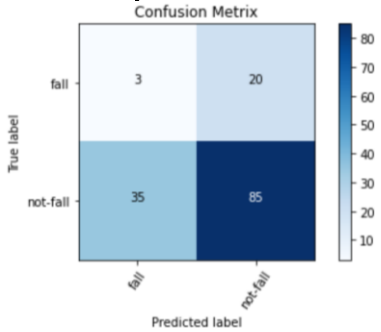

Figure 9: Confusion Matrix for DenseNet-201

The confusion matrix for DenseNet-201 is shown in figure 9. It has been clearly observed that DenseNet-201 was unable to understand the pattern of data for predicting the falls label. This may be due to the problem of class imbalance, since falls contains only 30 sequences whereas the not-falls contains 40 sequences. The accuracy and loss graph of the training and validation is shown in figure 10. The testing accuracy of the model is 79%, which is good but the model could not be able to predict the labels of falls. The accuracy and loss graph of DenseNet-201 is shown in figure 10.

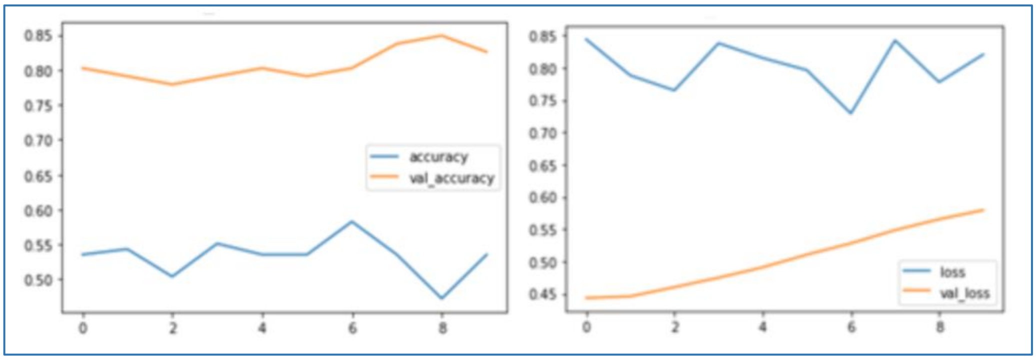

Figure 10: Accuracy and Loss graph of DenseNet-201

6 Discussion and Comparison of Developed Model (VGG16, ResNet-50)

Implemented Models	Accuracy of testing score
VGG-16	86%
ResNet-50	94%
DenseNet-201	79%

Table 4: Overall Accuracy

In this research, the comparative analysis has been made between 3 architectural model of CNN such as VGG-16, ResNet-50 and DenseNet-201. All the models are implemented with 10-cross fold validation. The feature extraction method and pre-trained networks were used to detect and classify the falls and not-falls optical flow images. Table 4 gives the comparison of test accuracy for the implemented models. The developed architectures of CNN were trained for 10 epochs. From table 4, it has been shown that ResNet-50 achieves better classification accuracy than VGG16 and DenseNet-201 architectures. The DenseNet-201 contains more layers than ResNet-50 and VGG-16. However, for classifying the optical flow images of fall and not-fall events, ResNet-50 pre-trained model performed well. This is due to ResNet-50 model understands the pattern and representation of the image more effectively and precisely. Similarly, VGG 16 also performed well in terms of identifying the fall labels when compared to DenseNet-201.

7 Conclusion

The experiment has been conducted on publicly available dataset namely UR fall detection. This vision-based fall detection system was created with transfer learning techniques from action recognition in order to overcome the impact of minimum number of samples in the fall dataset. The generic features from the optical flow images are learned through transfer learning for classifying the fall events. The correlation between consecutive frames were considered by staking them together for tracking the motion of human in the images. Furthermore, image augmentation has been applied to overcome the overfitting problem during training process. Then, optical flow images were fed as an input to the neural network. The proposed method is implemented with the CNN architecture which were tested on grey-scale images to classify fall and not-fall events. This research shows that ResNet-50 have outperformed and achieved the highest accuracy of 94% when compared to other two models. The accuracy of VGG-16 and DenseNet-201 has achieved the accuracy of 86% and 79% respectively. Whereas, in the DenseNet-201 architecture chances are higher that the real fall event could be misclassified as not-fall event which could not solve the real purpose of the fall detection system. Thus, the proposed transfer learning method will help to overcome these problems. Additionally, it has been observed that this method is appropriate for feature extraction from vision-based data and classification can be done when the data samples are small. The ResNet-50 model achieved 94% of accuracy in this study which is not enough for building an effective classification system, Thus, the accuracy of this model should be improved to build effective and accurate fall detection system.

7.1 Future work

The future work can be done by analysing the RGB images rather than optical flow images. Hence, the selection of deep learning algorithm should be more complex for observing the motion representations from RGB images. In order to continue this research, fine tuning can be applied to the classification algorithms for detecting the fall more accurately.

Acknowledgement

I sincerely thank my supervisor Christian Horn for supporting and guiding me for completing my research project. He guided me in every stage of this project by providing feedbacks throughout the conducted sessions. I would also like to thank my parents and sisters for believing and supporting me to achieve my goals.

References

Amjoud,A. and Amrouch,M. (2020)’ Convolutional Neural Networks Backbones for Object Detection’, International Conference on Image and Signal Processing, pp. 282-289,Springer Link. Available at: https://link.springer.com/chapter/10.1007/978-3-030-51935-3_30 [Accessed 17 August 2020]

Bhavya, K. R., Park, J., Park, H., Kim, H. and Paik, J.(2016),` Fall detection using motion estimation and accumulated image map `, *IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Seoul*, pp. 1-2, doi: 10.1109/ICCE-Asia.2016.7927288.

Cameiro, S. A., Da Silva, G.P., Leite, G. V., Moreno, R, Guimaraes, S. J. F. and Pedrini, H.(2019)` Multi-Stream Deep Convolutional Network Using High-Level Features Applied to Fall Detection in Video Sequences `, *International Conference on Systems, Signals and Image Processing (IWSSIP), Osijek, Croatia,* pp. 293-298, doi: 10.1109/IWSSIP.2019.8787213.

CARDI (2015) Illustrating Ageing in Ireland North & South Key Facts and Figures. Availableat:https://www.cardi.ie/publications/illustratingageinginirelandnorthsouthkeyfactsandfigures [Accessed 15 July 2020].

Connell,M., Kenny,R., Donoghue, O. ‘Measures of Health and Function that Predict Future Falls’, The Irish Longitudinal Study on Ageing, PP. 130-150, w3-key-findings-report. Available at: https://tilda.tcd.ie/publications/reports/pdf/w3-key-findings-report/Chapter%206.pdf [Accessed 15 July 2020].

Deng, J., Dong, W., Socher., R. and Li, L. (2009) ‘ImageNet: A large-scale hierarchical image database’, *IEEE Conference*, doi: 10.1109/CVPR.2009.5206848.

Doulamis.N., Doulamis.A. and Protopapadakis.E, (2018) ‘Deep Learning for Computer Vision: A Brief Review’, *Computational Intelligence and Neuroscience, 2018,* pp. 1-30, Hindawi. doi: https://doi.org/10.1155/2018/7068349.

Feng, Q., Gao, C., Wang, L., Zhang, M., Du, L. ans Qin, S.(2017)` Fall detection based on motion history image and histogram of oriented gradient feature `, *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS),* pp. 341-346, doi: 10.1109/ISPACS.2017.8266500.

Ge, C., Gu, I.Y. and Yang, J.(2017)` Human fall detection using segment-level cnn features and sparse dictionary learning `, *IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), Tokyo,* pp. 1-6, doi: 10.1109/MLSP.2017.8168185.

Giuffrida, D., Benetti, G., De Martini, D. and Facchinetti, T.(2019)` Fall Detection with Supervised Machine Learning using Wearable Sensors `, *IEEE 17th International Conference on Industrial Informatics (INDIN),* Helsinki, Finland, pp. 253-259, doi: 10.1109/INDIN41052.2019.8972246.

Hagui, M., Mahjoub, M. A and Elayeb, F.(2015)` Coupled Hidden Markov Model for video fall detection `, *11th International Conference on Natural Computation (ICNC), Zhangjiajie*, pp. 675-679, doi: 10.1109/ICNC.2015.7378071.

Health service (2019) About AFFINITY National Falls and bone health project 2018-2023. Available at: https://www.hse.ie/eng/services/list/4/olderpeople/falls-prevention-and-bone-health/ [Accessed 17 July 2020].

He, L., Liu, X. and Pei, L.(2017)` A High-Precision Optical Flow Method Based on TV-L1 Model and Neighborhood Patch `, *International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), Wuhan*, pp. 321-324, doi: 10.1109/ICIICII.2017.89.

Horn, K. and Schunck, B. (1981) ‘Determining optical flow’, *ELSEVIER*, 17(1-3), Science Direct. doi: https://doi.org/10.1016/0004-3702(81)90024-2.

Hsieh, Y. and Jeng, Y.(2018)` Development of Home Intelligent Fall Detection IoT System Based on Feedback Optical Flow Convolutional Neural Network `, *IEEE Access*, 6, pp. 6048-6057, 2018, doi: 10.1109/ACCESS.2017.2771389.

Huang, Z., Liu, Y., Fang,Y. and Horn B. K. P.(2018)` Video-based Fall Detection for Seniors with Human Pose Estimation `, *4th International Conference on Universal Village (UV), Boston, MA, USA*, pp. 1-4, doi: 10.1109/UV.2018.8642130.

Ke, Q., Liy, J., Bennamoun, M., An, S., Sohel, F. and Baussaid.F. (2018) “Computer Vision for Human–Machine Interaction”, *ELSEVIER*, pp.127-145, Science Direct. doi: https://doi.org/10.1016/B978-0-12-813445-0.00005-8.

Khoshgoftaar,T and Shorten, C. (2019) ‘A survey on Image Data Augmentation for Deep Learning’, *Journal of Big Data*, pp. 1-48, Springer Open. Avalilable at: https://link.springer.com/content/pdf/10.1186/s40537-019-0197-0.pdf [Accessed 17 August 2020]

Kido, S., Miyasaka, T., Tanaka, T., Shimizu, T. and Saga, T.(2009)` Fall detection in toilet rooms using thermal imaging sensors `, *IEEE/SICE International Symposium on System Integration (SII), Tokyo*, pp. 83-88, doi: 10.1109/SI.2009.5384550.

Lofti, A., Albawendi, S., Powell, H., Appiah, K. and Langensippen, C.(2018)` Supporting Independent Living for Older Adults; Employing a Visual Based Fall Detection Through Analysing the Motion and Shape of the Human Body `, IEEE Access, 6, pp. 70272-70282, 2018, doi: 10.1109/ACCESS.2018.2881237.

Ogawa, Y and Naito, K.(2020) ‘ Fall detection scheme based on temperature distribution with IR array sensor ‘, *IEEE conference*, Las vegas, USA, doi: 10.1109/ICCE46568.2020.9043000.

Plyer, A., Le Besnerais, G. and Champagnat, F.(2016)`Massively parallel Lucas Kanade optical flow for real-time video processing applications`, *J Real-Time Image Proc*, 11, pp. 713–730 doi: 10.1007/s11554-014-0423-0

Ren,L. and Peng,Y. (2019) ‘Research of Fall Detection and Fall Prevention Technologies: A Systematic Review’, *IEEE*, IEEE conference. doi:10.1109/ACCESS.2019.2922708.

Sadreazami, H., Bolic, M. and Rajan, S.(2019)` Residual Network-Based Supervised Learning of Remotely Sensed Fall Incidents using Ultra-Wideband Radar `, *IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan,* pp. 1-4, doi: 10.1109/ISCAS.2019.8702446.

Sase, P. S. and Bhandari, S. H.(2018)` Human Fall Detection using Depth Videos `, 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, pp. 546-549, doi: 10.1109/SPIN.2018.8474181.

Shieh, W. and Huang, J.(2009)` Speedup the Multi-camera Video-Surveillance System for Elder Falling Detection `, *International Conference on Embedded Software and Systems, Zhejiang,* pp. 350-355, doi: 10.1109/ICESS.2009.62.

Silva, J., Gomes, D., Sousa, I. and Cardoso, J. S.(2020)` Automated Development of Custom Fall Detectors: Position, Model and Rate Impact in Performance `, *IEEE Sensors Journal,* 20(10), pp. 5465-5472, doi: 10.1109/JSEN.2020.2970994.

Singh, A., Rehman, S. U., Yongchareon, S., and Chong, P. H. G. (2020)` Sensor Technologies for Fall Detection Systems: A Review `, *IEEE Sensors Journal*, 20(13), pp. 6889-6919, doi: 10.1109/JSEN.2020.2976554.

Thummala, J. and Pumrin, S.` Fall Detection using Motion History Image and Shape Deformation `, *8th International Electrical Engineering Congress (iEECON), Chiang Mai, Thailand,* pp. 1-4, doi: 10.1109/iEECON48109.2020.229491.

Wang,L., Xiong,Y., Wang,Z., Qiao,Y, and Tand,X. (2016)’ Temporal Segment Networks: Towards Good Practices for Deep Action Recognition’, *CoRR, abs/1608.00859,* arXiv. Available at: https://arxiv.org/abs/1608.00859

Yhdego, H et al.,(2019)` Towards Musculoskeletal Simulation-Aware Fall Injury Mitigation: Transfer Learning with Deep CNN for Fall Detection `, *Spring Simulation Conference (SpringSim), Tucson, AZ, USA,* pp. 1-12, doi: 10.23919/SpringSim.2019.8732857.

Zerrouki, N., Harrou, F., Sun, Y. and Houacine, A.(2018)` Vision-Based Human Action Classification Using Adaptive Boosting Algorithm `, *IEEE Sensors Journal*, 18(12), pp. 5115-5121, doi: 10.1109/JSEN.2018.2830743.