The University of Manchester

# Sensor Signal Processing for Human Gait Deterioration Analysis by Machine Learning

A thesis submitted to The University of Manchester

for the degree of Doctor of Philosophy (Ph.D.) in the Faculty of Science and Engineering

2021

**Abdullah S Alharthi**

Sensing, Imaging and Signal Processing Group

Department of Electrical and Electronics Engineering

# Contents

**Contents** 1

**List of Symbols** 6

**List of Figures** 8

**List of Tables** 14

**Abstract** 16

**Declaration of Authorship** 18

**Copyright Statement** 18

**Acknowledgements** 19

**Chapter. 1 Introduction** 20

1.1 Introduction 21

1.2 Aims and Objectives 22

1.3 Thesis Structure 23

1.3.1 Parkinson's disease gait 24

1.3.2 Cognitive Load for healthy gait 24

1.3.3 Multimodality fusion 25

1.3.4 Publications 25

1.3.5 Contribution of authors 26

**Chapter.2 Background and Methodology** 28

2.1 Introduction 29

2.2 Deep Learning 29

2.3 Deep Convolutional Neural Networks (CNNs) 30

2.3.1 Perceptron layer 31

2.3.2 Convolutional Layer 32

2.3.3 Pooling Layer ..... 33
2.3.4 Activation function ..... 33
2.4 Training and Testing ..... 34
2.4.1 Forward propagation ..... 34
2.4.2 Backpropagation ..... 35
2.5 Regularization Methods ..... 36
2.6 Evaluation Measure ..... 38
2.7 Long Short-Term Memory Networks ..... 39
2.8 Layer-Wise Relevance Propagation (LRP) ..... 40
2.9 Python ..... 44
**Chapter.3 Literature Review ..... 47**
3.1 Introduction ..... 49
3.2 Background ..... 51
3.2.1 Gait Parameters ..... 51
3.2.2 Applications of Gait Analysis ..... 54
3.2.3 Deep Learning for Gait Analysis ..... 55
3.3 Gait Modalities ..... 57
3.3.1 Video Sequence ..... 57
**3.3.1.1 Video Sequence Databases** ..... 58
**3.3.1.2 CNN Architectures** ..... 60
**3.3.1.3 Single Deep ANNs** ..... 61
**3.3.1.4 Dual Deep ANNs** ..... 63
**3.3.1.5 Transfer Learning** ..... 66
3.3.2 Wearable Sensors ..... 66
**3.3.2.1 Normal Gait Analysis** ..... 69
**3.3.2.2 Abnormal Gait Recognition** ..... 71
3.3.3 Floor sensors ..... 74

3.4 Multi-Modality Gait Sensor Fusion ..... 77
3.5 Discussion ..... 81
3.5.1 Spatiotemporal Character of Gait Data ..... 81
3.5.2 Multi-sensor Incentives for Deep Learning from Gait Data ..... 82
3.6 Literature Update ..... 83
3.6.1 Video Sequence ..... 84
3.6.2 Wearable and floor sensors. ..... 84
3.7 Conclusion ..... 85
**Chapter. 4 Parkinson's disease Gait ..... 86**
4.1 Introduction ..... 88
4.2 DCNN AND LRP Background ..... 90
4.2.1 Deep Convolutional Neural Network (DCNN) ..... 90
4.2.2 Layer-Wise Relevance Propagation (LRP) ..... 91
4.2.3 Perturbation ..... 93
4.2.4 Dataset ..... 95
4.3 DCNN Implementation and Results ..... 95
4.3.1 Data Pre-processing ..... 95
4.3.2 Proposed DCNN Architectures ..... 96
**4.3.2.1 DCNN** ..... 99
**4.3.2.2 Parallel DCNN** ..... 99
**4.3.2.3 Quadruplet DCNN** ..... 99
4.3.3 Feature Learning and Classifications ..... 100
4.3.4 LRP and Model selection ..... 101
4.3.5 Gait Event Assignment Using LRP ..... 104
4.4 Discussion ..... 106
4.5 Conclusion ..... 112
**Chapter.5 Gait Under Cognitive Load Part (1) ..... 114**

5.1 Introduction ..... 116
5.2 Methodology ..... 117
5.2.1 iMAGiMAT System ..... 117
5.2.2 Data Acquisition ..... 119
5.2.3 Convolutional Neural Network (CNN) ..... 120
5.2.4 LRP ..... 121
5.2.5 LRP Heatmapping Validation ..... 122
5.3 Results ..... 123
5.3.1 Gait spatiotemporal Classifications ..... 123
5.3.2 Gait Spatiotemporal Classification Decomposition ..... 124
5.4 Discussion and Conclusion ..... 127
**Chapter. 6 Gait Under Cognitive Load Part (2) ..... 133**
6.1 Introduction ..... 135
6.2 Background ..... 138
6.2.1 Deep CNNs and LRP ..... 138
6.2.2 iMAGiMAT System ..... 138
6.2.3 Gait ..... 139
6.3 Methodology ..... 140
6.3.1 Data Acquisition and Pre-Processing ..... 140
6.3.2 CNN Structure ..... 141
6.3.3 LRP ..... 142
6.4 Experiments and Results ..... 143
6.4.1 Overview ..... 144
6.4.2 Gait signature classification ..... 145
**6.4.2.1 Experiment 1: 21 Subject gait signature verification** ..... 146
**6.4.2.2 Experiment 2: Gender Classification** ..... 147
**6.4.2.3 Experiment 3: 21 Subjects Cognitive Load Classification** ..... 148

**6.4.2.4 Experiment 4: Single Subjects Cognitive Load Classification** ..... 149

**6.4.2.5 Experiment 5: Binary Classification under Cognitive Load** ..... 150

6.4.3 LRP Analysis of Gait Spatiotemporal Classific-ations ..... 152

6.5 Discussion ..... 155

6.5.1 Classification of Gait Signatures under Cognitive Load ..... 155

6.5.2 Interpretation of Classifications ..... 158

6.5.3 Benefits and Limitations ..... 163

6.6 Conclusion ..... 163

**Chapter. 7 Gait Speed from Multimodal Data ..... 165**

7.1 Introduction ..... 167

7.2 Data and Data-Processing ..... 167

7.3 Machine Learning Methods ..... 168

7.3.1 Vision Transformer ..... 169

7.3.2 Multi-stream CNN and Single-stream CNN ..... 170

7.3.3 Hybrid CNN+LSTM network ..... 171

7.3.4 Statistical Classifiers Algorithms ..... 171

7.4 Results and Discussion ..... 172

7.5 Summary and Conclusions ..... 175

**Chapter. 8 Summary and Contributions, Future Directions, and Conclusions 176**

8.1 Summary and Contributions ..... 177

8.2 Future direction ..... 178

8.3 Conclusion ..... 180

**References 181**

# List of Symbols

| | |
|---|---|
| $i$ | Iteration index {$i$=0,1,2,…} |
| $j$ | Iteration index {$j$=0,1,2,…} |
| $k$ | Iteration index {$k$=0,1,2,…} |
| $m$ | Number of training data examples {$m$=0,1,2,…} |
| $\sigma$ | Nonlinear monotonically increasing activation function |
| $U$ | Perceptron for layer |
| $x$ | Input data to Neural network |
| $l$ | Layer |
| $\omega$ | Neural network weight |
| $C$ | Convolution operator |
| $\circ$ | Element-wise multiplication |
| $d$ | Iteration index {$d$=0,1,2,…} |
| $J(\theta)$ | Cost function |
| $P$ | Performance measure |
| $\delta$ | Regularisation parameter |
| $\eth^{(L)}$ | The output layer error |
| $\eth_j^{(l)}$ | Error of node $j$ in layer $l$ |

| | |
|---|---|
| $\alpha_j^{(l)}$ | Activation of node *j* of layer *l* |
| $t$ | Training example |
| $f_c(x)$ | Output function for class c |
| $c$ | Gait class |
| $R$ | Relevance score of neural network |
| $R_i^l$ | Relevance score for $i^{\text{th}}$ neuron of layer *l* |
| $R_j^{l+1}$ | Relevance score for $j^{\text{th}}$ neuron in the previous layer |
| $h(x)$ | Hypothesis function |
| $\theta$ | Trainable parameters |
| $\hat{y}$ | Neural network output value |
| $y$ | Neural network actual output value |
| t | Training example |
| TP | True positive |
| TN | True negative |
| FP | False positive |
| FN | False negative |
| $o$ | Total number of frames to represent one gait cycle |
| $tb$ | The frame number in the sequence (moment of time) |

| | |
|---|---|
| $\psi, \varphi$ | Values in the 2D image coordinate |
| $F^{tb}$ | Binary silhouette of the subject at time $tb$ |
| α , β | Relevance parameters |
| $SA$ | Spatial Average |
| $s$ | The number of plastic optical fiber sensors |
| $n$ | The number of the data |
| $\mu$ | Mean value |
| œ | Standard deviation |
| F1 | F1 score |
| $q$ | The number of F1 scores accuracies. |
| $g$ | Neural network weight and bias |
| $T$ | Learning rate |

## List of Figures

Figure 2.1. Activation functions.....30

Figure 2.2. Illustration of a convolutional neural network with two-dimensional convolutional blocks and subsampling as pooling operation followed by Perceptron layer.....31

Figure 2.3. Working principle of a single layer perceptron.....31

Figure 2.4. Convolutional layer principle. (a) 1D convolution operation, (b) 2D convolution operation.....32

Figure 2.5. Machine learning training and validation learning curves. .....36

Figure 2.6. Dropout method, a) original layer units; b) layer after dropping units. .....37

Figure 2.7. Confusion matrix table. .....37

Figure 2.8. LSTM Memory block or cell. .....39

Figure 2.9. DNN and LRP signal processing flow. Red arrows indicate the relevance propagation flow [20]. .....40

Figure 2.10. Illustration of the LRP propagation procedure applied to a neural network. The prediction at the output is propagated backward in the network, using various propagation rules, until the input features are reached. The propagation flow is shown in re[24]. .....42

Figure 2.11. Anaconda Spyder environment. .....45

Figure 2.12. Google Colaboratory notebook. .....45

Figure 3.1. Important gait events and intervals in a normal gait cycle. .....53

Figure 3.2. From right to left: video sequence, silhouette images and EGI image [37]. ...58

Figure 3.3. [65]: High-level difference architectures for small view-angle differences: a) Siamese CCN with probe and gallery input; b) Triplet CNN with positive probe, negative probe and gallery; Low-level difference variants of a) and b) for substantial view-angle differences: c) single CNN with probe and gallery; d) Siamese CNN. .....61

Figure 3.4. Gait GEI images at 14 viewing angles [65]. .....65

Figure 3.5. Convolutional neural network to extract and classify gait features from wearable inertial measurement unit with accelerometer and gyroscope sensors [77]. .70

Figure 3.6. iMAGiMAT footstep imaging system. .....73

Figure 3.7. 3D CNN+LSTM architectural detail of the proposed multimodal human gait recognition using VS and WS fusion [116]. .....80

Figure 4.1. Example GRF data recorded at 100 frames per second for healthy subject and subjects with PD severity ratings 2, 2.5, and 3, with sample length of 500 time-frames xi $=$ [x1 … x18] $\in \mathbb{R}500 \times 18$. The signals with lower amplitude (xi $=$ [x1 … x16]<500 N)

represent pressure sensor signal under each foot (different colors for each of the 8 sensors). In each sample, the calculated sum of the 8 sensor outputs for each foot is also shown additionally (x17&x18>500 N, different colors for left and right foot). .....91

Figure 4.2. Proposed DCNN architectures; a) Single DCNN, b) Parallel DCNN. The color-coding of boxes: convolution layers and fully connected layer (blue); pooling layers (green); concatenation layers and flattening layers (brown); dropout layers (navy); input at the top and softmax output layers (gray). The diagrams are generated using Neutron from GitHub repository based on the models weights and biases. .....97

Figure 4.3. Proposed DCNN architectures Quadruplets DCNN. .....98

Figure 4.4. Single 2D-CNN predictions on 1281 sample is shown as confusion matrix.100

Figure 4.5. Parallel 2D-CNN predictions on 1281 sample is shown as confusion matrix. ..... 101

Figure 4.6. Quadruplet 2D-CNN predictions on 1281 sample is shown as confusion matrix. ..... 103

Figure 4.7. LRP method selection by perturbation steps progressively removing information with the highest relevance scores. Steeper initial decrease indicates better identification of gait events with most weight in the classifications. ..... 103

Figure 4.8. Semi-log plot of the perturbation effect on the proposed DCNNs architectures. The decline in accuracy results from progressively removing information from the input data based on LRP-SPF and re-predict, at each step, 100 steps total. ..... 104

Figure 4.9. Perturbation with Gaussian noise based on LRP relevance scores, used for LRP and DCNN selection. Top plot: healthy gait processed GRF signals; Middle plot: GRF with Perturbation noise; Bottom plot: LRP relevance scores for the selected LRP Sequential Preset a Flat (LRP-SPF) and Parallel DCNN. $xi = [x1 \dots x18]$ represent the 18 signals after data standardization (see figure 4.2 and section 4.3.1). The LRP plot is dominated by $x18$ because all the signals are plotted on top of each other in the temporal domain. LRP relevance score are highly dependent on the temporal changes, whereas spatial variation does not affect the model prediction. ..... 105

Figure 4.10. Gait events processed SA (see equation 4.5) signal top. The highlighted gray area in (a) is explained in (b) based on gait events for one foot from figure 4.2.1 as: A-Heel strike, B- Loading response or flat foot, C- Mid-stance or single support, D- Terminal stance or heel rising, E- Pre-swing or double-limb support, F- Initial swing and Mid-swing or toe-off, G-Terminal swing..... 106

Figure 4.11. LRP method applied on randomly selected samples for the four PD severity ratings. SA of gait spatiotemporal signals: green; SA for LRP relevance scores over the same temporal period: blue. Vertical red bars with number labels display consistency with gait events listed below with capital letters as per figs. 3.1 and 4.10: 1, 3 and 6 - Heal strike and foot flattening (A); 2- Mid-stance and single support (C); 4- Loading response after the double support interval (B), 5 and 8 - Terminal swing and ready for the heel strike (G), 7- (F) Initial swing and Mid-swing or toe-o ..... 109

Figure 4.12. A deconvolution decomposition method applied to explain the parallel DCNN model prediction for PD severity ratings. SA of gait spatiotemporal signals is in black plot and model decomposition using deconvolution for each class in red plot. The deconvolution plot spikes every 10 frames..... 110

Figure 4.13. A Deep Taylor decomposition method applied to explain the parallel DCNN model prediction for PD severity ratings. Similar to figure 4.11 and 4.12 the Deep Taylor decomposition plot spikes every 10 frames..... 111

Figure 4.14. A LRP sequential preset a flat (LRP-SPF) applied to explain the single DCNN model prediction for PD severity ratings. Similar to figure 4.11, 4.12 and 4.13 the single DCNN model decomposition plot spikes every 10 frames..... 113

Figure 5.1. iMAGiMAT design on the right and actual system on the left. ..... 118

Figure 5.2. CNN network architecture for gait classification, and LRP. ..... 119

Figure 5.3. Gait spatial average of gait spatiotemporal signals (see equation 5.3), y axis: transmitted light intensity [%]; x axis: frames. Gait events recorded by the sensors in a typical gait cycle[131]: 1- Heel strike, 2- Foot-flattening, 3- Single support, 4- Opposite Heel strike, 5- Opposite Foot-flattening, 6- Double support, 7-Toe-off, 8- Foot swing, 9-

Heel strike, 10- Double support, 11- Toe-off, 12- Foot swing, 13- Opposite Heel strike, 14-Single support, 15-Toe-off. ..... 124

Figure 5.4. CNN model confusion matrix classification of 15 subjects..... 125

Figure 5. 5. LRP heatmaps validation by perturbation technique. Information with the highest relevance scores is progressively removed and the test samples are re-predicted. Steeper initial decrease indicates better identification of gait events with most weight in the classifications. a) shows the model predictions in 20 steps based on removing relevance scores using Deep Taylor decomposition and random removal of information. b) shows the model performance after 1000 steps of information removal..... 126

Figure 5.6. Raw IMAGiMAT POF signals top, corresponding LRP scores signals bottom. ..... 127

Figure 5. 7. LRP methods are applied on three subjects' testing data (three pairs of graph rows), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. SA of gait spatiotemporal signals: green; SA for LRP relevance signals over gait temporal period: blue. Vertical red bars with numbers display consistency with gait events as per figure 5.4: 1- Initial foot-flattening, 2- Heel strike and Foot-flattening, 3-Single support, 4- Between opposite heel strike and double support, 5- Foot-flattening, 6-Between foot-flattening and single support, 7- Between heel strike and double support, 8- Toe-off, 9- Between foot-flattening and single support..... 132

Figure 6.1. Proposed CNN network architecture for gait classification. Color-coding of boxes: convolution layers and fully connected layer (blue); pooling layers (green); flattening layer (brown); dropout layers (navy); input at the top and softmax output layers (gray). The diagrams are generated using Neutron [192] from GitHub repository based on the models' weights and biases. ..... 137

Figure 6.2. Gait signature classification confusion matrix for 21 subjects. The diagonal squares are the true positives, in this case 100% of classification and elsewhere are the false positives (0%)..... 144

Figure 6. 3. Model training and validation loss in (a) and model accuracy in (b) for 21 subjects in 5 classes. ..... 145

Figure 6.4. Confusion matrix for classification under cognitive load: 21 subjects, 5 classes. ..... 149

Figure 6.5. Model loss (a) and training and validation accuracy (b) under cognitive load, 21 subjects and 5 classes. ..... 150

Figure 6.6. Representative gait cycle spatial average of a spatiotemporal signals (see equation 6.5). Gait events recorded by the sensors in a typical full gait cycle of two steps ([1] as A,B,C,D,E,F,G): 1- Heel strike, 2- Foot-flattening, 3- Single support, 4- opposite Heel strike, 5- opposite Foot-flattening, 6- Double support, 7-Toe-off, 8- Foot swing, 9- Heel strike, 10- Double support, 11- Toe-off, 12- Foot swing, 13- opposite Heel strike, 14- Single support, 15-Toe-off. ..... 151

Figure 6.7. LRP methods applied on a single subject from experiment 1 testing data (each column is one pair), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. SA of gait spatiotemporal signals: black; SA for LRP relevance signals over gait temporal period: blue; POF LI (Plastic Optical Fiber Light Intensity). Vertical red bars with numbers display correspondence to gait events as per figure 6.6: 1,5- Loading response or Foot flat and Double support, 2,3,4 - Loading response or Foot flat and Single support. ..... 154

Figure 6.8. Consistent with identifying gait events relevant for the CNN prediction, random subject from experiment 1 gait events are: 1,4- Loading response or Foot flat and Single support, 2,3- Foot swing and opposite Heel strike, 5- Loading response or Foot flat. ..... 158

Figure 6.9. LRP methods applied on normal gait samples (from different subjects) from experiment 3 testing data, to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. Gait events are: 1,2,3- loading response or Foot flat and Double support. ..... 159

Figure 6.10. LRP methods applied on a single subject from experiment 5 testing data (each column is one pair), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. Gait events are: 1- Heel strike, 2- Toe-off, 3-

between Foot swing and opposite Heel strike, 4- between Double support and Toe-off. ..... 160

Figure 6. 11. LRP heatmaps validation by perturbation technique for experiment 1. Information with the highest relevance scores is progressively removed and the test samples are re-predicted. Steeper initial decrease indicates better identification of gait events with most weight in the classifications. a) shows the model predictions in 30 steps based on removing relevance scores using LRP Sequential Preset a Flat (LRP-SPF) and random removal of information. b) shows the model performance after 300 steps of information removal. ..... 162

Figure 7.1. 52 reflective cutaneous markers, placed by anatomical palpation on the participants. Markers (front side – green, back side – red) are numbered by their order in the spatial domain. Electromyography signals recorded from probes of the right leg muscles are illustrated in black: E1-E8. ..... 169

Figure 7.2. CNN+CNN network architecture, Input 1: TRM, input 2: GRF, input 3: EMG, input 4: Forces and Moments. ..... 170

Figure 7.3. Hybrid CNN+LSTM network architecture. Input 1: TRM, input 2: GRF, input 3: EMG, input 4: Forces and Moments. ..... 171

Figure 7.4. Gait speed classification in 5 classes: confusion matrix using hybrid CNN+LSTM ..... 173

Figure 7.5. Gait speed classification in 4 classes. (a) confusion matrix and (b) ROC using Multi-Stream CNN. ..... 174

## List of Tables

Table 3.1. Results for Gait Recognition from VS. ..... 62

Table 3.2. Results for Gait Recognition from WS. ..... 68

Table 3.3. Results for Gait Recognition from Floor Sensors. ..... 75

Table 3.4. Results for Gait Recognition from Multi-Modality Sensor Fusion.....79

Table 4.1 Number of subjects with the severity rating. .....94

Table 4.2. Datasets subject's discerption. .....96

Table 4.3. Models F1-Score for each Dataset and F1-Score, Mean and Standard Error with Datasets Combine. .....99

Table 4.4. PD classification results on PhysioNet three datasets. ..... 107

Table 6.1. Models Classification Accuracy for Each Subject. ..... 146

Table 6.2. F1-score predictions for binary classification. ..... 147

Table 6.3. . F1-Score Predictions for Comparison of CNN with Classical Classifiers. ..... 148

Table 7.1. Size of Sample Frames and Dimensionality by Data Type..... 168

Table 7.2. F1 scores to predict gait speed for 10 random subjects. ..... 172

# Abstract

Gait analysis is a growing field of research that utilises non-invasive sensors and machine learning for various applications. The focus of this thesis is to provide sensor signal processing and sensor fusion methodologies, as well as highlight the importance of cognitive load on gait tasks. This is achieved by linking machine learning outputs to decisions based on visual observations by humans (or quantitative parameters of gait derived from such observations), which are tested and routinely implemented in current healthcare practice.

Firstly, the challenges put forward in literature are identified and critically analysed with a view to the variations in classification accuracies. This allows to identify the challenges originating from the sensor modality limitations and sensor data processing methods. It is concluded that deep learning achieves better classification results, compared to the classical statistical analysis; sensors under the foot, compared to wearable sensors; multimodality sensors fusion, compared to single modality sensors.

Secondly, the importance of cognitive load in healthy gait analysis is investigated. This results in more robust classification accuracies with 100% F1 scores for subject's identity verification, as well identifying the weaker features that are critical for models' predictions using Layer-Wise Relevance Propagation. The analysis of gait deterioration due to cognitive decline caused by Parkinson's disease allows robust results to be achieved by the proposed methods to rate the severity of Parkinson's disease (PD) from sensors under the foot. The classification accuracies achieved for PD are 98% F1 scores for each of PhysioNet.org dataset and 95.5±0.28 F1 scores for the combined PhysioNet dataset. The classification result is used to show the impact of PD severity on gait, by linking the clinical potentially observable features to the models' outputs.

Lastly, sensor fusion of four modalities is demonstrated for the prediction of gait speed, by implementing machine learning methods that fuse gait signals in the deep layers of a deep learning model. The models yield state-of-the-art results and demonstrate the robustness of using cognitive load for identifying healthy subjects; the staging of PD severity based on model's perturbation (by progressively removing information with the highest relevance scores and analyse the model outputs); the use of a large number of

people (never seen by the model) data to test the robustness of the models' sensors fusion. An additional finding of this work is that the effect of cognition deterioration on gait impacts the body balance and foot landing and lifting from/on the surface in both classification cases, cognitive load in healthy gait and cognitive decline in PD gait.

# Declaration of Authorship

I, Abdullah S Alharthi, declare that no portion of the work referred in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

# Copyright Statement

i. The author of this report (including any appendices and/or schedules to this report) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this report, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has entered into. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the report, for example graphs and tables ("Reproductions"), which may be described in this report, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this report, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=487), in any relevant Report restriction declarations deposited in the University Library, the University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's Guidance for the Presentation of Reports.

# Acknowledgements

All praise and thanks are due to the Almighty for blessing me with everything I have today.

Thank you to . . .

My supervisor Prof. krikor Ozanyan, it has been an enriching experience to be his PhD student. I appreciate all his contributions, time, ideas and insights to make my PhD experience productive and stimulating. His expertise was invaluable in formulating the research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

My co-supervisor Dr Alex Cason, I am greatly thankful for the excellent example of being a wonderful support system to me all the time.

Finally, and most importantly, I would like to thank my mother and father, for all their love and encouragement. They were supportive, encouraging, and patient parents who raised me up with a love of science and constantly supported me in all my pursuits. I appreciate all their faithful support during the final stages of this PhD.

I gratefully acknowledge the funding sources that made my PhD work possible. I was funded by the Government of Saudi Arabia for the 4 years of my studies.

# Chapter. 1 Introduction

# 1.1 Introduction

A repeating pattern is visually observed in nature and in art. While patterns in nature are often chaotic and often fractalated, patterns in language, science or mathematics may be recognised analytically, e.g., by mathematical modelling and pattern structure prediction. Further, capturing of a repeating pattern gained a significant interest in the last two centuries, by using some form of sensing mechanism.

Sensors are an essential element to facilitate our lives "in our offices, gardens, shopping centres, homes, hospitals, cars, toys, etc" [1]. These sensors are embedded in home appliances (e.g., lights control, air conditioner, TV, fire and smoke detection) and industrial applications (e.g. in advanced micromachinery, use of magnetoelastic force and torque, chemical and gas analysis, industrial assembly and production lines). Further use of sensors are in the areas of health monitoring e.g. temperature, blood oxygen, blood glucose, imaging (X-rays, MRI, ultrasound etc.), ingestible sensors, motion, ECG, heart rate and many other non-invasive and invasive implementations. Sensors are a necessary component exist in everyday life with the purpose of detecting signals, responding to signals, or simulate signals generated by events or changes in the environment. Generally, the detected signals are in the form of "nonelectrical value or an electrical value" [2]. Information of interest are captured by these sensors and sent to other electronics to fathom a signal. Some sensors' outputs interpretation is straightforward; such outputs are signals translated to visual outputs and can be understood by the user. However, the vast amount of sensors' outputs are often complex with characteristics too diverse to comprehend by the user. Nevertheless, the rapid rise of research fields of advanced computing, wireless communication technologies and machine learning has brought progress in sensors signals analysis, with more reliable accuracy of sensors pattern analysis for a certain task.

One of the repeating patterns in nature is gait which represents the manner of walking in humans. This "walking pattern has a cyclical nature of sequential movements with both lower limbs, described as a series of transitional states" [3]. In such it is a unique behaviour trait for every human being, influenced by mutually independent factors, such as weight, gender, health condition and age. Consequently, human gait sensing and recognition gained a significant interest in providing a quantity of importance for a

plethora of tasks in many applications, e.g. healthcare [4],[5], sport [6],[7], biometrics [8],[9],[10], human robot interaction [11],[12].

## 1.2 Aims and Objectives

In this thesis the aim is to study the inconsistency of gait patterns by implementing sensor fusion and pattern recognition using machine learning, with the motivation of identifying subjects and categorizing cognitive load decline. The challenges arise from gait being characterized by a sequence of periodic events as repetitive cycles occurring naturally in all healthy humans, with small variations in these repetitive cycle events can occur owing to temporary psychological or lasting physiological condition and longer term changes can take place as a result of ageing or a long-term healthcare condition. Therefore, the difficulty lies in claiming that visual observation, harmonic analysis or Fourier decomposition of gait, or their combination, would represent adequately the variations in the dynamics of gait cycle nonlinearity and non-stationarity. Repeating gait patterns are manifested in a sequence of spatial regions and time periods, lending themselves to methods to suppress noise with the ability to generalize and optimise fundamental features. For that reason, a comparatively small number of weakly manifested features can be picked up to allow a characterization in biometrics or indicate a health condition by detecting spatiotemporal deviations from gait previously categorized as “normal” for a specific individual. Artificial neural networks, emulating 'natural' intelligence, have become a helping tool to solve many challenging problems in essential areas of industry, engineering and commercial fields. Furthermore, with the rapid rise in the capabilities of analytical computing technologies, artificial neural networks applications are growing to encompass enhanced clinical diagnosis, improve treatment of diseases and disorders with the capacity of decision-making and decision-support, whereas previously these areas were thought to be only in the reach of healthcare practitioners [13]. Substantially, in this thesis the listed points below are implemented to fill in the research gaps due to the complexity of gait.

1. Reviewing the literature and identifying the gap in the current knowledge in gait analysis. The assessment of the limitations in theories and points of view, the variety of deep learning methods applied for gait spatiotemporal analysis in literature are reviewed to establish the advances in the current knowledge on gait.

2. Formulating the rationale to apply sensor fusion as means to improve gait classification with a view to detecting individual or group differences, as well as temporal changes.
3. Building machine learning algorithms, capable of sensor fusion to map adequately the complex character of gait from data harvested by floor sensors to simple classes represented with subject identity, gender, or cognitive demanding task with fast and reliable decision, with the motivation to study the inconsistency of gait patterns which can be associated with biometric and impaired gait.
4. Building models with minimal pre-processing of complex spatiotemporal data and capable to detect gait cognitive decline to achieve accurate state-of-the-art results for disease severity staging.
5. Highlighting the gait cycle events and intervals affected in particular stages of Parkinson's disease using explainable artificial neural networks based on the outputs of machine learning.
6. Quantifying the weight of gait cycle events in the output of machine learning methods to reveal the phases most affected by cognitively demanding tasks.

The thesis is structured on two parallel cases of gait analysis. Convolutional neural network and layer wise relevance propagation methods applied for healthy gait under cognitive load, Parkinson's gait due to cognitive decline, multimodal sensors data fusion to predict gait speed. Convolutional neural networks are applied to provide solutions for gait spatiotemporal pattern recognition problems with statistical confidence in the most essential metrics to generalize and optimise fundamental features.

## 1.3 Thesis Structure

The second chapter of this thesis is a brief review of machine learning methods and computing environment, while the third chapter and further contain research outputs from published journal papers and a conference paper during the PhD program. Chapter 3 is a thorough literature review of existing work on gait to identify the gap in the current knowledge in gait analysis. It includes assessment of the limitations in theories and points of view, and the variety of deep learning methods applied for gait spatiotemporal analysis. This chapter is published as a review paper in 2019, thus a literature update is

included in the end of the chapter to cover the last two years' key advances in gait analysis. Chapter 4 to 7 are detailed in the next subsections. Finally, chapter 8 presents summary and contributions, future directions, and conclusions of this thesis.

### 1.3.1 Parkinson's disease gait

Chapter 4 studies the effect of gait deterioration due to cognitive decline. The aim is to rate the severity of Parkinson's disease (PD) and analyse the impact on the gait. The investigation is carried out by using an external public database for PD patients PhysioNet with clinically defined severity benchmarks based on ground truth diagnostic. Deep learning models are then built to process the spatiotemporal signals harvested by sensors under the foot. Additionally, LRP is used to link gait event changes to PD staging of the disease for different severity levels using the temporal domain. Further in Chapter 4, the best performing model is used for perturbation to study the inconsistency of PD gait for designing a convolutional D-ANN model resilient to noise due to the inconsistency of gait.

### 1.3.2 Cognitive Load for healthy gait

Chapter 5 and Chapter 6 present analyses of results based on gait under cognitive load through dual tasking. The aim of these chapters is to understand the relationship between cognitive load and gait disturbance. State-of-the-art deep artificial neural networks (D-ANNs) are applied on raw spatiotemporal gait sensor data captured by the in-house iMAGiMAT footstep imaging system. Dual task gait is investigated by performing experiments on healthy volunteers (under ethical approval number 2018-4881-6782) while they walked on the iMAGiMAT system and performed cognitive demanding tasks. Further deep learning models are built from scratch with extensive testing of a variety of methods and the finalised models are applied to analyse gait spatiotemporal data. Additionally, the classified gait events are analysed in the temporal domain based on ANNs backpropagation methods for explainable AI using Layer-wise Relevance Propagation (LRP). Chapter 5 is a conference article presented on the basis of gathering more data and subsequently finalized with even more data and hypotheses in Chapter 6 (journal paper). The aim of the conference paper is to build the dataset with machine learning methods and present it to the sensing and data engineering community.

Consequently, the feedbacks are used to advance the hypotheses and improve the methods as presented in chapter 6.

### 1.3.3 Multimodality fusion

Chapter 7 investigates the fusion of sensor outputs from four modalities to predict gait speed. Deep machine learning models fuse the raw sensor signal during training to make a prediction of speed under the extreme conditions, and for people never seen by the model, to simulate a real world scenario. The investigation is carried out on an external public database gathered by clinicians with background check of health issues to exclude unhealthy gait.

### 1.3.4 Publications

**Journals**

1 **Abdullah S. Alharthi**, Syed U. Yunas and K. B. Ozanyan, "Deep Learning for Monitoring of Human Gait: A Review," in IEEE Sensors Journal, vol. 19, no. 21, pp. 9575-9591, 1 Nov.1, 2019. DOI: 10.1109/JSEN.2019.2928777. **One of the top 10 popular papers in IEEE Sensor Journal for 2019 up to date.**

2 **Abdullah S. Alharthi**, Alexander J. Casson and Krikor B. Ozanyan," Gait Spatiotemporal Signal Analysis for Parkinson's Disease Detection and Severity Rating," IEEE Sensors Journal, vol. 21, no. 2, pp. 1838 - 1848, 20 Aug 2020, DOI: 10.1109/JSEN.2020.3018262.

3 **Abdullah S. Alharthi**, Alexander J. Casson and Krikor B. Ozanyan," Spatiotemporal Analysis by Deep Learning of Gait Signatures from Floor Sensors," IEEE Sensors Journal, 2021, Early access, DOI: 10.1109/JSEN.2021.3078336

**Conference papers**

4 **Abdullah S. Alharthi** and Krikor B. Ozanyan, "Deep Learning for Ground Reaction Force Data Analysis: Application to Wide-Area Floor Sensing," 28th International Symposium on Industrial Electronics (ISIE), 2019 DOI: 10.1109/ISIE.2019.8781511

5 **Abdullah S. Alharthi** and Krikor B. Ozanyan, "Deep learning and Sensor Fusion Methods for Cognitive Load Gait Difference in Males and Females," 20th International Conference on Intelligent Data Engineering and Automated Learning, 14 - 16 November 2019. DOI: 10.1007/978-3-030-33607-3_25. **Won the best student paper IDEAL 2019.**

6 **Abdullah S. Alharthi,** Syed U. Yunas and Krikor B. Ozanyan, "Sensor Fusion for Analysis of Gait under Cognitive Load: Deep Learning Approach," IEEE Sensors Applications Symposium 2020. DOI: 10.1109/SAS48726.2020.9220046

7 Syed U. Yunas, **Abdullah S. Alharthi** and Krikor B. Ozanyan, "Multi-modality fusion of floor and ambulatory sensors for gait classification", 28th International Symposium on Industrial Electronics (ISIE). 2019 DOI: 10.1109/ISIE.2019.8781127

8 Syed U. Yunas, **Abdullah S. Alharthi** and Krikor B. Ozanyan, "Multi-modality sensor fusion for gait classification using deep learning," IEEE Sensors Applications Symposium 2020. DOI: 10.1109/SAS48726.2020.9220037

9 **Abdullah S. Alharthi** and Krikor B. Ozanyan, "Multimodal Gait Spatiotemporal Data of Different Walking Speeds Fusion," IEEE Sensors conference, 2021 (Accepted in IEEEXplore Proceedings of IEEE SENSORS 2021, in production.)

**Book Chapter**

10 Omar Costilla-Reyes, Ruben Vera-Rodriguez, Abdullah S Alharthi, Syed U Yunas, and Krikor B Ozanyan, "Deep learning in gait analysis for security and healthcare," Book Chapter in "Deep Learning: Algorithms and Applications", Pedrycz W. and Chen S.-M., Eds. Springer Nature, vol. 865, pp. 299- 334, 2019. DOI: 10.1007/978-3-319-89629-8.

### 1.3.5 Contribution of authors

The contributions of the co-authors of the manuscripts presented in this thesis are as follows:

### Abdullah S Alharthi

In all manuscripts led the major contributions in: Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft and writing – review & editing. Also contributed to supervision, project administration, resources.

### Krikor Ozanyan

In all manuscripts, contributed to:
Conceptualization, formal analysis, investigation, project administration, resources, supervision, and writing – review & editing.

### Alexander J. Casson

In chapter 6 " Gait Spatiotemporal Signal Analysis for Parkinson's Disease Detection and Severity Rating" and in chapter 8 "Spatiotemporal Analysis by Deep Learning of Gait Signatures from Floor Sensors" contributed to: supervision and review & editing.

### Syed U. Yunas

In chapter 3 " Deep Learning for Monitoring of Human Gait: A Review ", In chapter 5 "Sensor Fusion for Analysis of Gait under Cognitive Load: Deep Learning Approach" contributed to: Data curation, and review & editing.

# Chapter.2 Background and Methodology

## 2.1 Introduction

In this chapter, the concepts to analyse gait are presented to aid in the understanding and comprehension of this thesis. A brief review of theoretical concepts of the widely used machine learning frameworks and the conceptual operations necessary for training, validating, and testing convolutional neural networks are presented. Layer-Wise Relevance Propagation (LRP) concepts are presented for interpretability and explainability of artificial neural networks, followed by the computing environment used to implement these algorithms.

## 2.2 Deep Learning

The supervised deep structured learning or hierarchical learning is inspired by the biological neural networks structure and function. It is based initially on the concept of multi-layer Artificial Neural Network with the aim to learn data representations automatically through training; thus, Deep Learning becomes the method of choice where the classification features, if known at all, are complex, with no straightforward quantitative relation to the raw data. Typically, ANNs consist of three types of layers sequentially connected as follow: input layer, output layer and hidden layers. The input layer receives a sample and forward it to the next layer; the output layer contains a robust representations of the input data with reduced dimensions, which are classified based on labels of samples; hidden layers play the role of down sampling the input data by discovering and transforming the important features representations to an output layer. In each layer, there are several neurons connected to adjacent layers by nodes like synapses, each neuron has a weight, non-linear activation function, and a bias term, all are optimized during training to obtain better prediction performance. There are many types of hidden layer to result in various networks structures such as Deep Belief Networks (DBN), Feedforward Deep Networks (FDN), Boltzmann Machine (BM), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long-Short Term Memory (LSTM) a special kind of RNN[14]. Further we focus on models with practical significance for gait applications such as CNN, LSTM [15].

## 2.3 Deep Convolutional Neural Networks (CNNs)

State-of-the-art CNNs have established themselves in a variety of classification tasks, providing new insights into complex data. As its name, the Deep CNNs core building block is the convolution layer with the term 'deep' refers to the number of layers. The CNNs model is suitable for processing 1-D, 2-D or 3-D data that has a known grid-like topology [16]. The network has the ability to learn high level of abstraction and features from large datasets by applying convolution operation to the input data. The CNNs are built with convolution layers, perceptron layers, pooling layers, and normalization layers, with a set of filters and weights shared among these layers. The convolutional layers output a feature map, harvested automatically from the raw input data followed by perceptron layer based on neurons to map the features to an output. Each convolutional layer followed by pooling layers to reduce the computational cost by reducing the size of representation and make the convolution layer output more robust. The CNN model uses commonly two types of pooling layers such as max pooling and average pooling. Each layer has an activation functions (e.g. Sigmoid, Tanh, ReLU, Leaky ReLU, Maxout, ELU as shown in figure 2.1), these functions define the output of that node based on calculating the weight of neurons and adding a bias, and deciding on how the weighted sum of the input is transformed into an output [17]. The convolutional neural network shown in figure 2.2 is broken down in the following sections.

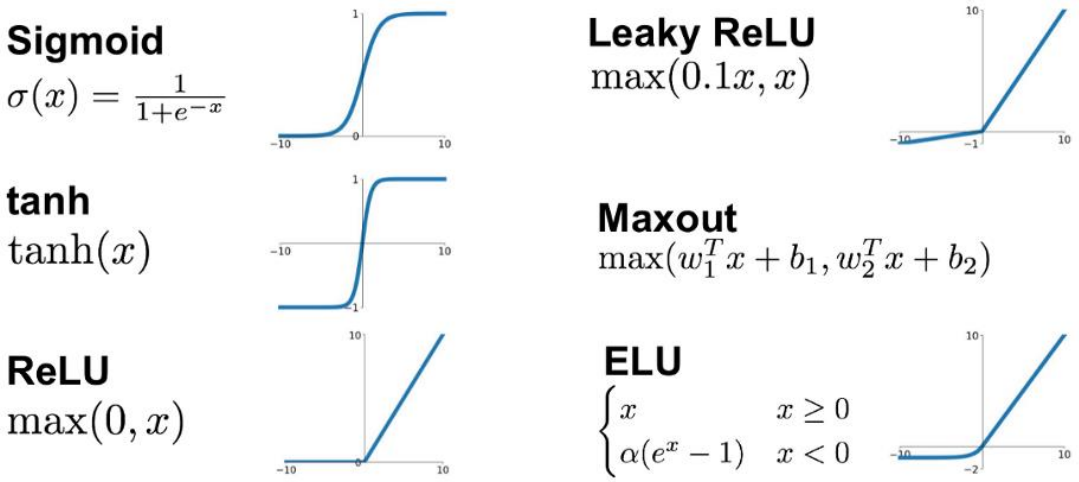



Figure 2.1. Activation functions

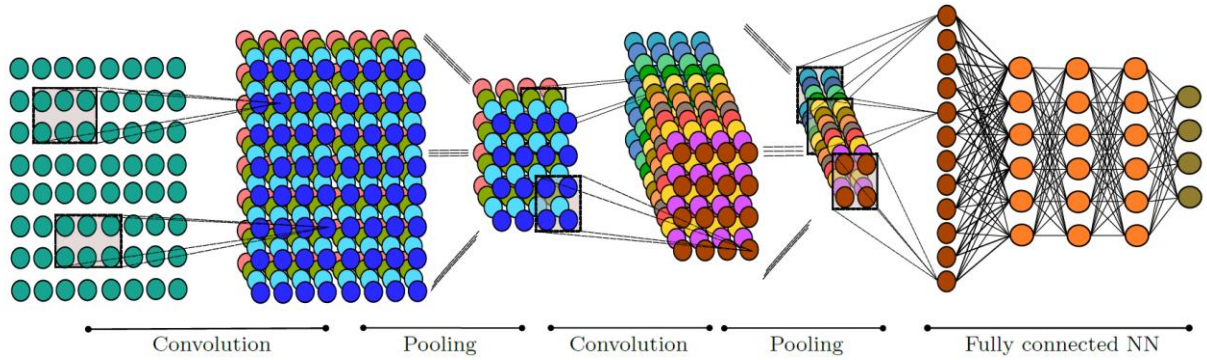



Figure 2.2. Illustration of a convolutional neural network with two-dimensional convolutional blocks and subsampling as pooling operation followed by Perceptron layer.

### 2.3.1 Perceptron layer

Figure 2.3 shows a single layer perceptron with its functionality principle. It is a Feed-forward network which is the simplest type of artificial neural networks. During training perceptron layer learns the pattern by changing connection weights, based on the amount of error in the output $\hat{y}$ compared to the actual result $y$. A mathematical representation of single layer perceptron $U$ with an input $x$ and an output $y$, a nonlinear monotonically increasing function $\sigma$, weight $\omega$ and $i$ iterator can be expressed as:

$$U = \sigma\left(\sum_i \omega_i x_i\right) \tag{2.1}$$

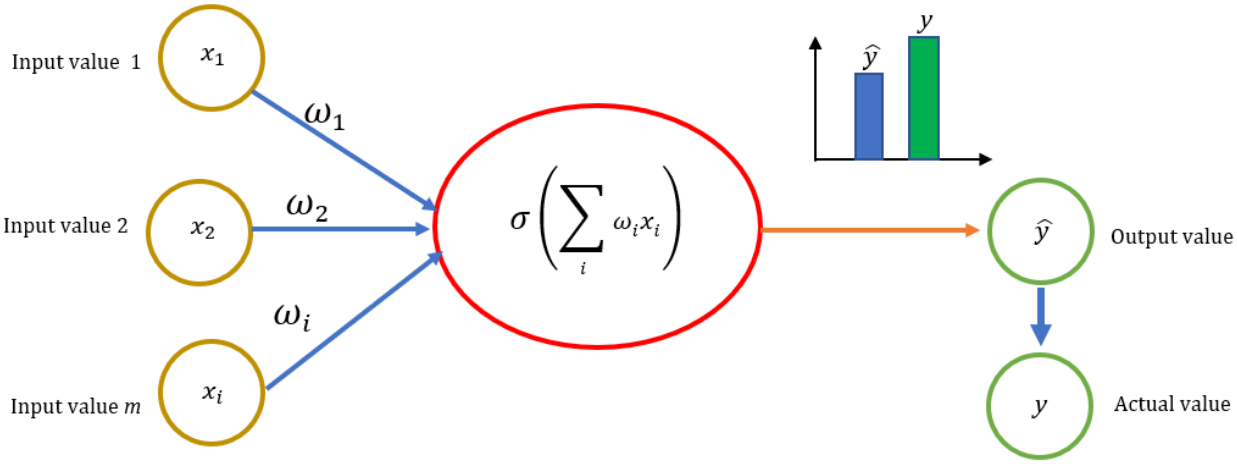
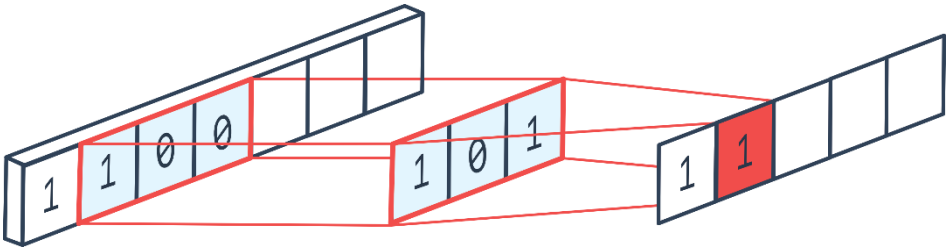
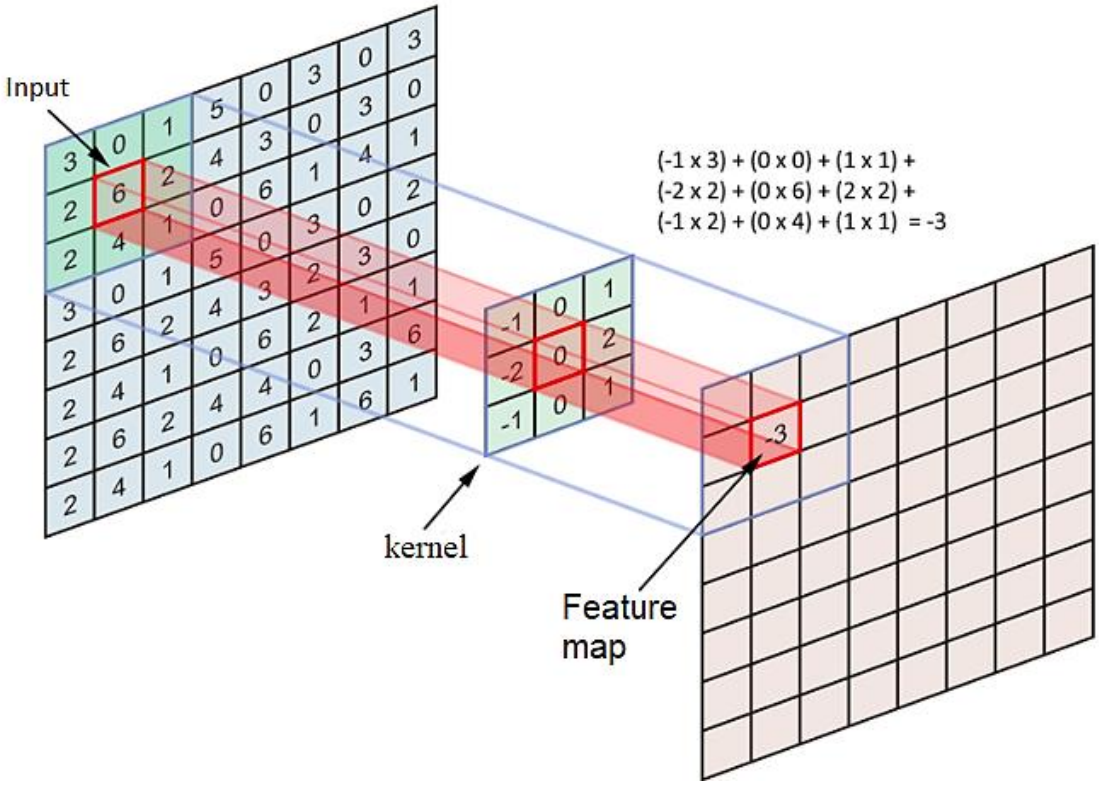



Figure 2.3. Working principle of a single layer perceptron.




(a)




(b)

Figure 2.4. Convolutional layer principle. (a) 1D convolution operation, (b) 2D convolution operation.

## 2.3.2 Convolutional Layer

A core building block in CNN model is the convolutional layer where the computational of data processing is taking place. This layer is based on convolution, a special kind of linear operation [16],[17]. Convolution operation is performed on the input data and a filter or kernel to produce a feature map as shown in figure 2.4. In this process, the filter slides over an input data and performs convolution. The convolution operations output feature maps**.** The learning occurs at the kernel by updating the kernel values during

training, similar to the weight in the perceptron layer. At the end, the feature map output consists of different feature maps produced by different kernels as convolution layer output. An activation function is utilized to produce nonlinear feature maps that can be optimized during training to pass valuable neurons values to the next layer. A mathematical representation of a convolution operation in one dimension with an input vector $x$, a kernel $\omega$ with *i, d* to denote iterators, and $(\circ)$ to denote the element-wise multiplication, can be expressed as *C(i)* with *i* is the index of an element in the new feature map (ch. 9 [16]):

$$C(i) = (\omega \circ x)[i] = \sum_d x(i-d)\,\omega(d) \tag{2.2}$$

Gait is captured as two dimensions signal as spatial and temporal as shown in figure 2.4, therefore the convolution operation in eq. 2.2 can be extended to two dimensions. Such that the spatiotemporal input is a large set of data points, and the kernel is a set of data smaller in size than the input. Then the convolution operation slides the kernel over the input and computes element-wise multiplication and add the values in a smaller future map as shown in figure 2.4. With a 2-D input $x$ and a 2-D kernel $\omega$ with $(i,j)$, $(d,k)$ are iterators, the mathematical representation of a convolution in two dimensions can expressed as $C(i,j)$ with $(i,j)$ is the index of an element in the new feature map (ch.9 [16]):

$$C(i,j) = (\omega \circ x)[i,j] = \sum_d \sum_k x(i-d, j-k)\omega(d,k) \tag{2.3}$$

### 2.3.3 Pooling Layer

There are two types of pooling layers, max pooling and average pooling. The objective of using this layer is to recombine the convolutional layer output and to reduce the network parameters which reduces the amount of computation in the network. In pooling layer, a filter slides over the convolutional layer output and the maximum or average value in the filter window are transformed as an element in an output matrix as pooling layer output.

### 2.3.4 Activation function

The output of a layer in a neural network is the input to the next layer and if the activation functions are linear in these layers then the network stacked layers are linear. Therefore, the network will only be able to learn linear relationship among data, while adding

nonlinear activation function is advantageous in making the network decision boundary non-linear. The aim is to make the neural network able to adapt to the nonlinear characteristics in real world pattern, by making the network more complex.

The widely used activation function in convolutional layer is Rectified Linear Units (ReLU) to only pass positive values, defined as (max(0,x)). The ReLU function becomes the method of choice to avoid vanishing gradients problems in the network, where the parameters of an earlier layer near the input are not tuning and learning, it can result in stopping the neural network from further training.

## 2.4 Training and Testing

In gait analysis, the goal of using neural networks is to develop a model to extract gait features and perform well on unseen real-world gait data. For appropriate training and testing, commonly the model is trained and validated on 70%-80% of the data for a number of epoch and tested on the remaining 30%-20%. In supervised training, the procedure is launched by initializing the weights randomly to learn a function $h(x)$ (called hypothesis) to a desired output function $f(x)$ to map the input $x$ to a desired output $y$. The model is trained on a batch of data for a number of iterations and validated on a batch of data to identify a performance measure $P$ in each epoch. $P$ is optimized by reducing the prediction error of $h(x)$ in mapping the input $x$ to the output $y$, which is expressed by the cost function $J(\theta)$. Here $J$ is the cost and $\theta$ is the parameters $\omega$, $b$, with other learnable parameters that can be adjusted during training with $b$ as the bias term.

### 2.4.1 Forward propagation

The cost function is approximated by using forward propagation. This allows the calculation and storage of intermediate variables (including outputs) for a neural network in consecutive order from the input layer to the output layer. The optimization function is used to minimize the cost function with the weights $\omega$ of $\theta$. Therefore, the error in our hypothesis $h(x)$ for a given set of parameters $\theta$ is found using the generalized cost function. The cost function or error function are the sum of loss functions, while the loss functions are defined based on predicting a data point to a label with error.

The widely used cost function for supervised classification problem is binary cross entropy. Essentially neural networks models are trained to learn to predict data D = $\{(x_1, y_1), (x_2, y_2), \dots (x_M, y_M)\}$ to binary labels *y = 0 or 1*, these labels can be categorical for multi-class (one-hot encoded label, such that label 3 in 4 classes problem would be [0 0 1 0]). In this thesis we used categorical cross entropy because gait is a multi-class problem. The binary cross entropy loss ($J_{bce}$) formula can be expressed as follows ([19], ch. 8 [16]):

$$J_{bce} = -\frac{1}{M}\sum_{m=1}^{M}[y_m \log (h_\theta (x_m)) + (1 - y_m) \log(1 - h_\theta (x_m))] \quad (2.4)$$

To calculate the cost function of the binary cross entropy, we need to average the cross entropy across all M data examples to make the loss function comparable across different size datasets. The categorical cross-entropy loss ($J_{cce}$), with $K$ to denote the number of classes and $y_m^k$ the target label for training example $m$ for class $k$, is expressed as follows ([19], ch. 8 [16]):

$$J_{cce} = -\frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M} y_m^k \log (h_\theta (x_m, k)) \quad (2.5)$$

### 2.4.2 Backpropagation

It is short for "backward propagation of errors", it is an algorithm based on gradient descent. The method moves in a reverse order from the output layer to the input layer while calculating the gradient of the error function based on the network weights, the aim is to minimize *J (θ)* using an optimal set of parameters in *θ.* It is based on performing the partial derivative to minimize the cost function. The partial derivative is expressed as $\frac{\partial}{\partial \theta_{i,j}^{l}} J(\theta)$. The output layer calculates the error of the network layers $L$ with: $ð^{(L)} = \alpha^{(l)} - y$, such that the error of node $j$ in layer $l$ is denoted as $ð_j^{(l)}$ and the activation of node $j$ of layer $l$ is denoted as $\alpha_j^{(l)}$ and $y$ is the output of the output layer, then the backpropagation can be expressed for neural networks as:

$$ð^{(L)} = ((\theta^{(l)})^{(T)} ð^{(l+1)}) \circ \alpha^{(l)} \circ (1 - \alpha^{(l)}) \quad (2.6)$$

Here the ð values of the output layer $L$ are calculated by multiplying the ð values in the next layer (in reverse direction) with the $\theta$ matrix of layer l, hence T denote matrix. We then perform element-wise multiply ($\circ$) with the $g'$ which is the derivative of the

activation function, which is evaluated with the input values given by $z^{(l)}$. Where $g'\left(z^{(l)}\right) = \alpha^{(l)} \circ (1 - \alpha^{(l)})$.

The partial derivatives needed for backpropagation is performed by multiplying the activation values and the error values for each training example *t* and *m* is the number of training data as:

$$\frac{\partial}{\partial \theta_{i,j}^{l}} J\,(\theta) = \frac{1}{m} \left[ \sum_{t=1}^{m} \alpha_j^{(t)(l)} \eth_j^{(t)(l+1)} \right] \tag{2.7}$$

## 2.5 Regularization Methods

The main objectives of using machine learning are to make the models perform well on unseen data; thus, models experience underfitting or overfitting which leads to poor accuracy. The underfitting is when the model performs poorly during the learning process on the training data and the validation data. The overfitting is when the model performs well on the training data and poorly on validation data in the learning process (See figure 2.5).

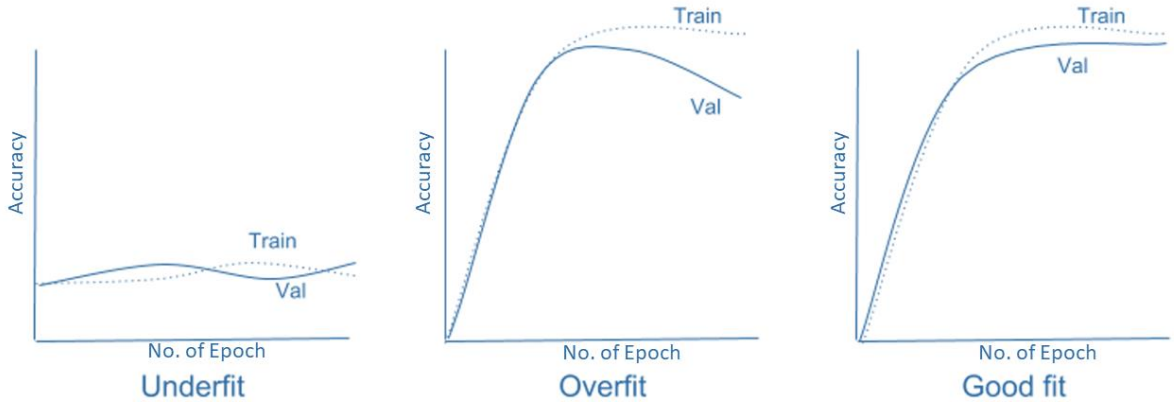



Figure 2.5. Machine learning training and validation learning curves.

These two problems can be triggered by defects in the data or the network architecture; however, the research on machine learning partially solved the problem by introducing a regularization method. There are a number of regularization methods e.g. weight regularization, activity regularization, weight constraint, early stopping, adding noise to the data, and dropout [15]. Deep neural networks comprise a large number of neurons and learning it all together is a challenging task, because few neurons have a strong

predictive capability while many neurons have a weaker connections. Therefore, during training few of the strong neurons (depends on the dropout factor) are ignored while the weaker connections are learned more in each epoch, such that the network consist of weaker neurons learned during training and optimized in validation for better prediction. This method is defined as 'dropout' and it has been utilized in this thesis to prevent overfitting. Figure 2.6 shows a neural network layer with dropout and a layer with all units connected.

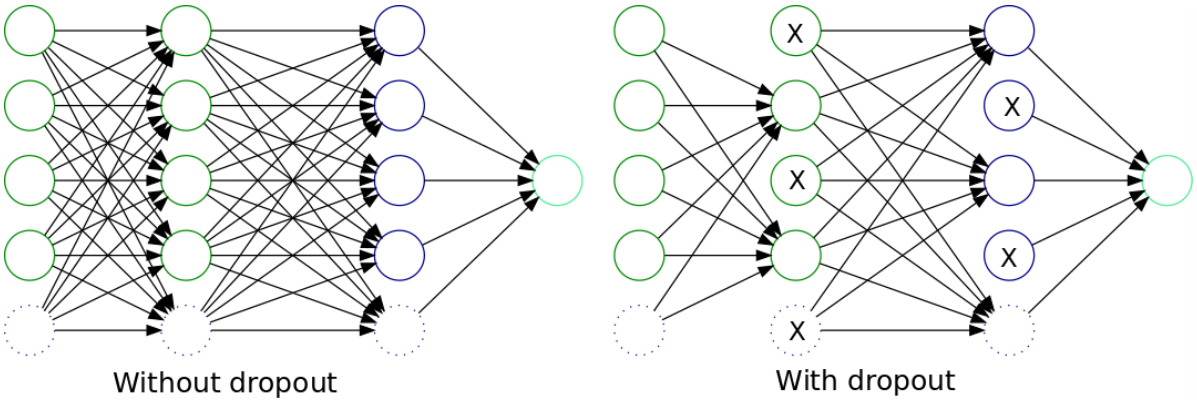



Figure 2.6. Dropout method, a) original layer units; b) layer after dropping units.

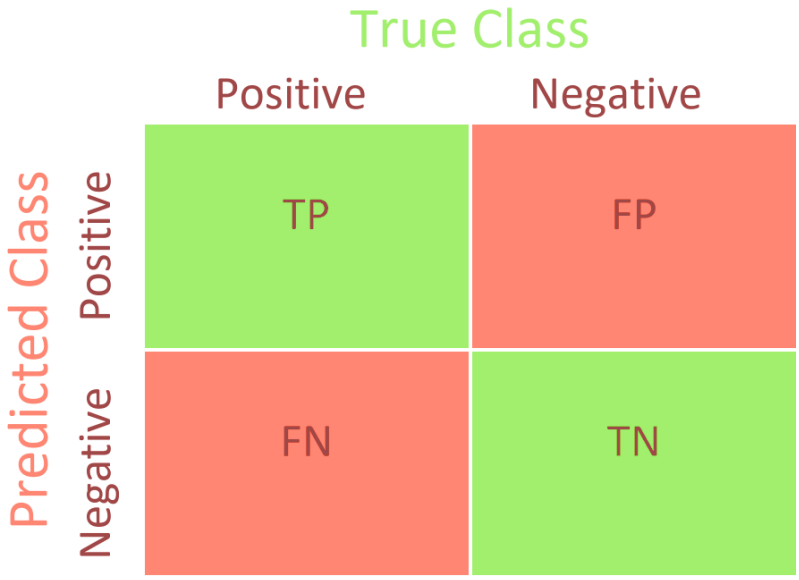



Figure 2.7. Confusion matrix table.

## 2.6 Evaluation Measure

The widely used accuracy measure for gait analysis is the confusion matrix [18]. It is a table to visualize the number of predictions classified correctly and wrongly for each class. The table consists of true positive, true negative, false positive, and false negative classification occurrences. One of the advantages of the confusion matrix display is that it is straightforward to identify the decision confusions, thus possibly concluding on the quality of the data involved (See figure 2.7). It shows each class prediction as follows:

- **True positive,** TP**:** It is the number of positive class correctly predicted as positive.
- **True negative,** TN**:** It is the number of negative class correctly predicted as negative.
- **False positive,** FP**:** It is the number of negative class incorrectly predicted as positive.
- **False negative,** FN**:** It is the number of positive class incorrectly predicted as negative.

From this confusion matrix table, the number of predictions classified correctly and wrongly are used to calculate different rates of measure to evaluate the performance of a machine learning model. Performance measures, such as accuracy, recall, precision and F1 values of a model can be defined using the following equations.

**Accuracy**: indicator of the ratio between the correctly predicted data to total number of samples in the dataset, defined as:

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{2.8}$$

**Recall**: the proportion of positive classes identified correctly, defined as:

$$\frac{TP}{TP+FN} \tag{2.9}$$

**Precision**: the fraction of positive cases correctly identified over all the positive cases predicted, defined as:

$$\frac{TP}{TP+FP} \tag{2.11}$$

**F1 Score**: the harmonic mean of Precision and Recall, defined as:

$$\frac{2*Precision*Recall}{Precision+Recall} \tag{2.12}$$

There are popular evaluation measure used for classification problems such as Area Under the Curve (AUC) and Receiver Operating Characteristic (ROC). In this thesis we use the confusion matrix over the Area Under the Curve (AUC) because the number of TP,TN,FP and FN samples are values of interest to understand the confusion in gait classes for further analysis using LRP.

## 2.7 Long Short-Term Memory Networks

The LSTM is a type of recurrent neural network, and it is favourable for processing time series data, where the order is of importance. Since the gait cycle is time related cue, the LSTM approach is found to be quite useful in gait data analysis. In essence, a spatialized neural network exploits recurrence, which is using information from a previous forward pass over the neural network. The network produces an output for every entity in the input using its internal memory, in this process the output is determined by the cell state based on both the current input and the previous output for a sequence of data[19]. The networks have memory blocks instead of neurons layers, the blocks are connected in each layer. Each block contains Forget Gate, Input Gate, and Output Gate (See Figure 2.8).

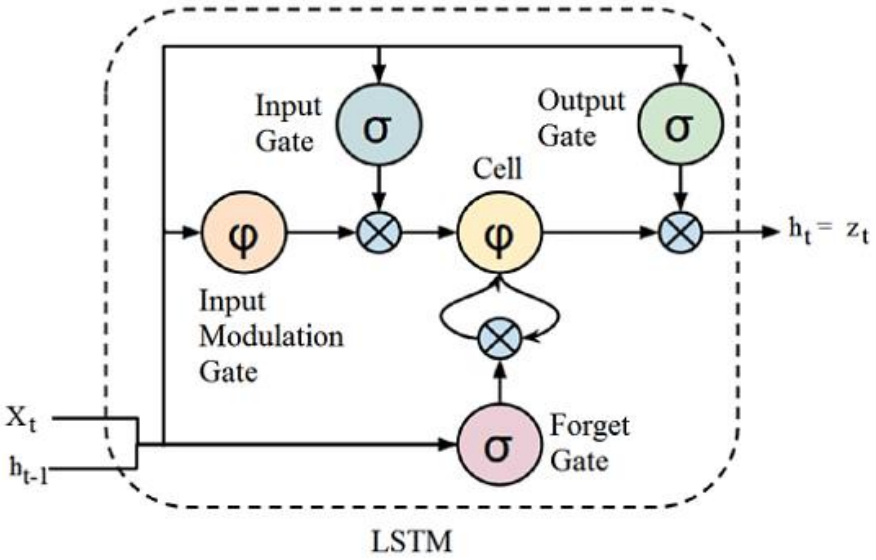



Figure 2.8. LSTM Memory block or cell.

## 2.8 Layer-Wise Relevance Propagation (LRP)

LRP [20], [21] [22] is a backward propagation method which identifies which parts of the ANN input vector carry most weight in the model prediction. In this thesis we quantify the contribution of a single component of an input $x_i$ (in our case, a sensor signal at a specific time frame) to the prediction of $f_c(x)$ ( $c$ denote a class of gait) made by the DCNN classifier $f$. The outputted gait class prediction is redistributed to each intermediate node via backpropagation until the input layer. The LRP outputs a "heat map" over the original signal to highlight the signal sections with the highest contributions to the model prediction, e.g. the data sections with the maximum variability given the classes. We first note that a neural network consists of multiple layers of neurons (feature maps in the case of a convolution layer), where neurons are activated as follows [20]:

$$a_k = \sigma(\textstyle\sum_j a_j \omega_{jk} + b_k) \tag{2.13}$$

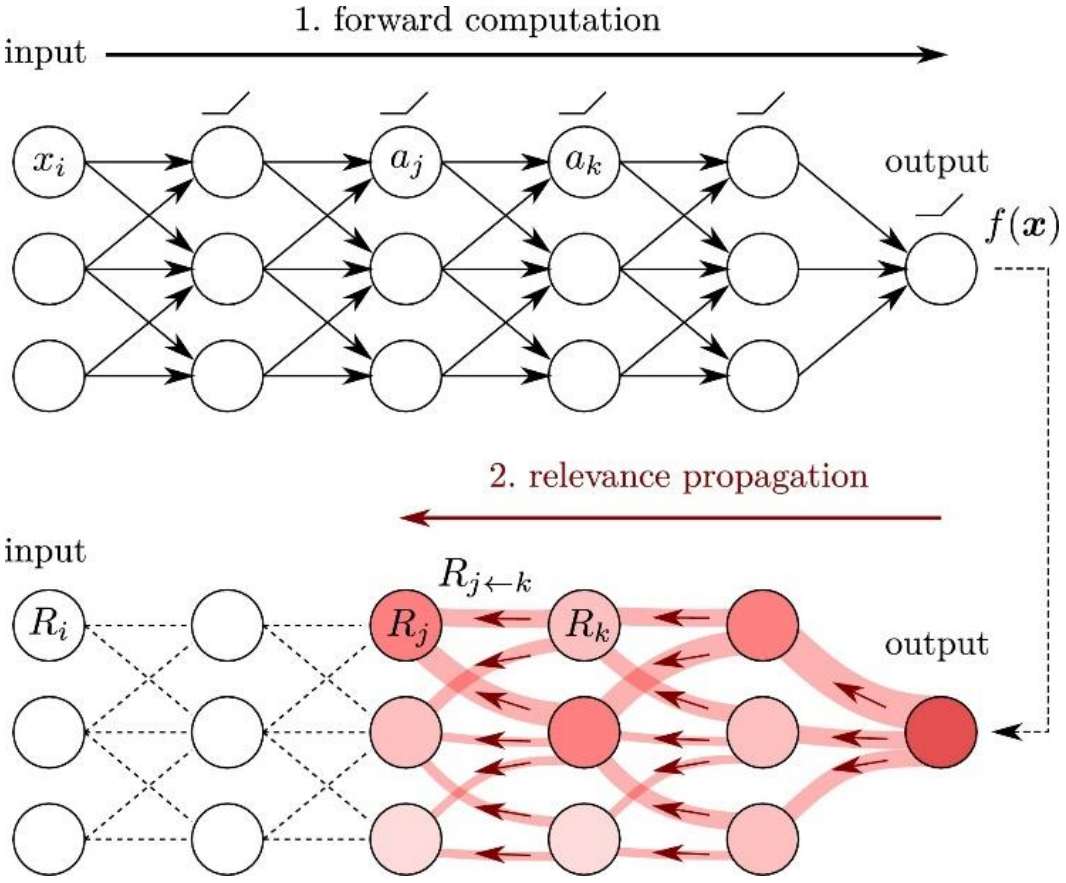



Figure 2.9. DNN and LRP signal processing flow. Red arrows indicate the relevance propagation flow [20].

Here, $a_k$ the neuron activation and $a_j$ is the activation of the neuron in the previous layer in forward direction; $\omega_{jk}$ denote the weight received in forward direction by neuron $k$ from neuron $j$ in the previous layer and $b_k$ is the bias. The sum is computed over all the $j^{\text{th}}$ neurons that are connected to $k^{\text{th}}$ neuron. $\sigma$ is a nonlinear monotonically increasing activation function. These activations, weights, and biases are learned by the DCNN during supervisory training. During training, the output $f_c(x)$ is evaluated in a forward pass and the parameters ($\omega_{jk} + b_k$) are updated by back-propagating using model error. For the latter, we base our computations on categorical cross entropy [23].

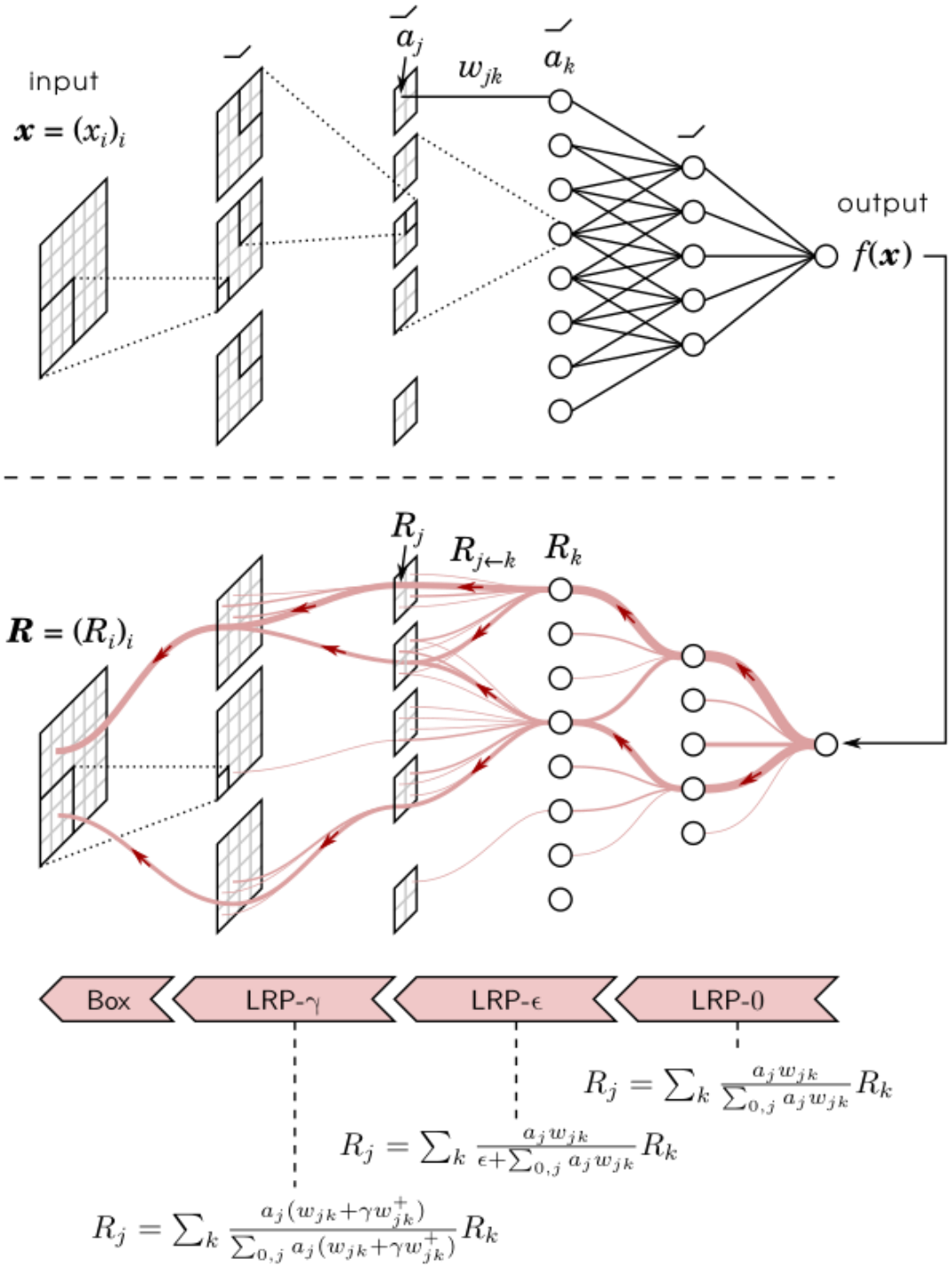



Figure 2.10. Illustration of the LRP propagation procedure applied to a neural network. The prediction at the output is propagated backward in the network, using various propagation rules, until the input features are reached. The propagation flow is shown in re[24].

The LRP approach decomposes the DCNN output for a given prediction function of gait class $c$ as $f_c$ for input $x_i$ and generates a "relevance score" $R$ for the $i^{th}$ neuron received

from $R_j$ for the $j^{th}$ neuron in the previous layer which is received from $R_k$, for the $k^{th}$ neuron in the lower layer (see figure 2.9), where the relevance conservation principle is satisfied as:

$$\sum_i R_{i\leftarrow j} = \sum_j R_{j\leftarrow k} = \sum_k R_k = f_c(x) \tag{2.14}$$

The LRP starts at the DCNN output layer after removing the *Softmax* layer. In this process, a gait class $c$ is selected as an input to LRP and the other classes are eliminated. The backpropagation for unpooling for the pooling layer is computed by redirecting the signal to the neuron for which the activation was computed in the forward pass. As a generalization, consider a single output neuron $i$ in one of the model layers, which receives a relevance score $R_j$ from an lower layer neuron *j,* or the output of the model (class $c$). The scores are redistributed between the connected neurons throughout the network layers, based on the contribution of the input signals $x_i$ using the activation function (computed in the forward pass and updated by back-propagating during training) of neuron $j$ as shown in figure 2.9. The latter will hold a certain relevance score based on its activation function and passes its value to consecutive neurons in the reverse direction. Finally, the method outputs relevance scores for each sensor signal at a specific time frame. These scores represent a heat map, where the high relevance scores at specific time frames highlight the areas that contributed the most to the model classifications. There are other propagation rules such as (αβ-rule) [20]

$$R_j=\sum_k(\alpha \frac{a_j\omega_{jk}{}^+}{\sum_j a_j\omega_{jk}{}^+} - \beta \frac{a_j\omega_{jk}{}^-}{\sum_j a_j\omega_{jk}{}^-})R_k \tag{2.15}$$

Where each sum corresponds to $R_{j\leftarrow k}$ a relevance message and $a_j\boldsymbol{\omega}_{jk}{}^+$ and $a_j\boldsymbol{\omega}_{jk}{}^-$ denote the positive and negative part of $a_j\ \omega_{jk}$ respectively. The parameters α and β are chosen so that $\alpha - \beta = 1$ and $\beta \geq 0$. A propagation rule can be chosen by selecting $\beta = 0$ to in result the following rule :

$$R_j=\sum_k \frac{a_j\omega_{jk}{}^+}{\sum_j a_j\omega_{jk}{}^+} R_k \tag{2.16}$$

There are other stabilizing terms can be used to avoid divisions by zero as explained in [20],[22]. For the LRP-$\gamma$ rule, let the neurons interconnection be as follow [24]:

$$a_k=\max\left(0, \sum_{0,j} a_j\ \omega_{jk}\right) \tag{2.17}$$

Here, $a_j$ denote input activation and $\omega_{jk}$ denote the weight received by neuron *j* from neuron *k* in the above layer. The sum is computed over all neurons $j$ in the lower layer plus a bias term $\omega_{0k}$ with $a_0$ = 1. The LRP-$\gamma$ rule as shown in figure 2.10 is given by:

$$R_j = \sum_k \frac{a_j(\omega_{jk}+\gamma\omega_{jk}^+)}{\sum_{0,j} a_j(\omega_{jk}+\gamma\omega_{jk}^+)} R_k \quad (2.18)$$

# 2.9 Python

The mathematical calculations of a neural network are impractical to perform by humans to recognize patterns for a certain application. However, Machines are capable of computing anything that is computable by executing instructions of sequences of arithmetic or logical operations automatically. Machine languages are established to enable humans to program these machines to perform certain tasks, e.g. MATLAB, C, C++, C#, FORTRAN, HTML, Javascript, Python and many others.

Python is an interpreted high-level general-purpose programming language used by many applications, e.g. machine learning computing. Python was designed with high capacity and capability to handle the implementation of complex mathematical models to process data and algorithm for training a structure defined by its nodes and connections.

In this thesis, Python was adopted as the programming language of choice for gait data processing and neural networks implementation. Python has many libraries containing useful functions to eliminate the need for writing codes from scratch. For this thesis, the neural networks implemented mainly using Scikit-learn, NumPy, TensorFlow, Keras, Pandas, Scipy, Matplotlib, other packages used for this project included in the Anaconda (Python distribution). All our codes are computed in Python using Anaconda (package management and deployment software) with Spyder (open-source cross-platform integrated development environment software) as shown in figure 2.11.

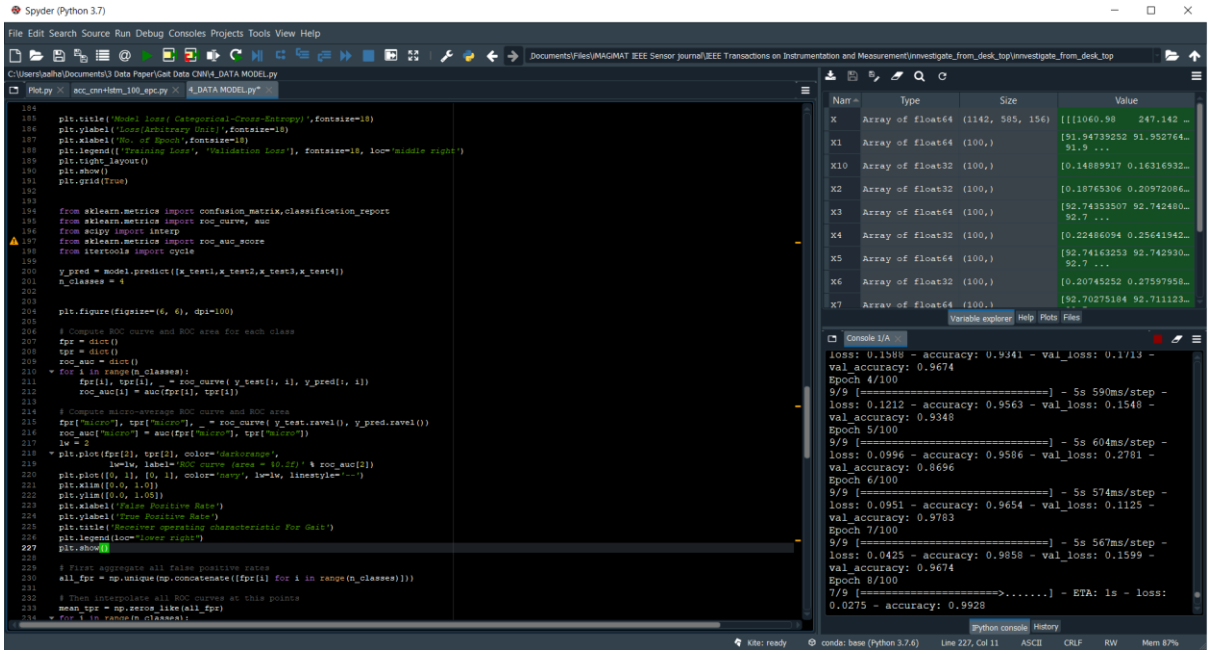



Figure 2.11. Anaconda Spyder environment.

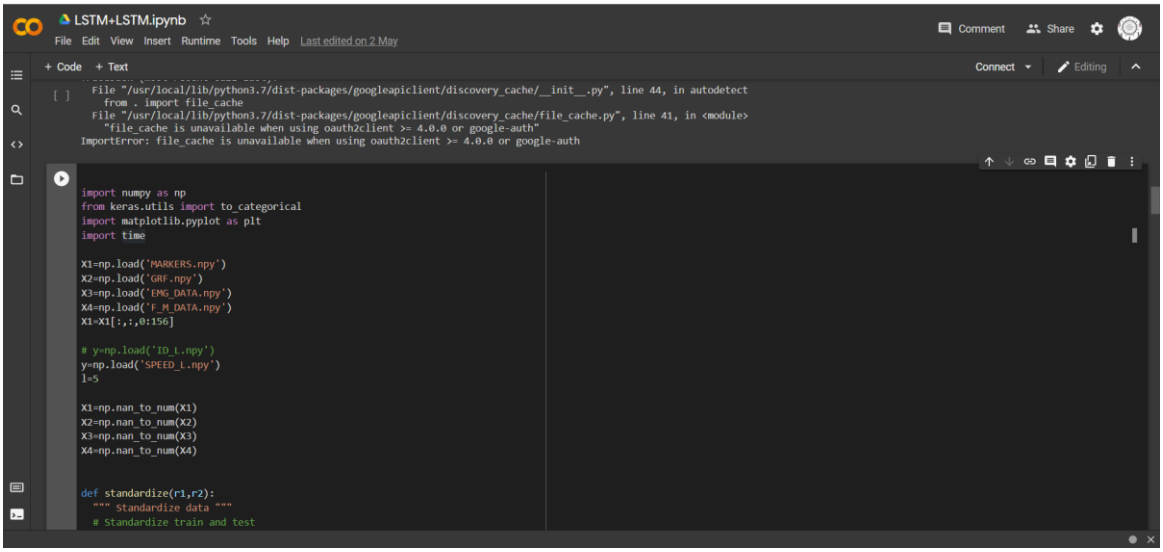



Figure 2.12. Google Colaboratory notebook.

Further, we used Google Colaboratory notebook to execute python codes as shown in figure 2.12. The advantages of using the Colaboratory is the computing resources, such that codes are executed for less time due to the use of Graphics Processing Unit and the Tensor Processing Unit (TPU) in Google Colaboratory. Tensor Processing Unit is an AI

accelerator application-specific integrated circuit (ASIC) developed by Google specifically for neural network and machine learning.

# Chapter.3 Literature Review

# Deep Learning for Monitoring of Human Gait: A Review

**Abdullah S. Alharthi, Syed U. Yunas, and Krikor B. Ozanyan**

## Abstract

The essential human gait parameters are briefly reviewed, followed by a detailed review of the state of the art in deep learning for the human gait analysis. The modalities for capturing the gait data are grouped according to the sensing technology: video sequences, wearable sensors, and floor sensors, as well as the publicly available datasets. The established artificial neural network architectures for deep learning are reviewed for each group, and their performance are compared with particular emphasis on the spatiotemporal character of gait data and the motivation for multi-sensor, multi-modality fusion. It is shown that by most of the essential metrics, deep learning convolutional neural networks typically outperform statistical analysis models. In the light of the discussed character of gait data, this is attributed to the possibility to extract the gait features automatically in deep learning as opposed to the statistical analysis from the handcrafted gait features.

Manuscript published at the IEEE Sensors Journal. Received March 26, 2019; revised July 10, 2019; accepted July 11, 2019. Date of publication July 15, 2019; date of current version October 4, 2019. The work of A. S. Alharthi was supported by the Government of Saudi Arabia. The work of S. U. Yunas was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC DTP) under Grant EPN509565/1. The associate editor coordinating the review of this paper and approving it for publication was Dr. Marco J. Da Silva. (Corresponding author: Abdullah S. Alharthi.) The authors are with the department of Electrical and Electronic Engineering, The University of Manchester, Manchester M1 3BU, U.K. (e-mail: abdullah.alharthi@manchester.ac.uk; syed.yunas@manchester.ac.uk; k.ozanyan@manchester.ac.uk). Digital Object Identifier 10.1109/JSEN.2019.2928777

## 3.1 Introduction

Gait refers to the displacement of the centre of gravity during locomotion. In humans, it is achieved through the synchronised movement of the lower limbs and the trunk, resulting in a move from one position to the other [3]. It is a unique behaviour trait for every human being, influenced by mutually independent factors, such as weight, gender and age.

The rich history of gait analysis is a record of a steady progression from descriptive studies to more sophisticated methods. Aristotle (350 BC) was the first to take note of animals and human gait [25]. However, useful descriptions of how humans walk were first achieved in the works of Newton, Galileo and Leonardo da Vinci. Borelli, a student of Galileo and the father of biomechanics [26], gave a considerable impetus to scientific approaches to gait analysis by measuring the centre of gravity of the human body and how humans keep balance while walking [27]. In 1836, the Weber brothers described gait as a periodic movement and defined the gait cycle on the basis of the pendulum-like forward leg motion [28]. In 1878, Muybridge used 12 cameras to capture racehorse gait to prove that all four horse hooves were off the ground while trotting. He also used a similar approach to capture a series of photographs of human movement [29]. The first substantial quantitative use of gait analysis was in 1895 [27] when Braune and Fisher used a photographic technique to determine a human body's velocity, acceleration, and dimensional trajectory to estimate the forces involved during the gait cycle. In 1930s, Bernstein studied the dynamic locomotion of 150 subjects to determine the centre of gravity of each limb segment of the subjects using a photographic technique [30].

Ground Reaction Force (GRF) was introduced in human gait understanding in 1924 when Amar [31] designed a force plate to measure the magnitude and the direction of GRF. The platform was improved by Elftman in 1938 using a high-speed cinematic camera to capture a pointer movement resulting from the force applied to the platform [32]. A substantial amount of knowledge was contributed to the human locomotion analysis in the 1950s, with the motivation to treat World War II veterans [33].

In the past two decades, the rapid rise in the capabilities of sensor systems involving analytical computing technologies has allowed the extraction of richer information from an increasing number of sensing modalities. In this context, developments in new gait-sensing instrumentation have underpinned the progress in the evaluation of different human locomotion parameters based on an ever-increasing volume and quality of data. Understandably, this has also raised awareness of challenges brought forward by the necessity to achieve multi-source, multi-sensor fusion from big data with diverse characteristics. Furthermore, it is unclear whether the complex character of gait maps adequately onto simple and widely used measurands, typically delivered by systems for fast and reliable diagnostics, recognition and classification. However, progress in machine learning technology has resulted in deep learning models that can be applied with minimal pre-processing on complex data and are capable of faster, more accurate results from databases that are constantly growing in volume and range. It presents new opportunities for detection, fusion and classification from different multi-source, multi-sensor data. Among these, gait spatiotemporal parameters are currently attracting attention due to the possibility of using such information in a variety of applications, e.g. healthcare [4],[5], sport [6],[7], and identification of individuals for security [8],[9].

Gait analysis is still on its way to maturity, and there is no gold standard sensing or data processing method. Further in this Review, we organise the modalities mostly used to study human gait into three groups, based on the sensing principle as well as the amount and character of the generated sensor data: video sequence (VS), wearable sensors (WS), and floor sensors (FS). We show that the sensing principle used for this grouping also shapes the choice of deep learning processing methodology: the VS solutions are based on action recognition using spatiotemporal information; WS systems typically comprise inertial sensors to acquire human body velocity, acceleration and orientation during physical human activity; FS characteristically monitor the GRF induced by floor contact during the gait cycle. The data captured from these modalities is analysed and classified using sophisticated supervised learning methods, based on appropriate assumptions.

This review is underpinned by an extensive literature search but only the most recent works, combining gait recognition with deep learning algorithms, are presented in more detail.

## 3.2 Background

To outline the contribution of deep learning in human gait analysis, it is necessary to understand how humans walk and the applications of gait giving rise to the set of methods utilized for analysis.

### 3.2.1 Gait Parameters

Gait can be perceived as a transformation of a brain activity to muscle contraction patterns resulting in a walking sequence. It is a chain of commands generated in the brain and transmitted through the spinal cord to activate the lower neural center, which will consequently result in muscle contraction patterns assisted by sensory feedback from joints, muscles and other receptors to control the movements. This will result in the feet recurrently contacting the ground surface to move the trunk and lower limbs in a coordinated way, delivering a change in the body center-of-mass position.

Gait is a sequence of periodic events characterized as repetitive cycles for each foot [27]. Each cycle is divided into two phases (see figure 3.1):

a) Stance Phase (approximately 60% of the gait cycle, with the foot in contact with the ground). This phase is subdivided into four intervals (**A, B, C, D**).

b) Swing Phase (approximately 40% of the gait cycle with the foot swinging and not in contact with the ground). This phase is subdivided into three intervals (**E, F, G**).

**Stance Phase:**

- **A.** Heel strike or Initial contact: It starts the moment the foot touches the ground, and it is the initial double-limb support interval. In the case of the right foot leading, the double support starts with left foot being on the ground when the right foot heel makes initial contact and finishes when the left foot leaves the ground with the left toe-off prepared to swing. At the end of this interval, the body weight is completely shifted onto the stance (leading) limb. This term is adopted in clinical psychology to denote the contact of the heel of the extended limb with the walking surface.
- **B.** Loading response or Foot flat: This is a single support interval following the initial double support interval. The bodyweight is transferred on to the supporting limb.

The trunk is at its lowest position, the knee is flexed, and a plantarflexion occurs at the ankle.

**C.** Mid-stance: This is a single support interval between opposite toe-off and heel-off. It starts from elevation of opposite limb until both ankles are aligned in coronal plane. The trunk is in its highest point and slowing its forward speed. The body center-of-mass is aligned with the forefoot (ball of the foot).

**D.** Terminal stance or Heel-off: It begins when the supporting heel rises from the ground in preparation for opposite swing. The trunk is sinking from its highest point, the knee has extant peak near the time of heel rise and ankle has dorsiflexion after heel rise.

**Swing Phase:**

**E.** Pre-swing: This is the second double-limb support interval. The opposite initial contact occurs, and the hip is beginning to flex, the knee is flexing, and the ankle is at plantarflexion. The toe is in last contact before the swing, finishing the push-off started in interval **D.**

**F.** Initial swing and Mid-swing: This interval begins with the toe-off into single support and starting to swing. The body weight is shifted to the opposite forefoot. In this instant, the knee joint gets the maximum flexion. The hip is flexing and the limb advances in preparation for a stride.

**G.** Terminal swing: This is the last interval of gait cycle and the end of the swing phase. The interval begins at maximum knee flexion and ends with maximum extension of the swinging limb forward. The hip continues flexion and the knee extends in regard to gravity, the ankle continues dorsiflexion to end neutral, ready for the heel strike.

With regard to the above gait events, the following parameters of human gait are usually analysed in clinical settings [34] for healthcare tasks, using various sensing and data processing methods:

- Cadence or rhythm (number of steps per unit time)
- Stride length
- Velocity

- Direction of leg segments
- Step angle
- Swing time for each foot
- Step width
- Support time
- Ground Reaction Force (GRF)
- Electrical activity produced by muscles
- Momentum and forces
- Body posture

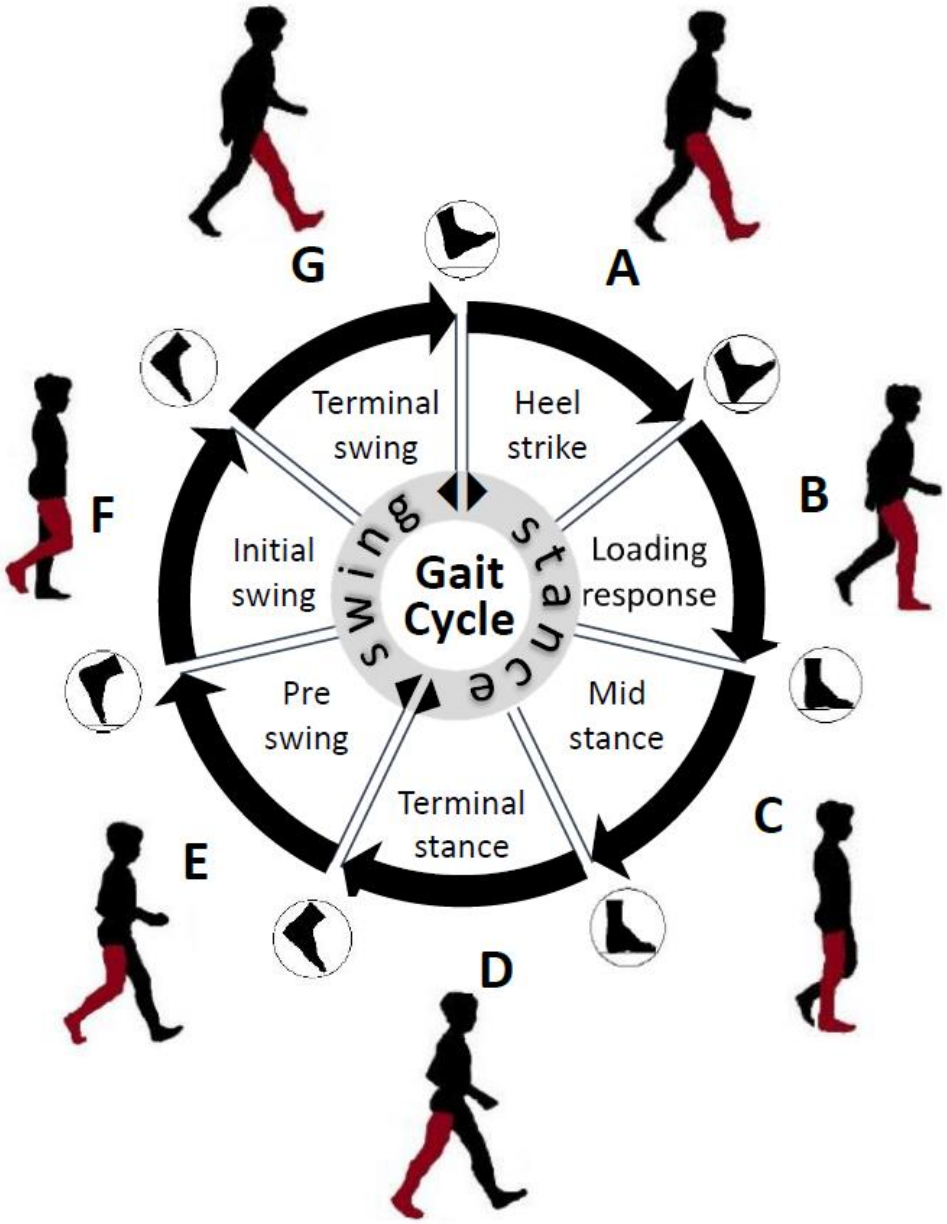



Figure 3.1. Important gait events and intervals in a normal gait cycle.

It is worth mentioning at this point, that while the listed parameters have clear observational value, it is difficult to claim that any of these, or their combination, would represent the maximum variability of the raw data due to a health condition. This difficulty has a direct impact on the ability to detect, with the lowest threshold affordable by the raw data quality, a meaningful deviations from the norm

### 3.2.2 Applications of Gait Analysis

The field of research in human gait is broad, with many specific applications. In medical applications, as gait abnormality affects a high percentage of the population, gait is studied to diagnose neurodegenerative diseases such as Parkinson's disease (PD), myelopathies, spinal amyotrophy, multiple sclerosis, cerebellar ataxia, brain tumors, cranioencephalic trauma, certain types of dementia, neuromuscular diseases etc. [34]. In fact, the ground reaction force of individuals during the gait cycle has been used to detect PD in [35]. The study shows that stance time, swing time, stride time and foot strike profiles can be used to distinguish PD patients from healthy controls. In addition, the spatiotemporal parameters of gait have been studied [36] to assess lower limb prosthesis users.

In security applications, gait analysis as a biometric has proven its success to distinguish and identify people, with minimum cooperation required from the subject. The aim is to identify individuals from a distant based on their walking habit. Typically, individuals' gait is captured by CCTV cameras as reported in [37]-[38]. In [39],[40], the ground reaction force has been found to be significant in identifying subjects based on their footstep signals and stepping behavior.

Injuries commonly occur during sports activity and some methods to evaluate athletes' recovery are based on gait, e.g. by analyzing forces exerted on each muscle through electromyography in [41]. The kinematic parameters of gait are used to analyse various indoor and outdoor activities, such as sports training and clinical rehabilitation of patients using a wearable sensor [6]. Even different gait characteristics assessment methods are used to assess athletes' ability to return to sport after surgery due to tear in the anterior cruciate ligament which causes knee instability [42]. Further, the gait dual-task paradigm for comprehensive athlete evaluation following a sports-related

concussion is reviewed in [7].

It is interesting to note that gait analysis is utilized to classify a person's gender based on their gait [43]. Furthermore, attempts to identify a person's emotional state, such as pride, happiness, fear and anger, have been based on gait [44].

### 3.2.3 Deep Learning for Gait Analysis

Supervised machine learning is a branch of artificial intelligence (AI) and a specific kind of machine learning. Algorithms or mathematical models are built and trained with a given set of inputs and desired outputs. A learning algorithm trains the model based on two learning styles, statistical analysis or deep learning, to produce a trained "machine" that carries out the desired task. The models are tested by exploring the data structure based on the learned mapping function to assign hypothesis class which is controlled by the user to evaluate the model performance [14]. Statistical analysis depends on handcrafted features learned in a predefined relationship between the inputs and the output, such as linear regression, logistic regression, decision tree, Support Vector Machine (SVM), random forest, naïve Bayes, and k-nearest neighbor.

Deep structured learning or hierarchical learning is inspired by the biological neural networks' structure and function. It is based initially on the concept of multi-layer Artificial Neural Network (ANN) with the aim to learn data representations automatically; thus, deep learning becomes the method of choice where the classification features, if known at all, are complex, with no straight forward quantitative relation to the raw data. Typically, the term 'deep' refers to the number of layers in the variety of possible networks structures: Deep Belief Networks (DBN), Feedforward Deep Networks (FDN), Boltzmann Machine (BM), Generative Adversarial Networks (GAN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long-Short Term Memory (LSTM) a special kind of RNN. A comprehensive presentation of the theory of ANNs and deep learning is not within the scope of this Review, and the reader is referred to established sources [15]. Further, we focus on models with practical significance for gait applications such as CNN and LSTM [14].

The CNN model is suitable for processing 1D, 2D or 3D data that has a known grid-like topology [16]. The network has the ability to learn a high level of abstraction and features from large datasets by applying a convolution operation to the input data. Commonly, the network consists of convolution layers, pooling layers and normalization layers, with a set of filters and weights shared among these layers.

The convolutional layers output a feature map harvested automatically from the raw input data. The pooling layers are utilized to reduce the size of representation and make the convolution layer output more robust [14],[15]. The CNN model uses commonly two types of pooling layers: max pooling and average pooling. All convolution layers and pooling layers have activation functions (e.g. Sigmoid, Tanh, ReLU, Leaky ReLU), these functions define the output of that node based on calculating the weight of neurons and adding a bias, and deciding on how the weighted sum of the input is transformed into an output [17]. LSTM networks are favorable for processing time series data, where the order is of importance, such as gait data sequences. In essence, they exploit recurrence, by using information from a previous forward pass over the network.

The computational complexities of deep learning are not specific to gait applications. The goal of using ANNs in gait analysis is to develop a model to extract gait features and perform well on unseen real-world gait data with high prediction accuracy. Commonly, for appropriate training and testing, the model is trained and validated on 70% of the data and tested on the remaining 30%. In supervised training, the procedure is launched by initializing the weights randomly, processing the inputs and comparing the resultant output against the desired output. During training, the weights and biases are adjusted in every iteration, until the error is minimized, and validation is used to estimate the model performance during training. Lastly, the model is tested with unseen data, allowing to identify over-training.

The widely used accuracy measure for ANN gait analysis is the confusion matrix [18]. It is a table to visualize the number of predictions classified correctly and wrongly for each class. The table consists of true positive, true negative, false positive, and false negative classification occurrences. One of the advantages of the confusion matrix display is that it is straightforward to identify the decision confusions, thus possibly concluding on the quality of the model and data involved.

## 3.3 Gait Modalities

The evolution of research in gait analysis suggests that, in order to capture the distinctiveness of gait, the various sensing modalities attempt to access biomechanical measures pertaining to the body's physical dimensions, body part masses, or the time-varying muscle-generated forces applied during the gait cycle. In the past decades, a number of modalities have proven their ability to capture gait characteristics and anomalies; however, the historically established methods used to analyze gait heavily rely on handcrafted features. With such an approach, salient features of the problem may be lost in the process of feature engineering, and the classification result can be data dependent. This can be mitigated by utilizing deep learning for its capability of automatic feature extraction, delivering high statistical confidence by learning rich features of gait patterns from sensor data. Sensing modalities for gait data capture used in conjunction with deep learning can be divided into three main groups: video sequence (VS), wearable sensors (WS), and floor sensors (FS); further, each of these is described in more detail. In addition, the different types of algorithms typically applied to analyze gait data are presented and their ability to adjust to the characteristics of a modality or and/or scenario is elucidated.

### 3.3.1 Video Sequence

Gait recognition based on VS has been driven by the advances in general machine learning and image processing methods. The most common aim is to distinguish the identity of a person from a distance. A typical VS system consists of several cameras with optics suitable for capturing the gait cycle. Common VS data sources are suitably positioned CCTV cameras. The information gathered in the form of sequential video frames is subjected to image processing techniques, such as threshold filtering, edge detection, pixel count, background segmentation, counting of light and dark pixels, and converting images to black and white [34]. Gait recognition based on VS in literature is sub-divided into skeleton model-based and skeleton model-free categories. (The above sub-division reference is to skeleton models, not machine learning models.)

The model-based approach is in essence fitting video sequences of gait to multi-segment skeleton models, as proposed in [45],[46]. This method is computationally expensive,

because of fitting skeletal segment models on sensor data, as well as the need to use the model-derived parameters to extract features. The extracted features are classified using statistical analysis methods. The model-free approach is based on extracting gait from VS using feature engineering, as proposed in [47],[48]. Here, deep learning is utilized to automatically extract gait

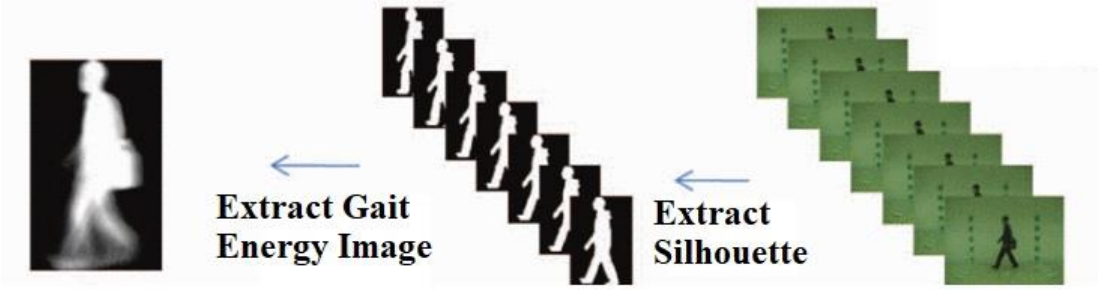



Figure 3.2. From right to left: video sequence, silhouette images and EGI image [37].

features from VS, which maximizes the use of data variability and eliminates the dependence on handcrafting. Most of the available model-free processed data is represented by Gait Energy Image (GEI), maps of optical flow and silhouettes [49] or Chrono-Gait Images (CGI) [50],[51]. These representations, extracted from VS, can capture both spatial and temporal information. As an example representation, GEI is defined mathematically as:

$$GEI(\psi, \varphi) = 1/o \sum_{tb=1}^{o} F^{tb}\ (\psi, \varphi) \tag{3.1}$$

Where $o$ is the total number of frames to represent one gait cycle, $F^{tb}\ (\psi, \varphi)$ is the binary silhouette of the subject at time $tb$ and $\psi, \varphi$ are values in the 2D image coordinate. Figure 3.2 shows schematically the extraction of GEI from the video sequence, in which a single GEI represents human motion sequence at specific time $tb$ using a sequence of silhouettes images. In other words, GEI image is a fused silhouettes images using equation 3.1.

#### 3.3.1.1 Video Sequence Databases

Once the VS representation algorithm is implemented, the machine learning model must be trained, validated and tested to assess its performance. The widely used benchmark is to train and test the algorithm with the following datasets (in chronological order of availability): CMU Motion of Body (MoBo) [52], USF Gait Based Human ID Challenge [53],

CASIA [54], OU-ISIR treadmill [55], OU-ISIR [56] and TUM-GAID [57].

The Carnegie Mellon University Robotics Institute Motion of Body (MoBo) dataset [52] encompasses 25 subjects performing four different walking patterns on a treadmill, namely slow walk, fast walk, incline walk and walking with a ball. The subjects' gait is captured by six high-resolution cameras, distributed around the treadmill.The University of South Florida Gait Based Human ID Challenge dataset [53] captures 122 subjects walking outside with shoes and clothes variations, as well as under different carrying load conditions. Gait is captured from a single viewing angle.

The Chinese Academy of Sciences Institute of Automation Gait Database CASIA [54] is divided into A, B, C, and D datasets. The CASIA A dataset contains 20 people; for each person, it contains 12 image sequences, four sequences for each of 3 angles (0, 45 and 90 degrees) to the image plane. The CASIA B dataset consists of 124 subjects' gait sequences captured from 11 views. The subjects performed normal walking, wearing a coat while walking, and carrying a bag while walking. The CASIA C dataset was captured by an infrared (thermal) camera from 153 subjects performing normal walking, slow walking, fast walking, and normal walking with a bag. The video sequence was taken from one angle at night time. The CASIA D dataset contains the video sequence and footprint images scans of 88 subjects with a wide age distribution. The video sequence is captured from a single angle and with no variations in clothing and carrying conditions.

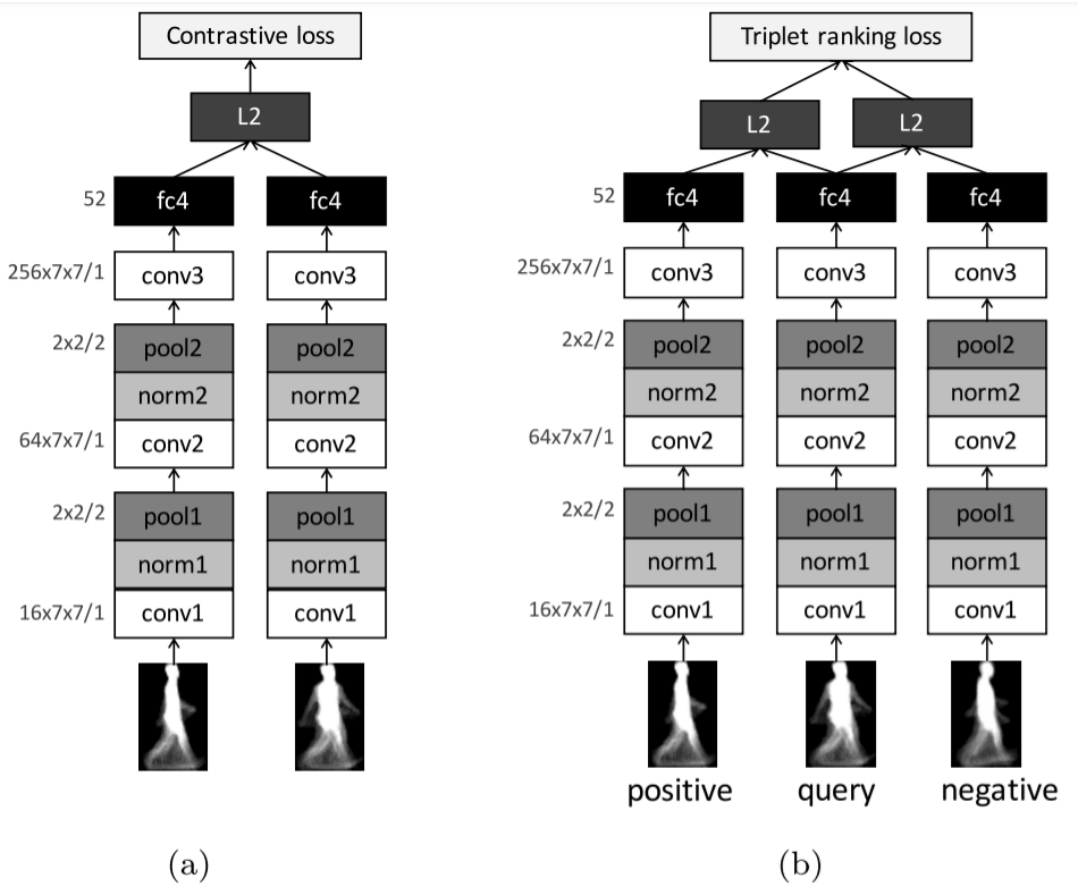
The Osaka University Institute of Scientific and Industrial Research treadmill dataset, OU-ISIR treadmill [55], contains 200 subjects' gait captured on a treadmill by 25 cameras from different angles, 34 subjects with walking at different speeds and 68 subjects with 32 clothing variations. The dataset is distributed in the form of silhouette sequences of subjects while walking on a treadmill. The same group's database on normal surface walking (not involving a treadmill), OU-ISIR [56] dataset, contains 4,007 (2135 males and 1872 females) with ages from 1 to 94 years. The dataset consists of silhouette sequences of the subject's gait captured by two cameras.

The Technical University of Munich Gait from Audio, Image and Depth database, TUM-GAID [57], contains 305 subjects' gait captured by video recording cameras at a single angle, while subjects walk indoors in both directions. Six walking conditions are captured

for each subject from the side view namely four normal walks: one with coating, shoes and one without (left and right), and two normal walks with carrying a backpack variation (left and right). 32 subjects of the cohort are recorded in two sessions (January and April), adding clothes variation.

### 3.3.1.2 CNN Architectures

Table 3.1 summarizes the results yielded by gait recognition VS models, comparing deep convolutional ANNs with automatic feature extraction to statistical analysis algorithms, where features are handcrafted. Deep learning models can be split into two groups, a single deep ANN and multiple deep ANNs joined in the last layer. The network inputs are single or a pair of processed silhouettes sequences. The latter case is mostly used for verifying individual's identity, with a view of 'probe and gallery' gait features. The 'probe' is an identified or verified subject, and the 'gallery' consists of templates as a browsing data set, where the probe is searched and matched to the closest instance in the gallery. These are examined in more detail below, for gait identification or verification.



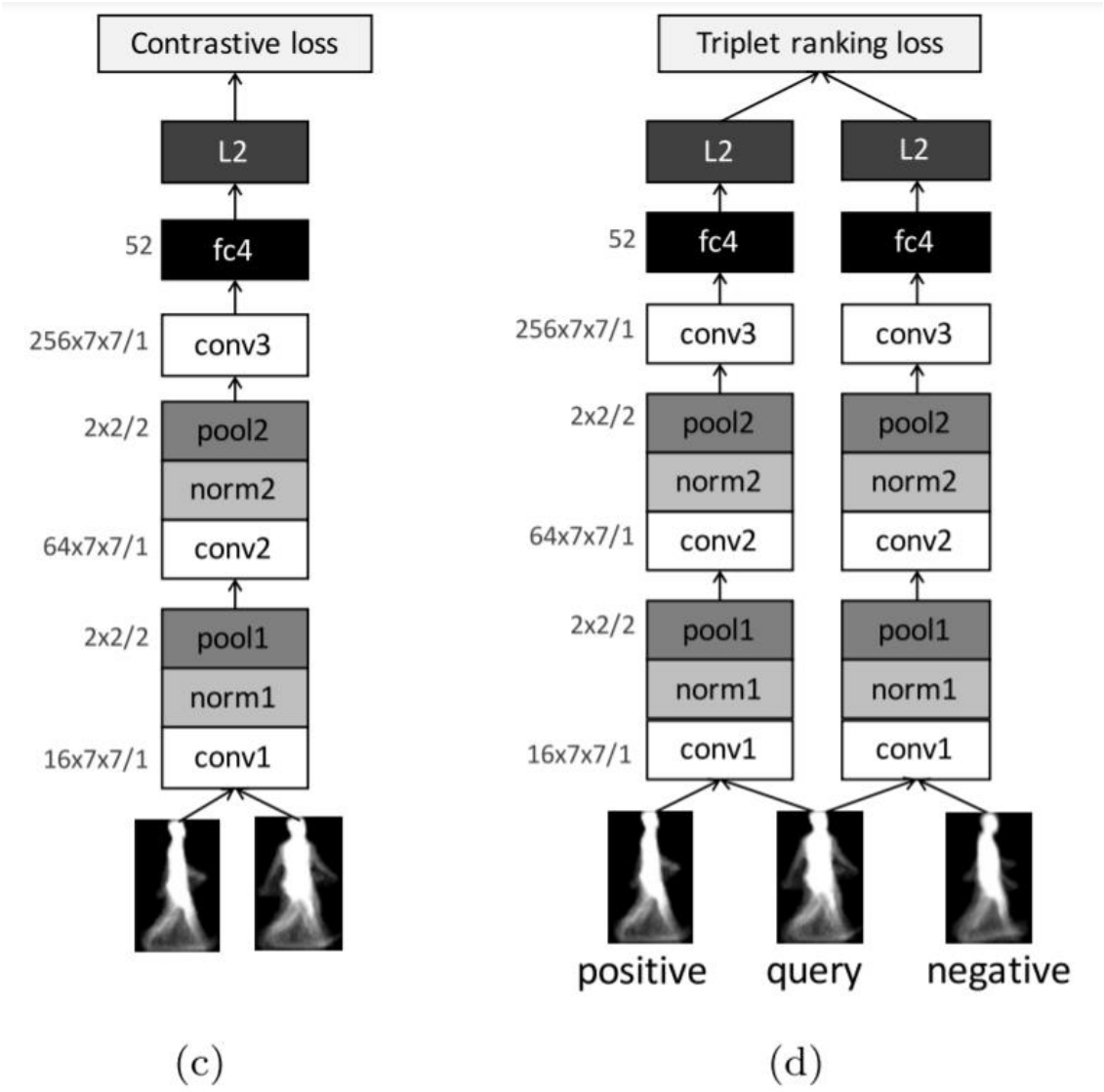



Figure 3.3. [65]: High-level difference architectures for small view-angle differences: a) Siamese CCN with probe and gallery input; b) Triplet CNN with positive probe, negative probe and gallery; Low-level difference variants of a) and b) for substantial view-angle differences: c) single CNN with probe and gallery; d) Siamese CNN.

#### 3.3.1.3 Single Deep ANNs

The single ANN input is a video sequence of images, on which the top *softmax* layer will perform classification based on the desired output for the given input. The *softmax* score outputs 1 for the true-match subject and 0 for false-match subjects. During validation, the loss is computed using cross-entropy between the *softmax* outputs and the corresponding desired output (the ground truth). Single CNN with a single input architecture has been investigated by a number of groups, with some examples outlined

below.

Yeoh et al. [58] used a CNN model trained on a single input as GEI. For testing, the *softmax* classifier in the last layer based on Euclidean distance is replaced by a Support Vector Machine (SVM) classifier to compute one-vs-all (probe vs gallery). The model, evaluated on OU-ISIR Treadmill dataset, yielded competitive performance in clothing-invariant for the identification of people.

Table 3.1. Results for Gait Recognition from VS.

| Accuracy (%) | Model | Reference |
|---|---|---|
| 70.51 (A) | wQVTM | Muramatsu et al. [62] |
| 71.76 (A) | LF+iHMM | Hu et al. [63] |
| 72.80 (A) | TCM+ | Muramatsu et al. [64] |
| 79.66 (A) | STIP | Kusakunniran et al. [48] |
| 80.50 (A) | CNN | Zhang et al. [66] |
| 85.51 (B) | CNN | Alotaibi et al. [8] |
| 85.70 (A) | CNN | Alotaibi et al. [8] |
| 90 (B) | CNN | Wu et al. [69] |
| 90.45 (A) | CNN | Shiraga et al. [59] |
| **92.50 (B)** | RBM | Hossain et al. [47] |
| **94.8 (B)** | Parallel CNN | Wu et al. [69] |
| **98.8 (A)** | CNN | Wu et al. [69] |
| **95.0 (A)** | CNN | Yan et al. [37] |
| **95.04 (A)** | CNN | Li et al. [72] |
| **97.35 (A)** | 3D-CNN | Wolf et al. [60] |
| **98.8 (B)** | Parallel +Triplet CNNs | Takemura et al. [65] |

(A): Different scenes; (B): Different views

wQVTM: with Quality Measures View Transformation Model

LF+iHMM: Local binary pattern Flow + individual Hidden Markov Model

TCM: Transformation Consistency Measures

STIP: Space-Time Interest Points

Yan et al. [37] proposed a CNN model with Multilayer Perceptron (MLP) classifier. The input is a single GEI for automatic extraction of gait features. The CASIA-B dataset is used for evaluating the methods. The model is trained using multitask learning to predict multiple human attributes. 95.88% accuracy for each task is achieved; however, it was realized that the changes of scenes or view could be generalized better by training on more data. Shiraga et al. [59] designed GEINet, which is a CNN with two sequential groups. The network input is a single GEI image (from OU-ISIR database) in the training stage. In the testing stage, the dissimilarity between a probe GEI and gallery GEI pair is computed using the distance between them at the fully connected layer. The model performs well on cross-view for gait verification and identification.

Wolf et al. [60] proposed a 3D CNN with a 3D spatio-temporal tensor as input, consisting of a grey-scale image for the first channel and optical flow for the second and third channels. The model is trained and tested using the CASIA-B dataset, MoBo database and UFS database. The approach was evaluated on variations in walking speed, clothing and the view angle. Based on this architecture, Castro et al. [61] used a spatiotemporal 3D tensor of the optical flow as the input of the CNN. The network was trained and tested using the TUM-GAID database with gait scenarios, clothing and carrying variations for each subject. Although the network accuracy was significantly improved using the optical flow rather than using silhouette-based input. However, it is difficult to generalize on which feature extraction method outperformed the other, since [60] and [61] are evaluated on different datasets. Nevertheless, it is clear that the optical flow feature can present robust gait spatiotemporal information for use in a CNN architecture.

#### 3.3.1.4 Dual Deep ANNs

The input into a dual network consists of two different images, as probe and gallery under similar conditions; however, different gait scenarios, viewing angle, as well as clothes and carrying conditions, may be involved. This architecture is effective in gait verification since the networks have the same weight and structure, which allow the extraction of gait features automatically in the same manner. The outputs are matched using contrastive loss to find the Euclidean distance. The latter can be compared to a threshold to identify matching pairs or to label an imposter if a match cannot be found. Below is an outline of architectures applied for CNNs with two inputs. Figure 3.3 [65] presents some dual

architectures used for verification and identification.

Zhang et al. [66] designed a shared parameters 'Siamese twin' CNN, each twin comprising a convolutional layer, a max-pooling layer and three fully connected layers to extract gait features automatically. The two twin outputs are connected to the contrastive loss layer. A pair of similar or dissimilar GEI images from the OU-ISIR database are used as an input to the Siamese network. In the training stage, the weights are shared simultaneously to optimize the network, and the model is fine-tuned by back-propagating with a contrastive loss. The gallery member with the nearest training sample is identified by testing to allow the feature metric computation of a discriminative loss function. The latter drives the similarity metric [67] to be small for pairs representing the same subject, and large for different subjects. Considering the changes of cross-view in real-world human identification scenarios, the model performs well in gait verification.

Wu et al. [68] proposed a CNN to extract gait features directly from the raw silhouettes' sequence for cross-view gait recognition. Gait sequences from the CASIA-B dataset are used to train and test the network. In the testing stage, the Euclidean distance is measured for similarity using the probe and gallery method, achieving an accuracy of 94.1%. Furthermore, in [69] several CNN that take two inputs as probe and gallery have been shown to outperform other approaches, including twin CNNs [65],[68]. Two GEI images are used for gait verification based on cross-view gait recognition. The dataset to train and test the proposed networks are the CASIA-B dataset, OU-ISIR database and USF database. The proposed methods outperformed the previous state-of-the-art methods by a significant margin in the three datasets.

For cross-view gait recognition, Takemura et al. [65] considered different architectures for verification and identification. This is based on the assumption that the absolute similarity scores are important for the verification task, while the relative similarity scores between a probe and the galleries are important for the identification task. For verification, a Siamese CNN with shared parameters is proposed (see figure 3.3a) to discriminate whether two inputs originate from the same subject or not, based on the contrastive loss value. For identification, three parallel CNNs are deployed as a triplet network (see figure 3.3b). The triplet input is three GEIs: a query (the probe subject), a positive (from the same subject) and a negative (from a different gallery member). A

triplet ranking loss is defined as the difference between two feature vector distances: the distance between positive and query and the distance between negative and query. The parameters of the triplet CNN are trained so that the dissimilarity between a probe and the same subject is relatively lower than that between a probe and different subjects. To accommodate possible substantial differences in the GEIs by viewing angle, low-level difference structures are introduced, as they are more directly affected by appearance differences due to taking the difference between a matching pair closer to the input level (see figure 3.3c) and figure 3.3d). Cross-view gait recognition is demonstrated on OU-ISIR and OU-ISIR Multi-View Large Population datasets, with 10,307 subjects' video sequences captured from 14 angles (see figure 3.4); however, the existing methods are difficult to evaluate on this dataset, and OU-ISIR LP is utilized to confirm the hypothesis regarding the network architecture.

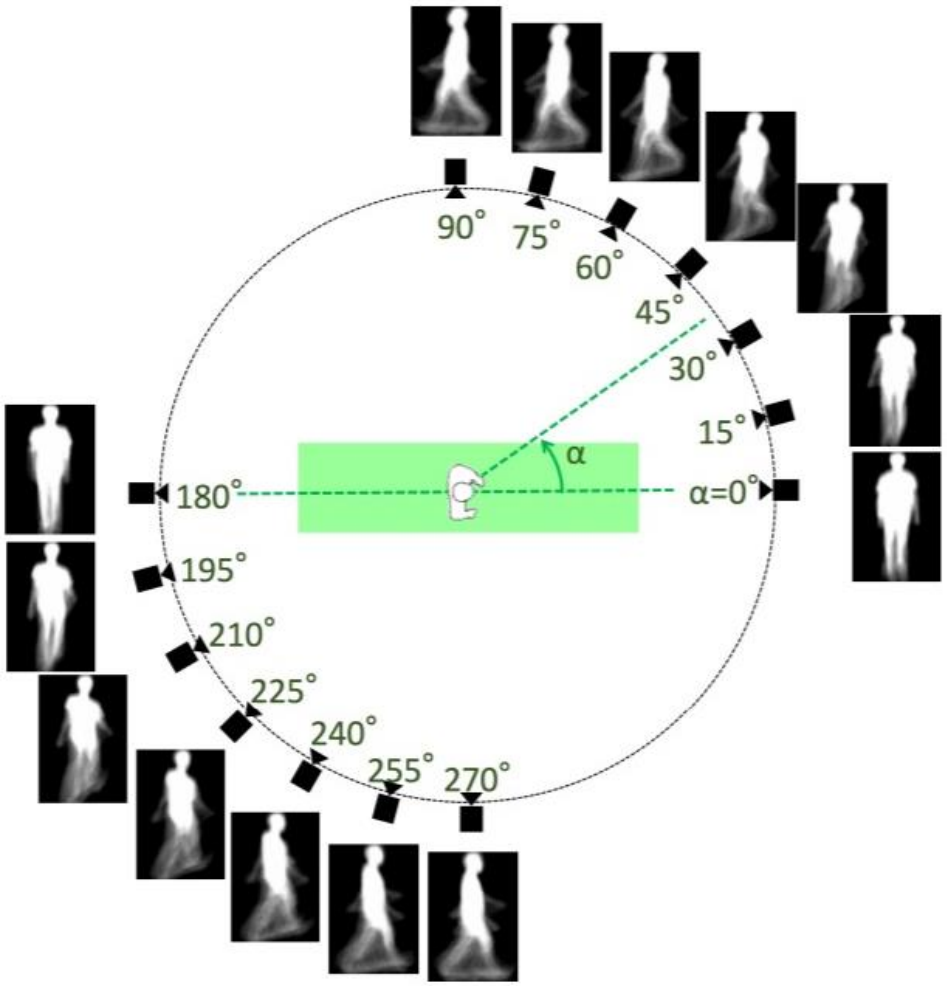



Figure 3.4. Gait GEI images at 14 viewing angles [65].

#### 3.3.1.5 Transfer Learning

Transfer learning is a comparatively new concept in ANNs and is the next strongest driver, after supervised learning, of the commercial success of machine learning [60]. Essentially, it is applying knowledge gained to solve a problem to a multiplicity of related problems. 'Pre-trained' models are beneficial as a starting point on specific ANN solutions, given the vast computing and time resources required to develop detailed physical models on these problems. Compared to starting from scratch, Transfer learning allows a substantial jump in the starting point for the delivery of a related ANN model [71].

Li et al. [72] used supervised pre-training of a VGG-D CNN (Visual Geometry Group) model and evaluated the efficacy of learned features on gait recognition tasks. The network consists of 16 convolutional layers and 3 fully connected layers with a nearest neighbor classifier. The silhouette images from the OU-ISIR dataset are used to train and test the network without fine tuning to capture gait spatiotemporal aspects. The probe and gallery method is used to identify people in a cross-view setting, significantly outperforming prior state-of-the art methods for both verification and identification.

Alotaibi et al. [8] determined empirically the appropriate CNN architecture for automatic gait feature extraction from GEI images using the CASIA-B dataset. They applied two transfer learning methods to the network pre-trained with 24 subjects. 'Fine-tuned CNN' involved adding one more subject (new total of 25 subjects) and dropping the weights of the *softmax* layer followed by re-training of the entire model; 're-learn *softmax* only' involved 'freezing' the weights of the convolutional layers and the weights of the *softmax* layer were re-learned. While the computational time for pre-training was 124.82 s, adding a single subject by fine-tuned CNN took 42.41 s and only 22.12 s by *softmax* re-learning.

### 3.3.2 Wearable Sensors

WS are an obvious means to acquire human gait due to their convenience, efficiency and lower price. Unlike other gait capturing systems, WS impose upon the user to cooperate wearing the device in a non-invasive way to provide gait signals. The advances in

electronic devices and signal processing techniques have extended the applications of WS sensors to produce a measurement of human body orientation, position and specific force in space and time. The inertial measurement unit (IMU) is a type of WS system that has been extensively used due to its small size, cost, light weight, and good precision characteristics. A typical IMU provides the most widely used combination of sensing modalities to capture human activities, including gait. It comprises of an accelerometer, a gyroscope and often a magnetometer, which gives the heading direction. Additional components such as batteries, microprocessors and communication modules are arranged to jointly operate an IMU system.

Gyroscope sensors measure the angular velocity as the rate of change of the sensor's orientation, while accelerometer sensors measure the acceleration of the body resulting from the acting forces in the opposite direction. A combination of these sensors can create a comprehensive report on the human body orientation, gravitational forces, velocity and acceleration [29].

Furthermore, it has been found convenient to use the gyroscope and accelerometer, usually integrated in a smartphone, benefiting from predictable availability and positioning, as well as eliminating the need for additional hardware. Mobile users' authentication is an acceptable approach when other gait authentication is not deployable. In the healthcare domain, IMU-equipped smartphones allow inexpensive prediction of falls due to neurological disorders or freezing of gait in patients [73]. The computing power on-board of a smartphone can be used as a standalone system to perform all tasks required for decision making and communicating with healthcare providers in any life-threatening situation.

The analysis of WS signals is a challenging task considering the large number of observations recorded per unit time. This is due to the spatiotemporal nature of the gait cycle and the difficulty to relate in a straightforward manner WS signals to a known gait characteristic. Manual feature extraction is the classical way for gait analysis using WS, and it is time-consuming and depends on knowledge of the context in which the signals are acquired. Since performance is key in real world applications, deep learning has emerged as a promising data processing method by extracting automatically reliable discriminative features of human gait, outperforming the approaches based on

handcrafted features. Table 3.2 summarizes the results yielded by gait recognition models based on WS using various deep ANN models.

Table 3.2. Results for Gait Recognition from WS.

| Reference | Model | Evolution | System | Feature |
|---|---|---|---|---|
| Zebin et al. [75] | CNN(L=7) | **97%** | Single IMU sensors | Activity + gait (A) |
| Ordóñez et al. [76] | DeepConv LSTM | 95.8**%** | Five IMU sensors | Activity + gait (A) |
| Gadaleta et al. [77] | CNN+SVM | **FP, FN <1%** | Single IMU sensors | Gait cycle (A) |
| Gadaleta et al. [78] | CNN+SVM | FP, FN <0.15% | Smartphone | Gait cycle (A) |
| Zhao et al. [79] | CNN(L=7) | 61.0**%** | Smartphone | Gait cycle (A) |
| Dehzangi et al. [81] | CNN (L=10) | **97.06 %** | Five IMU sensors | Gait cycle (A) |
| Lorenzi et al. [82] | ANNs | - | Single IMU sensors | Gait irregular steps PD (B) |
| Camps et al. [83] | CNN(L=8) | 90.6**%** | Single IMU sensors | Gait cycle (B) |
| Kim et al. [84] | CNN(L=7) | 91.8**%** | Smartphone | Gait cycle FOG (B) |
| Xia et al. [85] | CNN(L=7) | 90.60**%** | Three IMU sensors | Gait cycle FOG (B) |
| Rad et al. [87] | CNN (L=10) | 0.65**%** ± 0.01**%** | Three IMU sensors | Gait cycle FOG (B) |
| Murad et al. [90] | RNN | 94.1**%** | Three IMU sensors | Gait cycle FOG (B) |
| Ravì et al. [88] | Deep ANNs | 95.8**%** | Three IMU sensors | Gait cycle FOG (B) |
| Hannink et al. [91] | CNN(L=5) | **0.97 CC** | Two IMU sensors | Stride length (B) |
| Aicha et al. [93] | Conv LSTM | 94**%** | Single IMU sensors | Gait cycle (B) |
| Hu et al. [94] | LSTM | **97.0%** | Single IMU sensors | Gait on uneven and flat surface (B) |

A: Normal gait; B: Abnormal gait; CC Spearman correlation. Accuracy is above 60%.L: number of layers.

The sensor position on the body and the number of sensors comprising the system are an essential factor for the quality of the harvested data. In a systematic review analysis, Panebianco et al. [74] assessed accuracy and repeatability using 17 algorithms for their ability to monitor temporal parameters of human gait from 5 IMUs: one on the back, two pairs on the shanks and two pairs on the feet. For estimates of stance time, algorithms based on the acceleration of the shank and foot perform better than those based on the lower back; however, the sensor position did not affect the step estimation. For toe-off and heel strike detection, algorithms estimating angular velocity performed better overall, with notable dependence on the sensor positioning. Analysis has concerned mostly with the distinction between normal and abnormal gait, as follows below.

#### 3.3.2.1 Normal Gait Analysis

Analysis of normal gait parameters using WS has immensely attracted the interest of researchers and clinicians. The following are different methods and techniques that have been proposed and implemented for various applications.

Zebin et al. [75] proposed a system comprising 5 IMU sensors, worn on the lower back, thighs and shanks, for activity recognition including gait. A CNN based model is used to extract the features automatically from time-series raw data and achieve higher accuracy compared to the handcrafted features with statistical analysis. In another work, 7 IMUs positioned on the chest, arms and legs along with the12 accelerometers close to the limb joints, were used by Ordóñez et al. [76]. A DeepConvLSTM model is trained in a fully-supervised manner on human activities including gait. The DeepConvLSTM model outperforms previous results on the same dataset. However, increasing the number of sensors exacerbated the extraction of gait features compared to the use of WS attached to the pelvis and lower limb only [75].

For gait authentication, Gadaleta et al. [77] used a CNN model (see figure 3.5) to extract gait features from a single WS placed on the shank for each subject. Data from 15 subjects' gait is used in the training stage and 9 in the testing stage. In the latter, the network weights are frozen, and the CNN model is used to extract features, further the features are feed to SVM for classification. Thus, increasing the training dataset was suggested for improving the model performance. In a later work by Gadaleta et al. [78], the proposed

CNN model is used to extract gait feature vector from a single subject automatically, the gait feature are used to train a single-class SVM. The system can distinguish between an impostor and the user whose gait is used for training. The IMU signals acquired from smartphones are tested on a user against 14 impostors, yielding false positive and false negative rates less than 0.15%.

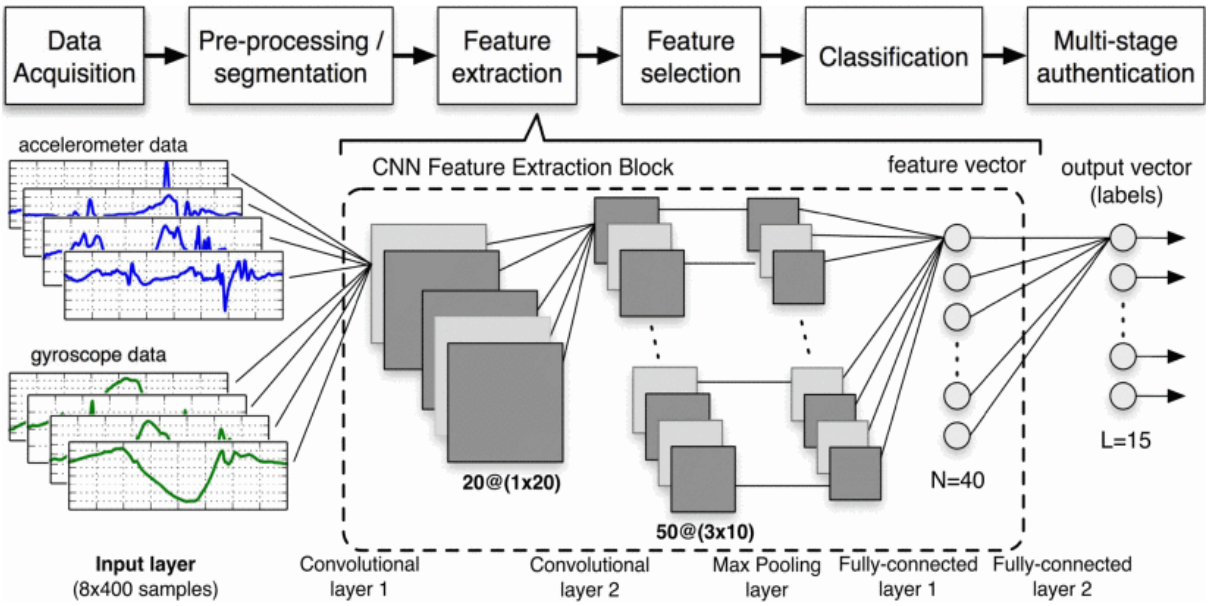



Figure 3.5. Convolutional neural network to extract and classify gait features from wearable inertial measurement unit with accelerometer and gyroscope sensors [77].

Zhao et al. [79] proposed a CNN model for gait labelling and authentication. The input to the network for automatic gait features' extraction is an Angle-Embedded Gait Dynamic Image (AE-GDI), which is a transformation of a WS data series. This allowed comparison with the state-of-the-art performance on VS (OU-ISIR) and WS (MCGILL [80]) datasets.

Similar to [75], Dehzangi et al. [81] placed 5 WS at various body locations. WS signals obtained from the sensors at chest, right wrist, knee and ankle, as well as the lower back of the subject, allows the study of CNN performance on time-frequency image transformation of raw signals. A total of 10 subjects' gait data were used to train and test the network; accounting for the multi-sensor character of the data, early and late fusion methods were applied, achieving state-of-the-art in subject identification. The deep learning approach to sensor fusion is addressed in more detail in Section 3.4.

#### 3.3.2.2 Abnormal Gait Recognition

Deviations from normal gait are extensively studied by WS, the main targets being to classify neurodegenerative conditions, or to prevent falls in older adults. While the assumptions underlying various algorithms differ, in practical applications it often appears more convenient to use a single WS for capturing a discriminative gait feature. The sensor system embodiments used for abnormal gait analysis can be grouped into dedicated IMU systems and smartphones. Lorenzi et al. [82] used a single IMU unit positioned on the head, to collect gait patterns during the gait cycle, aiming to distinguish normal gait from the freezing of gait and irregular steps in Parkinson's disease (PD), using dynamic time warping to select the input features to the ANN.

Deep learning recommended itself as an improved approach to recognizing the abnormality in human gait, in terms of classification accuracy and computational requirements. Camps et al. [83] used a waist-positioned IMU and an 8-layers CCN to achieve an accuracy of 90.6% to detect freezing of gait (FOG) detection in PD patients. The optimal architecture implemented with two convolution layers and 20 convolution filters. In [84], The gait of 32 patients was recorded by a smartphone accelerometer and gyroscope casually placed in the subject's trouser pocket. The CNN detected the FOG events in Fourier space with 91.8% accuracy, which is slightly higher than the CNNs methods proposed in [83].

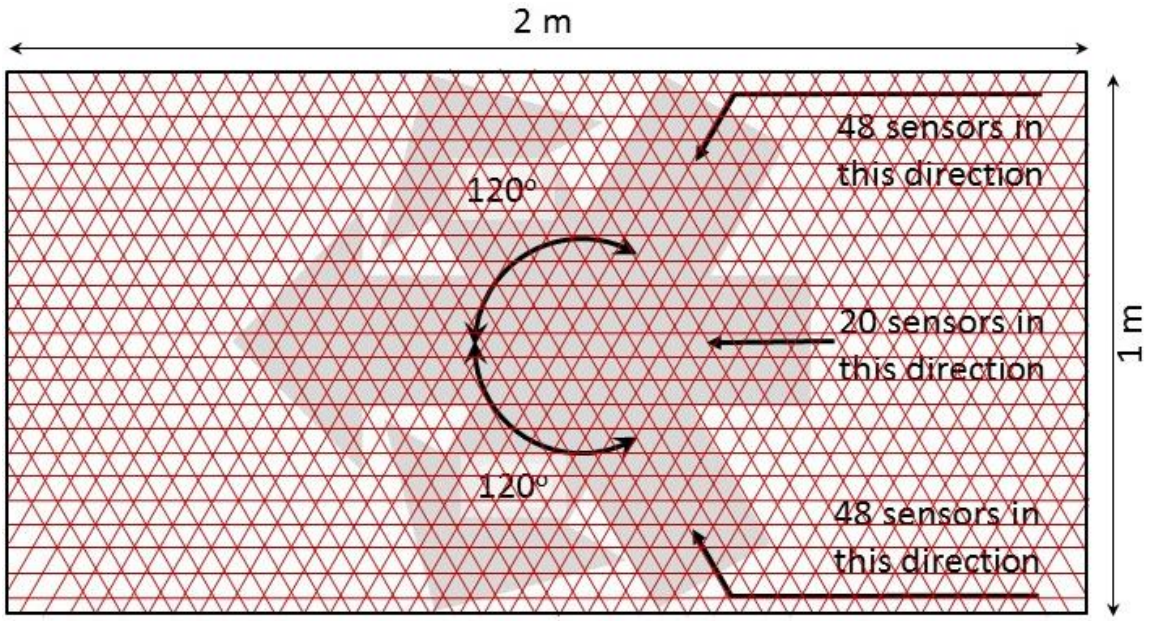
In a recent study, Xia et al. [85] proposed a CNN to extract gait features from three accelerometers positioned above the hip, knee, and ankle. Against the aim to distinguish FOG events from normal gait, evaluation on the Daphnet FOG dataset [86] from 10 subjects yielded an accuracy of 90.60%. Several other deep ANNs [87],[88] have been trained and tested for human activity recognition from raw spatiotemporal datasets, including the FOG dataset used in [86]. Rad et al. [87] and Hammerla et al. [89] used a CNN performing well in human activity recognition; however, the performance on the FOG dataset was weaker. Murad et al. [90] improved the FOG recognition accuracy to 94.1% with their proposed deep RNN trained on the Daphnet FOG dataset. Ravì et al. [88] argued that deep learning models do not perform well when a small number of activity segments are available and proposed feature fusion, where shallow features are fused with features derived by deep learning in the fully connected and the *softmax* layers. With

Daphnet FOG data, this method yielded for 'freeze' and 'no freeze' precision of 67.89% and 97.40%, as well as recall of 59.52% and 98.15%, respectively.

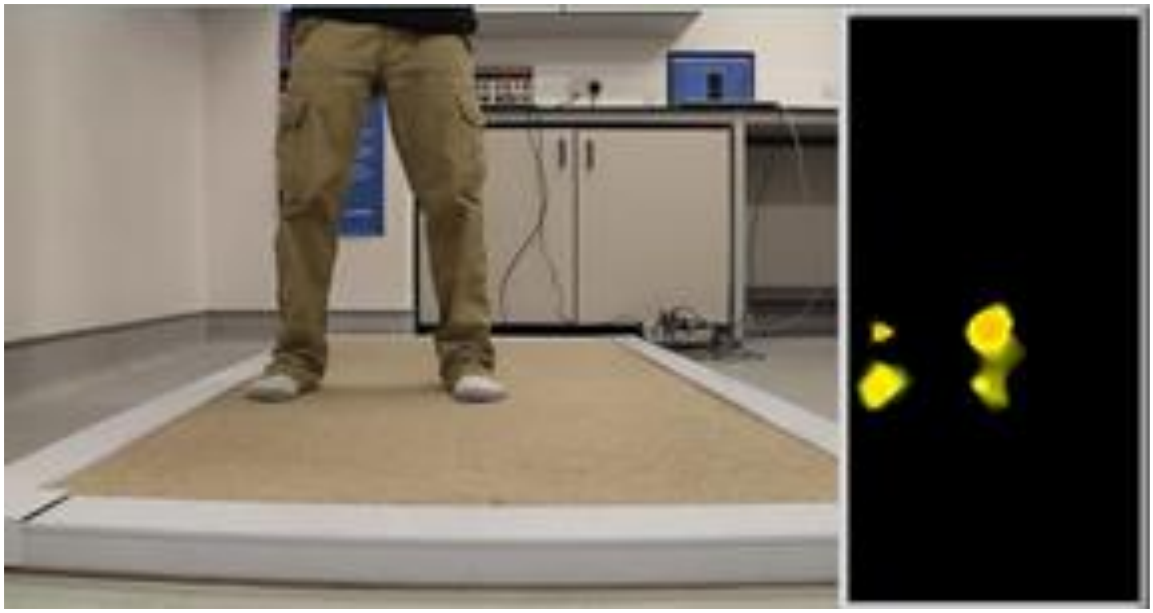
As an alternative use of deep learning, stride length estimates are derived in clinical settings to indicate, an early or further progression stage of neurological disorders. In the work reported by Hannink et al. [91], stride length is estimated automatically using WS and deep CNNs. The WS set consists of a 3D-accelerometer and a 3D-gyroscope attached below each ankle joint. The aim of this approach is to extract spatiotemporal gait parameters to aid the physician in scoring gait impairment objectively. The CNN performance was evaluated on the eGAIT dataset [92], using 10-fold cross validation on three different stride types. It was observed that the performance was dependent on stride definition and the better results were achieved for mid-stance to mid-stance intervals. Importantly, the CNN analysis of WS data was not affected by the use of a four-wheeled walking aid, where the data processing became problematic with the GAITRite walkway sensor system (see section 3.5.2.).

Gait analysis using WS has been extensively studied for the detection of falls in older adults. Most of the reported work is based on handcrafted features deep learning is appeared as an improved approach in terms of increased classification accuracy and reduced computational load. Aicha et al. [93] reported work on CNN, LSTM, and ConvLSTM models used to extract gait features from raw accelerometer signals positioned on the lower back. The model trained and tested on 296 participants' gait to predict fall risk as the main task and user identity as an auxiliary task. The models' performance with features extracted using deep learning was observed to be marginally better compared to handcrafted features.

Hu et al. [94] attempted to capture the higher risk of falling while walking on uneven surfaces as compared to the flat surfaces walk. Essential here is the ability of subjects, as a function of age, to produce the stability required to avoid a fall. A single IMU unit positioned on the trunk delivered raw signals from 35 users: 17 older adults (age: 71.5 ± 4.2 years) and 18 young adults (age: 27.0 ± 4.7 years) used as an input to the LSTM network. Automatically extracted spatiotemporal gait parameters are used to classify age-related differences in walking on a flat or uneven surfaces.




a)



b)

Figure 3.6. iMAGiMAT footstep imaging system.

a) geometry and number of the POF sensor elements (after [108]).

b) tomography image reconstruction (right panel) as top view of deformation by a person standing (seen in the image in the left panel) with weight on the ball of right foot and the left heel [95]. The orientation of a) is at 90 degrees with respect to the two panels in b).

### 3.3.3 Floor sensors

One of the key points in monitoring human gait is to capture the forces placed on the ground by the foot during gait cycle. The interaction of the human body with the walking surface is the point of contact with the environment, which cannot be avoided or modified at will. This interaction is typically described in terms of the GRF. Figure 3.1 emphasises that the details of the GRF dynamics follow the gait cycle, as the 7 intervals are defined by the contact of one or both feet with the walking surface. This interaction is highly individual: in the short term it can vary as a result of a temporary psychological or physiological condition and longer term changes can take place as a result of ageing or a long-term healthcare condition. Gait monitoring with floor sensors requires minimal, if any, cooperation or attention by the user and is amenable to embodiments for long period, continuous data capture. This motivates the advances in sensor technology for footsteps capturing systems and processing of GRF data to extract distinctive information on gait events, evolution of walking habits and reaction to physical and psychological interventions. Typical applications of FS are in the fields of biometrics, healthcare, sports, safety and security.

GRF data obtained from force plates has been successfully used for biometrics in [39],[95],[96]. Vera-Rodriguez et al. [95] have assembled the largest to date footstep database, SFootBD [97], containing about 9900 single strides from 127 volunteers. In the healthcare context, GRF sensor data has been used for flat foot diagnosis in children [98], for falls detection in a smart home environment [99] and for monitoring performance on dual cognitive tasks [100],[101]. Discrete switches [100],[102],[103], a row-column contact wire mesh [101] and pressure sensors [98],[99],[101], have been most commonly used as floor GRF sensors to derive stride length, width and duration; stride variability; cadence; velocity and other spatial characteristics of gait [100],[101],[102],[103] as well as time-on-heel to time-on-toe ratio [104]. While these features are of common use in healthcare practice, they are not straightforward to extract and interpret from substantial volumes of raw data. Consequently, data-mining methods and shallow machine learning have been introduced in the past couple of decades to process data from FS. Table 3.3 summarizes the results yielded by gait recognition models based on floor sensor using deep ANNs compared to statistical analysis.

Table 3.3. Results for Gait Recognition from Floor Sensors.

| Reference | Model | Accuracy (%) | Sensor System | Feature |
|---|---|---|---|---|
| Vera-Rodriguez et al. [96] | SVM | 6-10 EER | Piezoelectric sensors | Identifying |
| Leusmann et al. [99] | Same variable properties | 72 | Piezoelectric sensors | Stepping |
| Suutala et al. [102] | Gaussian process | 84.26 | Binary switches sensors | Identification |
| Yun et al. [103] | MLP | **96** | Switch sensors | Identification |
| Middleton et al [104] | Euclidean distance | 80 | Switch sensors | Gait (Parameter) |
| Singh et al. [105] | CNN+RNN | 87.66 | Conductive polymer fibre sheet | Identifying |
| Costilla-Reyes et al. [108] | Random forest | 90.84 ± 2.46 | Plastic optical fiber | Gait pattern signal subject |
| Costilla-Reyes et al. [109] | 3D-CNN | **97.88 ±1.70** | Plastic optical fiber | Gait pattern 13 subject |
| Costilla-Reyes et al. [110] | Deep residual ANN | **0.70-1.70 EER** | Piezoelectric sensors | Identifying |

EER: equal error rates; Parameters: stride length, stride cadence, and time on toe to time on heel ratio.

The recently demonstrated success of deep learning in processing of VS and WS data has induced interest in applying CNNs on data from FS. Singh et al. [105] proposed a pre-trained 17 layers CNN and gated recurrent units to extract gait features automatically from images of footstep GRF. The images were obtained on a 1 cm pixel grid covering an area of 80 cm x 80 cm, as point measurements of resistance between the upper and lower surfaces of a conductive polymer fibre sheet. The raw sensor data is used in image format as an input to the Inception-v3 model. The model is tested on identifying 13 people and yielded an accuracy of 87.66%. The limited volume of the training dataset was identified as the main hurdle towards better performance of the proposed method.

Cantoral-Ceballos et al. [106] used a principally different approach to floor sensing: instead of point measurements, they used a distributed Plastic Optical Fibre (POF) sensor

layer sandwiched unobtrusively between the top pile layer of a commercial carpet and deformable underlay, implementing Guided-Path Tomography [107] (iMAGiMAT, see figure 3.6). With frame rates of 256 Hz and spatial sampling adequate for inverting the data into footstep image frames, it was possible to capture in substantial detail the dynamics of an uninterrupted sequence of at least 4 footfalls at a time.

Costilla-Reyes et al. [108] demonstrated that, in the classification of 10 manners of walking from temporal data subsets, deep learning models (Deep Feed Forward ANN with 10 hidden layers and a RNN) outperformed statistical analysis, with some exceptions attributed to the shortage of training data. This was partially mitigated in a further work [109] where the UoM-Gat-13 dataset was introduced, as a full set of spatiotemporal raw signals (1400 frames at 256 Hz from each of the 116 sensors) from 10 manners of walking and 3 dual tasks. The raw signals were down-sampled, reshaped, and normalized to form a spatio-temporal input sequence for a CNN. The latter consisted of two convolutional layers, followed by one average pooling and one max pooling layers. The network, trained and tested on the UoM-Gait-13 dataset using a spatiotemporal Raw Sensor Matrix (RSM) representation, achieved classification accuracy of 97.88 ± 1.70%. For comparison, tomography images were reconstructed from the raw data and classifications by shallow and deep learning models were obtained for the three input options: raw spatiotemporal sequences, RSMs and reconstructed images. The deep learning approach with RSM input outperformed by a margin all others, on all measures: accuracy, precision, recall and F-score.

A deep residual ANN based on ResNet architecture was proposed by Costilla-Reyes et al. [110] for footstep biometrics. The pressure magnitude exerted by footsteps is sampled by two floor mats, with 88 piezoelectric sensors each, arranged to capture most of the full gait cycle. Different representations are adopted for the raw spatial and temporal components of the data. For the spatial component, each footstep frame is reshaped into a 2D matrix with the sensors of the two mats concatenated and pixels re-calculated as accumulated pressure. The temporal component representation optimises the data variability against training time by selecting frames corresponding to the heel strike, flat foot and heel-off intervals. Correspondingly, the network architecture consists of spatial and temporal streams; each stream has convolution, batch normalization, max pooling

and fully connected layers. Features learnt by the model are classified in the final *softmax* layer using a one-vs-one linear SVM. Class-score level fusion, applied on the outputs from the classifiers of the spatial and temporal streams, were proven to perform better than lower level feature fusion. Biometric verification was demonstrated on the SFootBD database [97] for three benchmarks, driven by common security scenarios: airport concourse (40 stride footsteps for 40 users and 763 impostors), office area (200 stride footsteps for 15 users and 2697 impostors) and private dwelling (500 stride footsteps for 5 users and 5603 impostors). In all three benchmarks, the deep residual ANN outperformed shallow CNNs and FNNs, as well as the handcrafted feature approach in [98]. In the private dwelling scenario alone, the models improved the Equal Error Ratio (EER), more than 3 times in validation and more than twice in evaluation, over the state-of-the-art. The superior performance was assigned to combining together the ResNet and SVM models, as well as the distinct representations for the spatial and temporal components.

## 3.4 Multi-Modality Gait Sensor Fusion

In its narrow sense, multi-sensor data fusion is combining data captured from multiple information sources, where the resulting information pool produces a new representation, distinct from those captured by individual sensors [111]. Gait feature fusion has been extensively used to study human gait features and anomalies associated with forces generated during the gait cycle. Deep learning is called for to combine multi-sensor data from all three modalities reviewed in Section 3.3. Several WS data are fused in the ANN layers to deliver body orientation, position and specific force in space and time. FS are based on sensor fusion since the proposed methods are based on using a set of switch sensors, pressure sensors or POF sensors to log the forces associated with foot ground contact. Further, we focus on the fusion of gait spatiotemporal sequences captured from at least two modalities, e.g. lower limb joint angle trajectories captured by VS or WS, and forces generated by the foot contact captured by FS or sensors under the foot. Table 3.4 summarizes the results yielded by gait recognition models based on sensor fusion using deep ANNs.

In the healthcare context, deep learning has been used for data fusion to address gait-

phase detection. Ding et al. [112] performed real-time gait-phase detection using one IMU sensor mounted on the shank to measure the absolute heading and angular velocity, as well as three foot-switches to label gait activities using deep learning. An LSTM-based gait-phase recognition algorithm is used to train the labelled data. Results showed 96.1% accuracy as compared to 89.1% and 91.8% for statistical analysis techniques such as SVM and MLP, respectively. The reported results shows a strong correlation between gait phase and the kinematic of the shank.

Deep learning was used to fuse sensors data by Vu et al. [113] for gait-phase detection to assist in taking full control of gait for transtibial prostheses users. This will result in sufficient control of active prosthetic devices in real-world applications. The proposed algorithms detect gait-cycle percentages and predict future gait percentages in the case of a delay in the system. An Exponential Delay Fully connected ANN (ED-FNN) is developed for this purpose. It is based on short and long delay to predict fast changes in the gait cycle progression on flat and 15-degree inclined surfaces.

The model was trained and tested to detect gait-phase from raw IMU signals, positioned on the lower shank. Two force-sensitive resistors (FSR) were placed under the foot for accurate heel strike and toe-off detection. Although, strictly speaking, FSR data was not used in the ED-FNN processing; however, it contributed to a better quality data input to the network. The model performs well in an offline setting as compared to other methods based on handcrafted features. Furthermore, this methodology uses less computational power, which is an essential factor to deploy on autonomous systems.

Multi-channel redundant fusion for generating bipedal gait was proposed by Mazumder et al. [114] with the aim to obtain a robust stride time and gait phase using a Radial Basis ANN. The stride time is calculated and fused to derive a robust fail-safe timing information based on which joint trajectory mappings. The proposed methodology estimates the user's intention to start, stop or change a particular gait pattern. A set of sensors are used for test data, namely an IMU sensor, foot pressure sensors and a myoelectric sensor for electromyography (EMG). The four EMG signal channels are fused with the pressure and IMU sensors signals to estimate stride time using the ANN. The proposed method is tested on five subjects walking on a treadmill, yielding classification accuracy with minimum square error $< 0.05$.

Table 3.4. Results for Gait Recognition from Multi-Modality Sensor Fusion.

| Reference | Model | Acracy (%) | Sensor System | Feature |
|---|---|---|---|---|
| Ding et al. [112] | LSTM | 96.1 | IMU sensor + footswitches | Gait-phase |
| Ding et al. [112] | SVM | 89.1 | IMU sensor + footswitches | Gait-phase |
| Ding et al. [112] | MLP | 91.8 | IMU sensor + footswitches | Gait-phase |
| Vu et al. [113] | ED-FNN | 2.1± 0.1 | IMU + force-sensitive resistors | Gait-cycle percentages |
| Mazumder et al. [114] | Deep Radial Basis ANN | square error < 0.05 | IMU + pressure sensors + EMG | Stride time and gait phase |
| Mun, et al. [115] | ANNs | 95 | IMU+ pressure sensors | Gait cycle events |
| Kumar et al. [116] | 3DCNN+ LSTM | 91.3 | IMU + pressure sensor+ camera | Gait recognition |
| Vera-Rodriguez et al. [118] | SVM+RBF | 1.6 EER | pressure sensors + camera | Identification |

EER: equal error rates; RBF: radial basis function; EMG: electromyography

In a similar approach, Mun et al. [115] used a deep ANN to estimate and quantify spatiotemporal gait parameters from foot characteristics. This was achieved with a footstep feature measurement system that scans the foot while a subject performs various motion tasks, and a set of IMU sensors integrated in a commercial motion-capture system (Xsens MVN, Enschede, The Netherlands), to detect heel strike and toe-off off events during gait cycle. The sensors data is fused in the deep layers of the ANN to estimate the gait features during the gait cycle, namely: stride length, step length, velocity, stride time, step time, single-limb support time, double-limb support time, as well as swing time and stance time. The proposed methodology yielded an accuracy of 95% in tests with 42 patients with predicted output of fast, normal and slow walk.

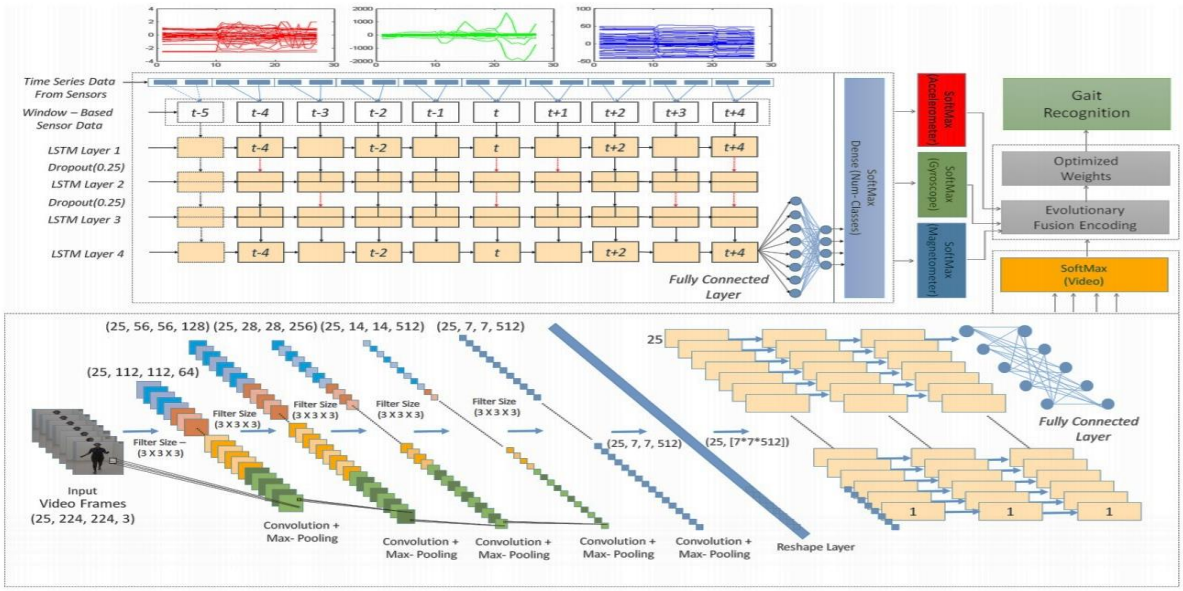



Figure 3.7. 3D CNN+LSTM architectural detail of the proposed multimodal human gait recognition using VS and WS fusion [116].

Gait recognition from two modalities has been proposed in studies of the lower limb trajectory by fusing VS and WS features extracted by deep learning network layers. Kumar et al. [116] proposed evolutionary 3DCNN+LSTM (see figure 3.7) to extract features captured by VS and WS with IMU and pressure sensor signals. The system used to capture gait consists of a video camera, 17 precision IMU nodes and two pressure insoles included in a Shadow Motion wireless body suit. The proposed methods were tested on 19 males and 4 females performing four different walking styles, namely normal-walk, fast-walk, walking while listening to music and walking while watching video on mobile. The CNN is utilized to extract gait spatiotemporal features from VS. Two LSTMs models were used: one to process the CNN output and the other to extract gait spatial-temporal features from the WS. In the final stage, a Grey Wolf Optimizer [117] is used to fuse the LSTMs outputs. The model achieved an average accuracy of 91.3% on gait labelling.

It is also worth noting that Vera-Rodriguez et al. [118] have suggested that their fusion of FS and VS modalities, implemented on handcrafted features as an input to statistical analysis methods, is amenable to deep learning methods for automatic extraction of fused features to improve the accuracy.

## 3.5 Discussion

Gait analysis does not benefit from the advantage of deploying traditional and proven methods, such as harmonic analysis where functions are represented in a full and orthogonal base, e.g. as a superposition of sine and cosine functions. Unfortunately, a full and orthogonal function base for gait is difficult to define, or in other words, the gait primitives are largely undefined. The gait features commonly used in practice (see Section 3.2.1) neither are fully independent nor do they exhaust all the possible members of the set. On this backdrop, the automatic extraction of those gait features which allow to distinguish a certain target case with the best accuracy, e.g. in healthcare or biometrics, appears to be a winning strategy. However, the nature of human gait requires distinct approaches depending on the character and volume of the recorded sensor data (sensing modality, availability of datasets, computational cost, etc.). This causes variations in the optimal choices made: the modalities for complementary fusion, the data representations, the optimal deep learning models, as well as the manner of their overall deployment.

### 3.5.1 Spatiotemporal Character of Gait Data

Because of the ambulatory nature of human gait, events defining the gait cycle are recurrent in space and time. Common patterns are manifested in a sequence of spatial regions and time periods, lending themselves to methods to suppress noise, e.g. by applying statistics over long data sequences. As the gait cycle duration is in the order of 1 s, acquired time sequences are usually abundantly sampled and require down-sampling to optimise computational resources and improve signal-to-noise. The approach to spatial sampling, however, is much less standard and is strongly modality-dependent (see Section 3.5.2.). Estimating correctly the spatial resolution limit determined by the data is crucial for the choice of suitable data representations and for interpreting the calculated accuracy.

A direct comparison shows that classification with features automatically extracted from fused spatiotemporal FS data (see Section 3.3.3.) yielded better accuracy compared to features from spatially-integrated temporal data or time-integrated footprints reconstructed from the same dataset. Thus the benefits of spatiotemporal data fusion by

deep learning CNNs appear to be beyond doubt. However, that has been achieved either at the data representation level or at the classification score level and not in the deep layers of the CNN. Furthermore, when trained on reconstructed footprint images only, deep and shallow models exhibited comparable performance. Arguably, this is because the spatial reconstructions involve solving an ill-posed and ill-conditioned inverse problem. This generated a much larger feature vector of pixel values, compared to the substantially smaller number of values in a single measurement frame, used for spatiotemporal fusion where deep CNNs outperform shallow models.

An interesting consequence of the spatiotemporal character of gait is the drive towards accurate time-stamping. This has resulted in the combination of modalities, such as IMUs plus shoe-sole switches, to detect heel-strike and toe-off for correct labelling of gait phases from WS data.

### 3.5.2 Multi-sensor Incentives for Deep Learning from Gait Data

Manual fusion may take place at the data pre-processing stage when constructing the data presentation, e.g. complementary fusion in the case of RSMs, or collaborative fusion to generate a sinogram image of the Radon transformed FS data [98], which then can be used for data inversion into 'centre-of-mass" coordinates or footstep images (see figure 3.6). In contrast, automatic extraction of fused features may take place in the convolutional layers of the deep ANNs. However, the preference has been to fuse the spatial and temporal components at the classification score level (see Section 3.3.3.)

Deep learning from gait data is most mature for VS, due to the ease of borrowing methodology from well populated research areas, such as face recognition. Translation of approaches such as object segmentation and detection, as well as 'probe and gallery' methods in image recognition are a few examples. Other reasons for the notable progress made with VS is the relative abundancy of reliable databases with variations in the viewing angle and scenes – all facilitated by the variety of ubiquitous sensing technology and fuelled by security and surveillance applications [119]. This also may explain why the drive to fuse VS with other gait modalities is comparatively weak; in fact, VS GEI arguably capture already fused data (see figure 3.2) and gait monitoring uses only some

of that data, as it does not concern with the part used for face recognition. Since VS for gait is less demanding in terms of pixel resolution and close-up, it can be speculated that accurate identification from gait could result in dis-applying facial recognition, wherever practical. Nevertheless, the fusion of VS with other modalities has been successful, mainly in the context of laboratory work (see Section 3.4).

WS offers the widest range of sensor and data types, as well as varying degree of consumer market penetration and level of collaboration by the user. Sensor fusion within the WS subset of modalities is systematically researched to deliver the next generation of health monitoring systems, and to satisfy the growing interests in activity recognition for commercial ecosystems comprising of smartphones, smartwatches, activity bands and other health/fitness monitoring devices. Progress will depend on the technology available, but it is very likely that such devices will use massively deep learning for sensor data fusion in order to personalize their owner's experience.

FS need to cover substantial areas where floor contact may be effected, thus they typically employ collaborative fusion from a large number of identical discrete or distributed sensors and measurands, deep learning from FS data is comparatively new and reports of complementary fusion with other gait modalities have been rare, which should be judged in view of the scarcity of appropriate datasets. Anyhow, the information resulting from sensor fusion with FS would be undoubtedly richer than just by VS and/or WS, thus delivering even more accurate classifications. Fusing FS with VS for security application can help to overcome the challenges of gait recognition in different scenes and limited angular views in VS; reciprocally, VS data can be invaluable to resolve the challenge of simultaneous users captured by FS systems. There is no doubt that fusion of all three gait modalities reviewed here will be more advantageous; however, this has to be measured against the practicality or designing, building, deploying and maintaining such complex systems.

## 3.6 Literature Update

Gait recognition constitutes a significant research field with a steady progression in the variety of studies. It mainly focuses on biometric and gait impairments for possible diagnoses and monitoring the severity of injuries or diseases. This chapter is published

in 2019 and the recent advancement ( with emphasises on machine learning methods only) since the publication of this chapter are update in this section.

### 3.6.1 Video Sequence

In the past two years the focus is to solve the view- and clothing invariant problems for video sequences with more advanced machine learning methods, such as generative adversarial network [120],[121]. Zhang et al. [120] designed a view transformation generative adversarial network (VT-GAN) to translate gaits between any two views only using single model. The proposed VT-GAN model includes three models namely, a generator, a discriminator, and similarity preserver. The model achieved a competitive results on the cross-view gait task using CASIA-B dataset described in section 3.3.1.1. The other use of GAN in the work of Babaee et al. [121] focuses on gait recognition from incomplete gait cycle. GAN is used in this work to reconstruct complete GEIs from incomplete GEIs. The GAN architecture is composed of a generator with an autoencoder to reconstruct a complete GEIs from incomplete GEIs, the other component is a discriminator. The model is evaluated on OU-ISIR dataset (see section 3.3.1.1) and achieved a competitive results in the case of incomplete gait cycle. A more recent use of GAN is in the work of Chen et al. [122], the proposed Multi-View Gait GAN (MvGGAN) for cross-view gait. The models are evaluated on CASIA-B and OUMVLP [123] datasets and achieving improved results for gait recognition performance in reality scenes.

### 3.6.2 Wearable and floor sensors.

The recent work on wearable and floor sensors has been applied for medical applications such as the impact of muscle fatigue on gait characteristics [124], health monitoring [125] and age-related differences [126]. In recent work on wearable sensors by Turner et al. [127], LSTM network is proposed to process and classify pressure sensors signals. The sensors were placed inside the shoes and participants were asked to walk eight walking trails. The aim of this study was to analyse artificially induced gait alterations. The results are promising for a potential use in the diagnosis of gait abnormalities or other neuromuscular movement disorders in patients. In a recent work on wearable sensors. Tran et al. [128], proposed multi-model LSTM and CNN to classify IMUs spatiotemporal

signals. The proposed models outperformed previous results on the whuGAIT [129] and OU-ISIR [130]datasets using hybrid network.

## 3.7 Conclusion

The character of gait data poses the problem of identifying features suitable for gait classifications, desirable in a number of application areas. The three gait-sensing modalities covered in this Review have produced data which is most amenable to the use of deep learning, to address the automatic extraction of such features. Deep learning CNNs typically outperform statistical analysis models in the most essential metrics. Furthermore, multi-sensor and multi-modality fusion results in better accuracy and robustness. This is achieved by employing the available flexibility in data representations, ANN architectures and the choice of model hyper-parameters. Gait analysis benefits from methods introduced and tested in other applications of deep learning. However, it requires particular attention due to its spatiotemporal character, the options for ubiquitous gait sensing and the privacy concerns they raise, as well as the cost of achieving research, development and commercialization objectives. Deep learning from multi-sensor, multi-modality gait data offers new options in the strong drive towards personalized healthcare, as well as towards more robust and un-intrusive biometrics for safety and security. These are some of the challenges of the day, but the state-of-the art indicates a promising step reaching further into the future, rather than just the current horizon.

# Chapter. 4 Parkinson's disease Gait

# Gait Spatiotemporal Signal Analysis for Parkinson's Disease Detection and Severity Rating

**Abdullah S. Alharthi, Alexander J. Casson, and Krikor B. Ozanyan**

## Abstract

Deep learning models are used to process and fuse raw data of gait-induced ground reaction force (GRF) for Parkinson's disease (PD) patients and healthy subjects with the aim to categorize PD severity. This is achieved by learning automatically, end-to-end, the spatiotemporal GRF signals, resulting in an effective PD severity classification with mean performance F1-score of 95.5% and F1-score standard errors of 0.28%. Layer-wise relevance propagation (LRP) is used to interpret the models' output and provide insight into which features in the spatiotemporal gait GRF signals are most significant for the models' predictions. This allows their assignment to gait events, implying that while for the classification of healthy gait the heel strike and body balance are the most indicative gait elements, foot landing and body balance are those most affected in advanced stages of PD. The proposed models are resilient to noise and are computationally efficient for processing and classification of large longitudinal GRF signal datasets, therefore they can be useful for detecting deterioration in the postural balance and rating PD severity.

Manuscript published at the IEEE Sensors Journal. Received July 22, 2020; accepted August 15, 2020. Date of publication August 20, 2020; date of current version December 16, 2020. The work of Abdullah S. Alharthi was supported by the Government of Saudi Arabia. The associate editor coordinating the review of this article and approving it for publication was Dr. Marco J. Da Silva. (Corresponding author: Abdullah S. Alharthi.) The authors are with the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M1 3BU, U.K. (e-mail: abdullah.alharthi@manchester.ac.uk; alex.casson@manchester.ac.uk; k.ozanyan@manchester.ac.uk). Digital Object Identifier 10.1109/JSEN.2020.3018262.

# 4.1 Introduction

Gait is the manner of walking in humans, it is produced by the biphasic forward propulsion of the center of gravity of the human body to move forwards. Gait is achieved by the synchronized recurrent contact of the feet with a surface, produced by the lower limb's movement. Gait analysis is an active research area for a plethora of tasks in many applications including sports, biometrics, and healthcare [131]. Quantitative gait analysis has received attention in recent years in the context of cognitive decline due to aging and the progression of disease, e.g. freezing of gait [132], irregular step in Parkinson's disease (PD) patients [82], and prediction of falls in older adults [93]. PD is the second most common neurodegenerative disorder, caused by premature death of neurons in the midbrain [133]. A definitive pathological diagnosis of PD is limited by the lack of reliable neuropathological criteria. Historically, 80–90% of the clinical diagnoses have been confirmed at autopsy, thus the time from initial diagnosis until death can be long [134]. PD has been associated with cancer [133], and 25–40% of PD patients eventually develop dementia, which leads to reduced life expectancy [134].

Today, clinical practitioners examine PD patients visually depending on established criteria to deliver evidential PD diagnosis and severity ratings, this is sometimes followed by a survey in which PD patient subjectively examined for mentation, behavior, and mood. The most widely used rating criteria are Unified PD Rating Scale, Hoehn and Yahr staging, and the Schwab and England rating of activities of daily living [135]. Deviation from healthy gait is one of the critical implications of PD, thus advanced stages increase falls risk [135]. However, in the early stages of PD visual gait evaluation may lead to inconclusive results. This is mainly due to the manifested slow walking and short stride being potentially also related to age, depressive mood, or other conditions [136]. Consistent with this, PD causes tremor, muscle stiffness and slowness of the movement. The gait cycle (see figure 3.1) is routinely monitored visually in health care practices to inspect gait abnormality, and form the basis of PD diagnosis, as well as the rating of severity of the disease. Thus these methods are semi-subjective.

To make the methods less subjective, sensing technology development has presented various sensor choices for capturing spatiotemporal information of gait. The hypothesis is that these could be used to aid PD diagnosis and rate the severity. In particular, the

ground reaction force (GRF) resulting from floor contact is a quantity often used for complex recordings of gait intervals and events, due to its resilience to variations in clothing, illumination and availability of a direct view onto the subject [131]. However, in the literature to date, GRF analysis has employed manual feature-extraction methods [35], [136], [137],[138] which rely heavily on domain knowledge of how to utilise such features' information content for PD diagnosis. These established methods may not capture the maximum variability of the raw spatiotemporal GRF signal from a practical application; they are prone to artefacts due to different sensing modalities and experimental conditions, as well as subject to choice by the human experts involved.

The recent advance in deep learning offers an alternative approach to processing GRF signals allowing to achieve reliable gait classifications. Deep Artificial Neural networks (ANNs), based on raw sensor data, avoid the use of subjectively engineered features by implementing automatic feature extraction [131]. This results in a highly attractive method for analysing gait data, demonstrated by the application of Long Short Term Memory (LSTM) deep networks for the analysis of gait affected by PD [139]. However, a well-known limitation of basic deep ANNs is that they do not allow a straightforward link back to what features in the input data were instrumental for the models to arrive at a particular decision [140]. This limits their cross-disciplinary reach, since opacity hinders the understanding [20] of the model predictions in terms of the individual disciplines' state-of-the-art [21]. It also means that information on the utility of classification is difficult to feed back to the sensor design and data processing community, thus hindering systematic improvement of gait characterization.

To overcome this issue, a machine learning method, layer-wise relevance propagation (LRP) [20] has recommended itself for demonstrating the relevance of deep ANN predictions to the input data. By using the relevance conservation principle ([141], [20]) it produces a relevance map, attributing portions of the model predictions to the raw input signal using backpropagation methods. This allows areas of importance to the classification to be identified in the input signal, and possibly used to interpret the model's predictions. LRP has shown notable success in image classification [142], [143], as well as in subject identification based on gait [144]. Further, region perturbation based on the LRP maps is an essential step for the evaluation of the classification result. This

approach tests the validity of a deep ANN to generalize on unseen data, as well as the adequacy of LRP maps.

This article presents a deep convolutional neural network (DCNN) for the analysis of GRF data from Parkinson's patients. This builds upon our previous work [145] limited only to the classification performance of a DCNN on GRF data with LSTM and statistical analysis. The hypothesis is that we can use spatiotemporal gait GRF signals to classify participants as a healthy subject or PD patient with a rating the severity of the disease and link the key known events in the gait cycle to cognitive deterioration due to PD severity. To our knowledge this is the first application of the LRP approach to the analysis of spatiotemporal gait GRF signals in PD. In addition, we validate the methods for how effective the generated classification result is, by applying perturbation noise using LRP relevance scores. Section 4.2 describes the technical background of the DCNN and the LRP method used, together with the GRF dataset obtained from PD patients. Our specific implemented networks, their classification performance, and LRP relevance maps are then presented in Section 4.3. Finally, our results are discussed in Section 4.4.

## 4.2 DCNN AND LRP Background

### 4.2.1 Deep Convolutional Neural Network (DCNN)

Supervised learning is by far the most widely used artificial intelligence algorithm for gait analysis [131]. The main objective of such machine learning methods is to find the gait pattern by training a network on supervisory mapping of ground truth input data to an output label. Here the ground truth label is healthy or PD with Hoehn and Yahr staging [135]. In this process, the models learn the different stages of gait deterioration.

State-of-the-art DCNNs have established themselves in a variety of classification tasks, providing new insights into complex spatiotemporal signals. The models learn a high level of abstraction and pattern by applying convolution operations to extract features and perform classification. The core building block of a DCNN model is the convolution operation. A DCNN model is built by combing multiple layers of this convolution operation, pooling's layers and fully connected layers with interconnections.

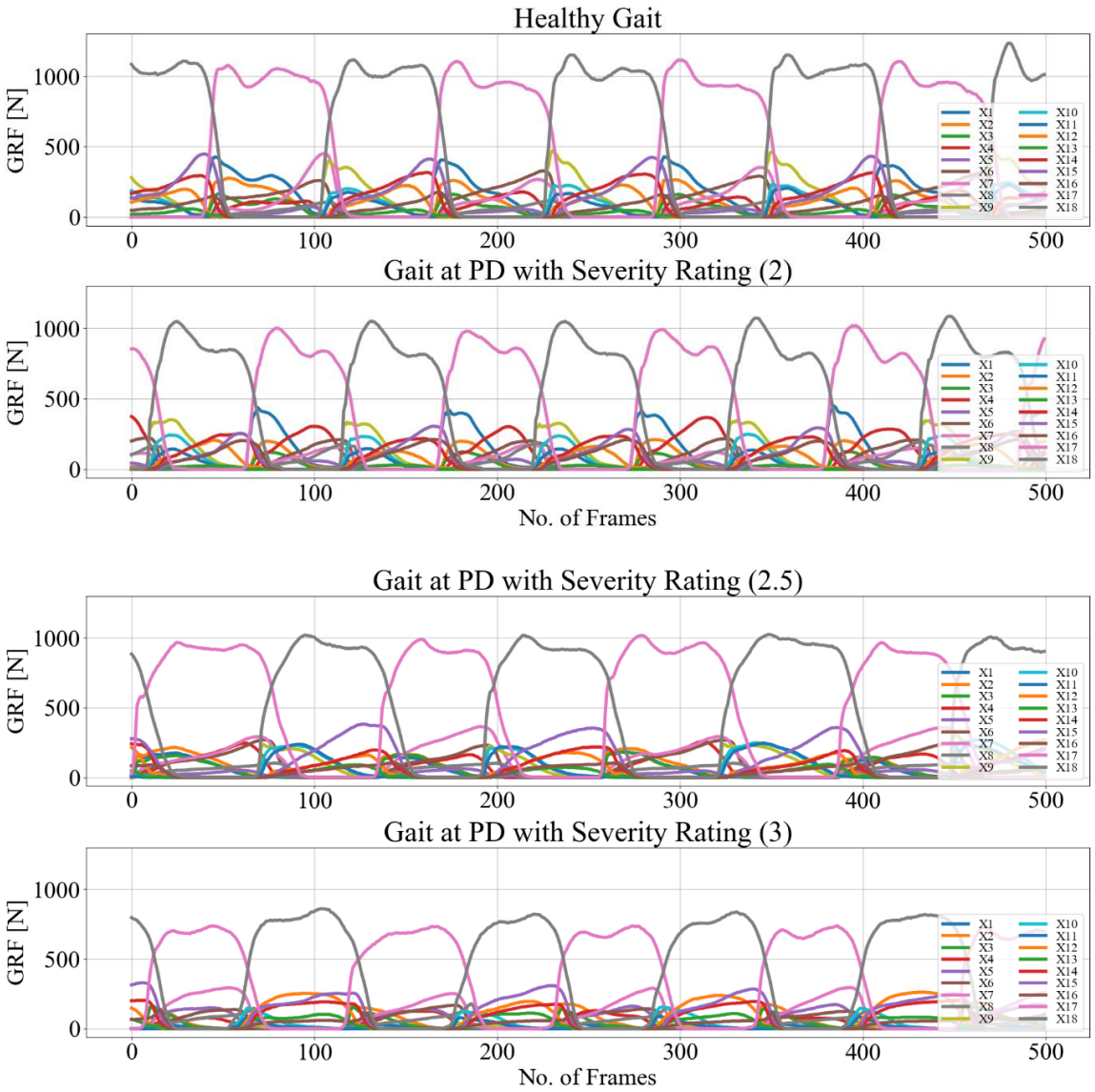



Figure 4.1. Example GRF data recorded at 100 frames per second for healthy subject and subjects with PD severity ratings 2, 2.5, and 3, with sample length of 500 time-frames $x_i = [x_1 \quad \cdots \quad x_{18}] \in \mathbb{R}^{500\times18}$. The signals with lower amplitude ($x_i = [x_1 \quad \cdots \quad x_{16}]$<500 N) represent pressure sensor signal under each foot (different colors for each of the 8 sensors). In each sample, the calculated sum of the 8 sensor outputs for each foot is also shown additionally ($x_{17}$&$x_{18}$>500 N, different colors for left and right foot).

## 4.2.2 Layer-Wise Relevance Propagation (LRP)

LRP [20] is a backward propagation method which identifies which parts of the ANN input vector carry most weight in the model prediction. In this paper we quantify the

contribution of a single component of an input $x_i$ (in our case, a sensor signal at a specific time frame) to the prediction of $f_c(x)$ ( $c$ denote a class of PD severity rating, based on gait) made by the DCNN classifier $f$. The outputted PD severity rating prediction is redistributed to each intermediate node via backpropagation until the input layer. The LRP outputs a "heat map" over the original signal to highlight the signal sections with the highest contributions to the model prediction, e.g. the data sections with the maximum variability given the classes.

We first note that a neural network consists of multiple layers of neurons (feature maps in the case of a convolution layer), where neurons are activated as follows [21]:

$$a_j^{1+l} = \sigma\left(\sum_i a_i^l \, \omega_{ij}^{(l,l+1)} + b_j^{(l+1)}\right) \tag{4.1}$$

Here, $a_i^l$ denote the activation of a neuron $i$ in the previous layer in forward direction; $\omega_{ij}^{(l,l+1)}$ denote the weight received in forward direction by neuron $i$ from neuron $j$ in the previous layer; $a_i^l \, \omega_{ij}^{(l,l+1)}$ denote the contribution of neuron $i$ in layer $l$ to the activation of the neuron j in layer $l + 1$. The sum is computed over all neurons that are connected to neuron $j$. The function $\sigma$ is a nonlinear monotonously increasing activation function, a rectified linear unit ReLU (max(0,x)) and $b_j^{(l+1)}$ is a bias term. These weights and biases are learned by the DCNN during supervisory training. During training, the output $f_c(x)$ is evaluated in a forward pass and the parameters ($\omega_{ij}^{(l,l+1)} + b_j^{(l+1)}$) are updated by back-propagating using model error. For the latter, we base our computations on categorical cross entropy.

The LRP approach decomposes the DCNN output for a given prediction function of gait class $c$ as $f_c$ for input $x_i$ and generates a "relevance score" $R_i^l$ for $i^{\text{th}}$ neuron of layer $l$ and $R_j^{l+1}$ from the $j^{\text{th}}$ neuron in the previous layer, where the relevance conservation principle is satisfied as:

$$\sum_i R_{i \leftarrow j} = \sum_j R_j = f_c(x). \tag{4.2}$$

The LRP starts at the DCNN output layer after removing the *Softmax* layer. In this process, a gait class $c$ (e.g. healthy gait or one of three PD severity ratings) is selected as an input to LRP and the other classes are eliminated. The backpropagation for unpooling the

pooling layer is computed by redirecting the signal to the neuron for which the activation was computed in the forward pass. In the case of average-pooling, the neuron which has the average activation value is considered.

As a generalization, consider a single output neuron *i* in one of the model layers, which receives a relevance score $R_j$ from an upper layer neuron *j,* or the output of the model (class *c*). The scores are redistributed between the connected neurons throughout the network layers, based on the contribution of the input signals $x_i$ using the activation function (computed in the forward pass and updated by back-propagating during training) of neuron *j*. The latter will hold a certain relevance score based on its activation function and passes its value to consecutive neurons in the reverse direction. Finally, the method outputs relevance scores for each sensor signal at a specific time frame. These scores represent a heat map, where the high relevance scores at specific time frames highlight the areas that contributed most to the model classifications. The relevance propagation rule $R_i$ for layer *l* can be represented as [20], [22]:

$$R_i^l = \sum_j \frac{a_i^l \omega_{ij}^{(l,l+1)}}{\sum_i a_i^l \omega_{ij}^{(l,l+1)}} R_j^{(l+1)} \tag{4.3}$$

There are other propagation rules such as (αβ-rule) [21],[22]:

$$R_i^l = \sum_j (\alpha \cdot \frac{(a_i^l \omega_{ij}^{(l,l+1)})^-}{\sum_i (a_i^l \omega_{ij}^{(l,l+1)})^-} + \beta \frac{\left(a_i^l \omega_{ij}^{(l,l+1)}\right)^+}{\sum_i (a_i^l \omega_{ij}^{(l,l+1)})^+} .) R_j^{(l+1)} \tag{4.4}$$

where $(a_i^l \omega_{ij}^{(l,l+1)})^+$ and $(a_i^l \omega_{ij}^{(l,l+1)})^-$ denote the positive and negative part of $a_i^l \omega_{ij}^{(l,l+1)}$respectively, such that $(a_i^l \omega_{ij}^{(l,l+1)})^+ + (a_i^l \omega_{ij}^{(l,l+1)})^- = a_i^l \omega_{ij}^{(l,l+1)}$ in equation 4.4, and the parameters α and β are chosen so that $\alpha - \beta = 1$ and $\beta \geq 0$. Other stabilizing terms can be used to avoid divisions by zero as explained in [20].

## 4.2.3 Perturbation

Human gait is inconsistent between individuals, and even for a single individual, requiring models to be reliable and robust to variance in the input data. Applied to the LRP analysis, the interpretation of important input data points needs to be robust to noise and variance in the input data stream. Considering this, a random perturbation noise analysis on the LRP relevance score can aid in the choice of LRP method, as well as in

designing a DCNN model resilient to noise due to the inconsistency of gait. This perturbation analysis is achieved as described in the following sections.

The "greedy" iterative procedure proposed in [22] allows to choose the appropriate LRP method and to evaluate the quality of gait classification relevance scores. This is achieved by step-wise removal of information from the spatiotemporal input signal. At each step, regions with the highest relevance scores are replaced by Gaussian noise by a "most relevant first" (MoRF) approach [22]). The change in the model performance is then monitored at each step by running the model to re-predict the test data with the perturbation accumulated by that step. The most desirable LRP method is the one with the strongest drop in accuracy at the first few steps [187], where the most relevant information is removed by the perturbations, and slower decline further as less important regions are removed. The accuracy drop is quantified using the area over the most relevant perturbation curve (AOPC) [22].

In terms of assessing the significance of DCNN model architecture, this task is achieved by progressively removing the highest relevance scores yielded by the best LRP method selected using the method described above and testing the model performance by re-predicting on the test data for each model. Models which substantially drop performance after only few perturbation steps are considered to be the most amenable to exploiting LRP. This is because the decline in performance allows to assert that those few removed regions are critical for accurate classification performance, therefore it is indicative of meaningful relationships between input patterns and learnt classes. In contrast, if removing a region with little impact on the classification performance, the implication is that it is of lesser interest in terms of seeking such relationships.

Table 4.1 Number of subjects with the severity rating.

| Severity (0) Healthy | Severity (2) | Severity (2.5) | Severity (3) | Group |
|---|---|---|---|---|
| 18 | 15 | 8 | 6 | Ga[146] |
| 26 | 12 | 13 | 4 | Ju [147] |
| 29 | 29 | 6 | 0 | Si [148] |

### 4.2.4 Dataset

To assess GRF data from PD patients we used the open access benchmark from PhysioNet.org [149]. It consists of data from 93 PD patients (mean age: 66.3 years; 63% men), as detailed in table 4.1, with different levels of PD progression, as detailed in table 4.2. Data from 73 healthy controls (mean age: 66.3 years; 55% men) are also present.

The dataset consists of GRF measurements collected as participants walked for approximately two minutes. Each subject had eight sensors placed underneath each foot to measure force [N] as a function of time. The output of the 16 sensors was recorded at 100 frames per second. Also, the sum of the eight sensors of each foot is added to each subject sample and the timestamp, yielding 19 columns in total. The data set was collected by three research groups, namely: Ga group [146], Ju group [147] and Si group [149] with the sub-parts of the dataset named after these groups. The Ju and Si groups recorded usual healthy walking at a self-selected speed. The Ga repeated this, and included additional samples for each subject, where they performed dual task while walking [146].

## 4.3 DCNN Implementation and Results

### 4.3.1 Data Pre-processing

Each sample recorded in the dataset contains 19 columns of data (see 4.2.4) with varying column length, as for some subjects gait was recorded for a longer time (12119 frames) than for others (less than 1000 frames). In order to make the input data length consistent, the datasets were split to equal size parts of 500 frames such that single long recording are divided to several chunks of 500 frames. The timestamp columns were deleted as it doesn't report information about gait. The final sample size is 18 columns and 500 rows or frames, as shown in figure 4.1. This choice is justified since the gait cycle is approximately one second and the sample captures heel strike and toe off for both feet over five gait cycles. The input dataset is a tensor with dimensions $m$×500×18 where m= 2698 for Ga group, 2198 for Ju and 1509 Si group (see example sequences in figure 4.1). The input is reshaped for the 2D-DCNN as K×50×15×12 building upon our previous work with different inputs and algorithms [145].

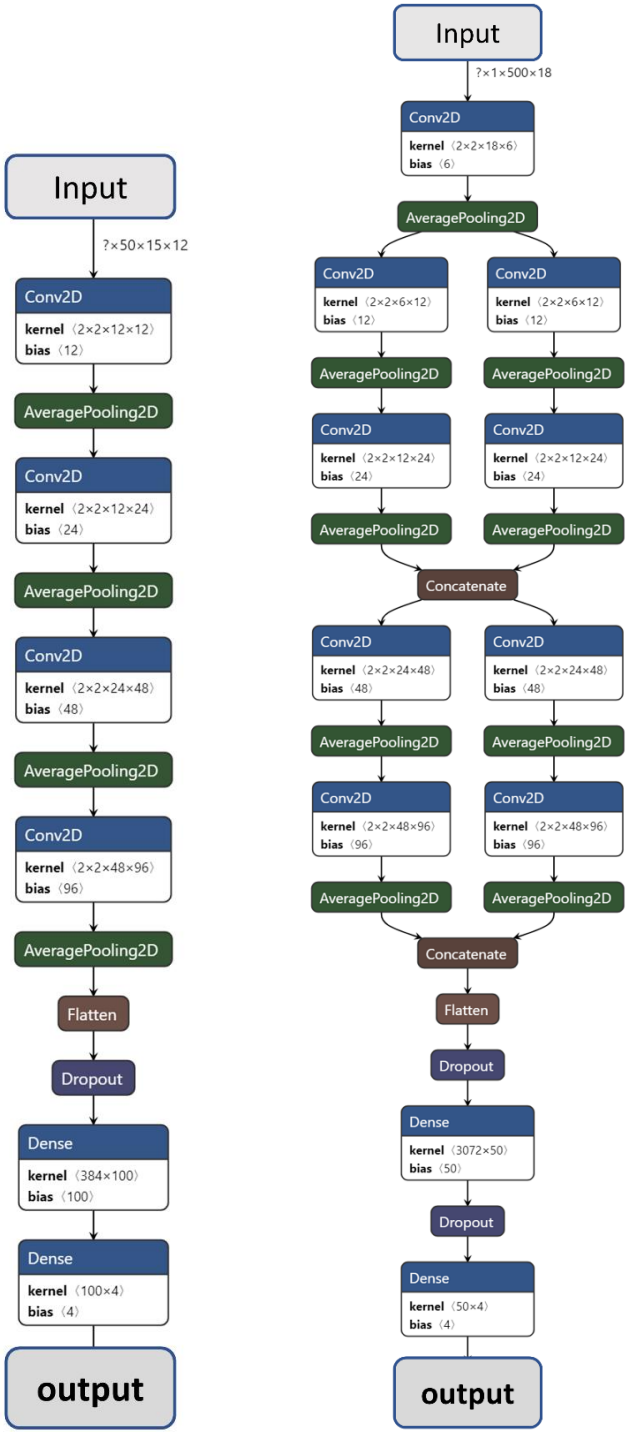
Table 4.2. Datasets subject's discerption.

| Subjects | Number | Male | Female | Group |
|---|---|---|---|---|
| PD patients | 29 | 20 | 9 | Ga [146] |
| Healthy Subjects | 18 | 10 | 8 | Ga [146] |
| PD patients | 29 | 16 | 13 | Ju [147] |
| Healthy Subjects | 26 | 12 | 14 | Ju [147] |
| PD patients | 35 | 22 | 13 | Si [148] |
| Healthy Subjects | 29 | 18 | 18 | Si [148] |

Data standardization is performed as a pre-processing step to reduce the redundancy and dependency among the data, such that the estimated activations, weights, and biases will update similarly rather than at different rates during the training process. The standardization involves rescaling the distribution of values with mean at zero and rescaling the standard deviation to unity such that $x_i^{st} = \frac{x_i - \mu(x_i)}{\vartheta(x_i)}$. Here $x_i^{st}$ is PD data rescaled such that $\mu(x_i)$ is the mean values of $x_i$ and $\boldsymbol{\vartheta}$ is the standard deviation of $x_i$. Next, the dataset is randomly split into training 60%, hold-out validation 20% and testing 20% with a *random state* parameter with different seed.

## 4.3.2 Proposed DCNN Architectures

In this work the input to the DCNNs, used to process and classify gait GRF signals, is the spatiotemporal 3D matrix of the raw sensor signals. An extensive experimentation with several DCNNs models' hyper-parameters in our previous work [145],[159],[151] suggests three different networks architectures, shown in figure 4.2. All networks are implemented using the Python Keras open source library [152] and iNNvestigate GitHub repository [153] for LRP. All codes are computed using a Dell Precision T5600 workstation.




(a) (b)

Figure 4.2. Proposed DCNN architectures; a) Single DCNN, b) Parallel DCNN. The color-coding of boxes: convolution layers and fully connected layer (blue); pooling layers (green); concatenation layers and flattening layers (brown); dropout layers (navy); input at the top and softmax output layers (gray). The diagrams are generated using Neutron from GitHub repository based on the models weights and biases.

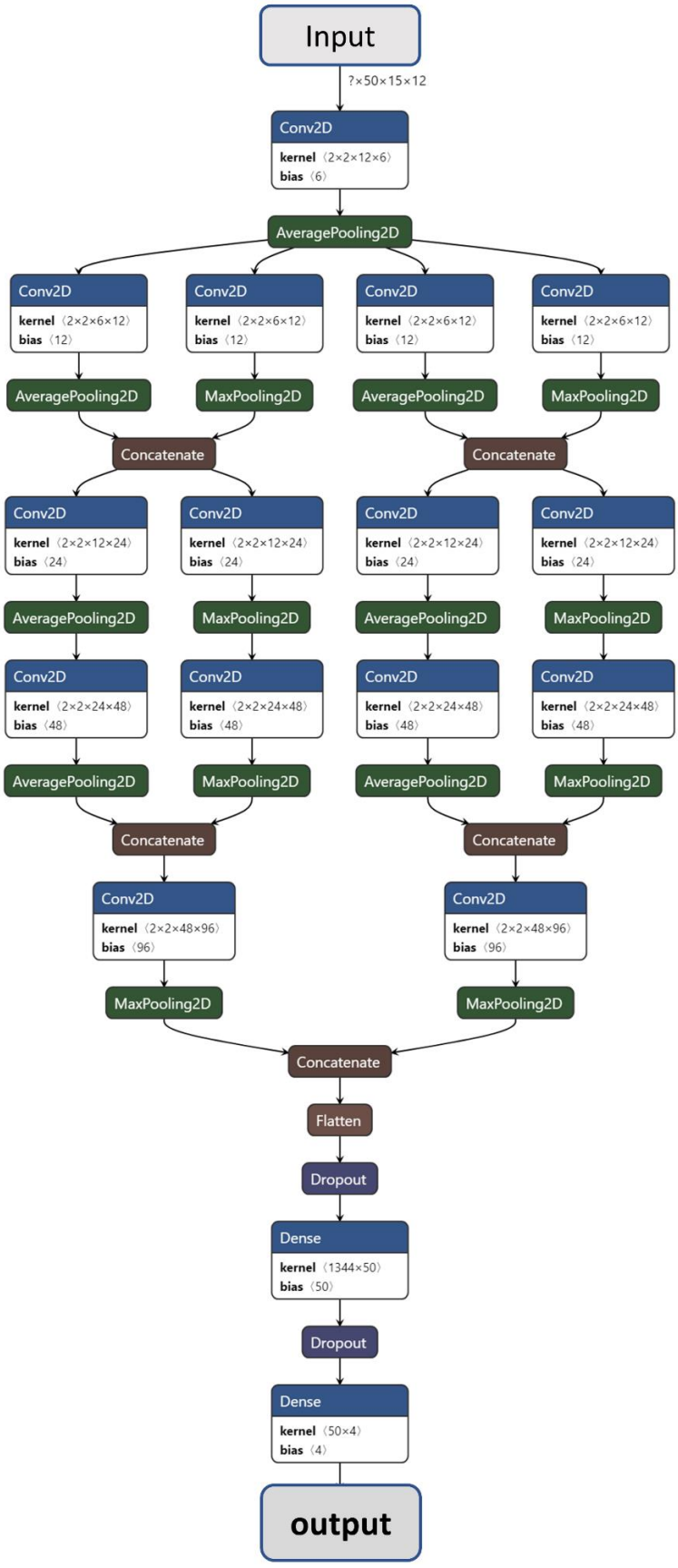



Figure 4.3. Proposed DCNN architectures Quadruplets DCNN.

#### 4.3.2.1 DCNN

A 2D-DCNN model (figure 4.2(a)) built for PD severity classification consists of four convolutional layers, each followed by an average pooling and two fully connected layers, yielding a total of 10 stacked layers. The four convolutional layers have *n* channels each that assign one frame of the input to a single convolutional layer channel. The convolutional layers use a stride of 1, same-padding, and a 2 x 2 filter for the average-pooling layers.

#### 4.3.2.2 Parallel DCNN

This network architecture (figure 4.2(b)) has been proposed specifically in our previous work [150] to process GRF signals. It is inspired by inception neural network architectures [154], the aim is to have filters with learnable parameters operate on the same level to recognize the salient parts in the sample. The network consists of two stages with parallel streams fused with concatenation layers, where each stream has its weights and biases launched uniformly and updated during training via backpropagation. The network is topped up with fully connected layers and a softmax layer, giving a total of 18 stacked layers.

Table 4.3. Models F1-Score for each Dataset and F1-Score, Mean and Standard Error with Datasets Combine.

| CNN Model | Ga | Ju | Si | GąUJuUSi Seed 42 | GąUJuUSi Seed 100 | GąUJuUSi Seed 200 | GąUJuUSi Seed 2020 | GąUJuUSi Mean performance | GąUJuUSi Standard error |
|---|---|---|---|---|---|---|---|---|---|
| Single | 98% | 98% | 98% | 95% | 93% | 96% | 96% | 95% | 0.70 % |
| Parallel | 96% | 97% | 96% | 96% | 95% | 95% | 96% | 95.5% | 0.28 % |
| Quadruplet | 97% | 97% | 98% | 95% | 94% | 94% | 95% | 94.5% | 0.28 % |

#### 4.3.2.3 Quadruplet DCNN

The quadruplet network shown in figure 4.3 is original model and its architecture implemented with multi-parallel streams by generalizing Siamese [155] and triplet networks [156]. This network consists of convolutional layers, max-pooling and average-pooling with each stream has its activations, weights, and biases launched uniformly and

updated separately via backpropagation. The goal of this network is to learn the spatial sensor signals and temporal signals separately and simultaneously (similar to inception neural network [154]), using two types of pooling layers. This allows the capturing of gait pattern and the ability of the network to generalize on unseen data.

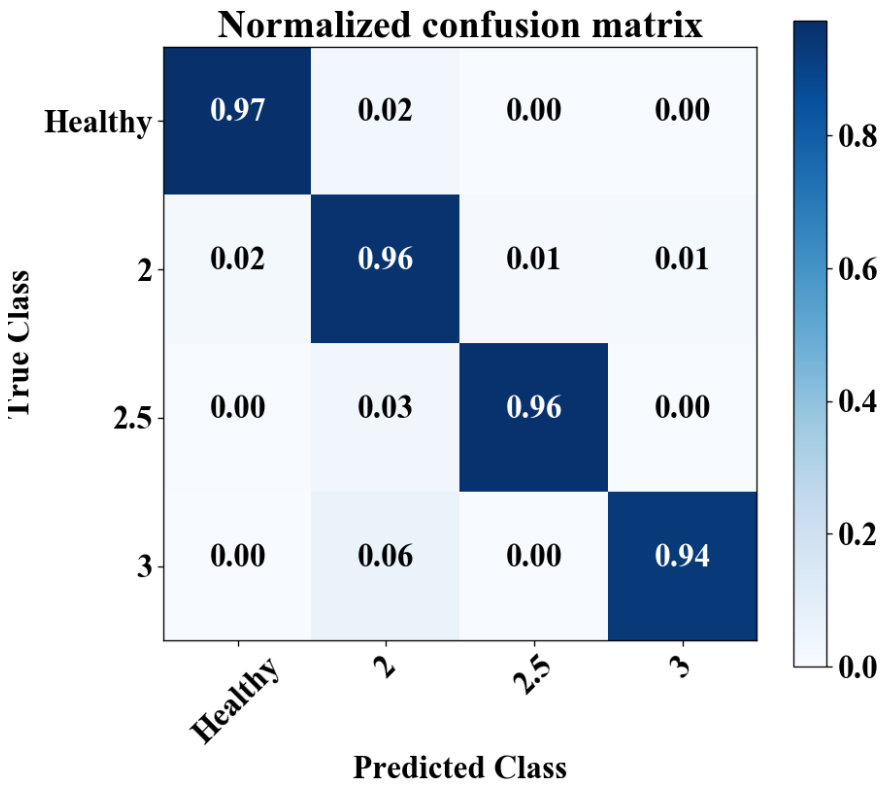



Figure 4.4. Single 2D-CNN predictions on 1281 sample is shown as confusion matrix.

### 4.3.3 Feature Learning and Classifications

The three models are trained, validated and tested separately, using a batch size of 200 samples for each iteration, 200 epochs are optimal to train the models, determined by a trial way (on the three datasets combined) and the error using categorical cross-entropy. A method for stochastic optimization (ADAM [157]) is used to train the proposed models. The optimizer parameters are adjusted as follows: $\alpha$ =0.002, $\beta 1$ =0.9, $\beta 2$ =0.999, $\varepsilon$=1e-08. Where $\alpha$ is the learning rate or the proportion that weights are updated; $\beta 1$ is the exponential decay rate for the first moment estimates; $\beta 2$ is the exponential decay for the second-moment estimates; $\varepsilon$ is a small number to avoid any division by zero in the implementation.

The loss computed by the categorical cross-entropy in every iteration is used to validate the models and update the weights and biases. To improve the model's performance a regularization method is utilized together with dropouts, as shown in figure 4.2. and 4.3. The models are trained, validated and tested, three times separately (for each dataset ) and four times(datasets combined) with different random state, to test the models' ability in classifying the three datasets and with the later combined. Accuracy is reported as confusion matrices (Figure 4.4, 4.5, 4.6), as well as precision, recall and f1-score.The mean performance and standard errors of the different random state runs are reported in table 4.3.

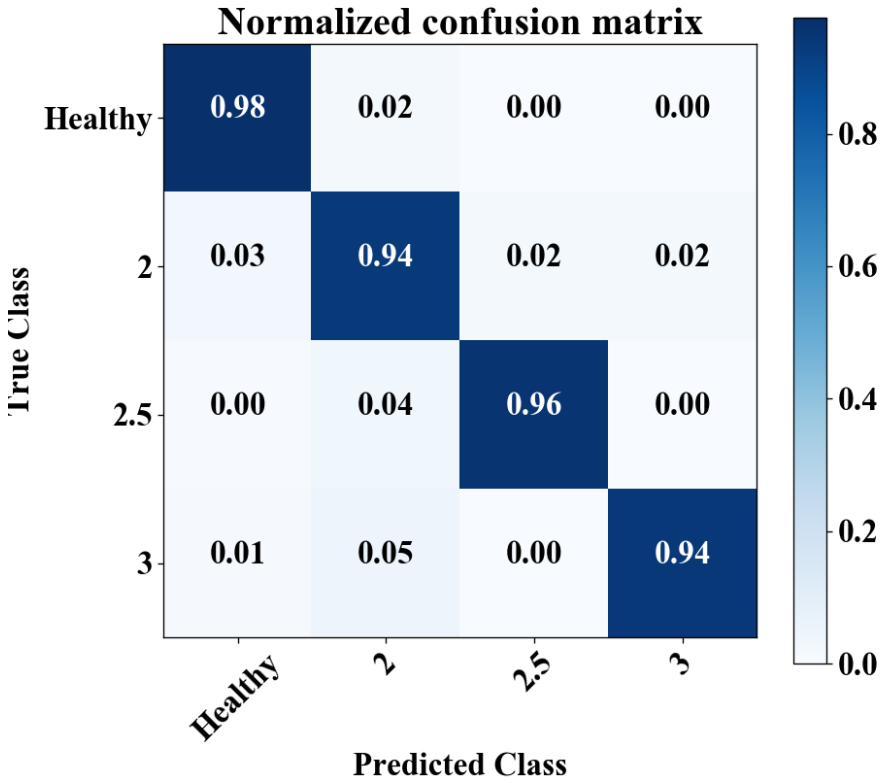



Figure 4.5. Parallel 2D-CNN predictions on 1281 sample is shown as confusion matrix.

## 4.3.4 LRP and Model selection

A number of LRP methods were tested with the three DCNNs models. All the models returned the same results; therefore we report the result from single DCNN, to identify the best performing backpropagation method implemented in iNNvestigate GitHub repository: Deep Taylor [21], Deep Taylor bounded [141], deconvnet (deconvolution)

[158], guided backprop (guided backpropagation) [159], and LRP sequential preset a flat (LRP-SPF)[141]. The DCNN classification accuracy is evaluated for each of the above LRP methods separately, by performing a sequence of perturbation steps, as described in section 4.2.3. (progressively replacing MoRF regions of size 7×7, representing 0.544% (0.00544 = 7×7/ 50×15×12) of the input stream, with Gaussian noise), and observing for each LRP method the cumulative change in the model performance. The baseline for comparison is established by replacing regions of the input data with random Gaussian noise regions instead of replacing the regions based on LRP methods. Next, we subtract the LRP maps accuracy from the input with the randomly replaced regions accuracy to show only the LRP accuracy change. As shown in figure 4.7, in our case the LRP curves recover after around the 15th perturbation step, because the remaining spatiotemporal regions are less and less relevant and the baseline accuracy is reached around the 25th perturbation step, as then all remaining regions are unimportant for the classification. As expected, the exhibited rate of change is proportional to the importance of the information perturbed at each step.

The choice of a DCNN model most suitable for LRP is justified by utilizing the same MoRF protocol [22], whereby each of the three DCNN models is perturbed step wise with Gaussian noise by progressively replacing MoRF regions of size 7×7 and re-predict gait class for 100 steps. In contrast with the LRP method selection illustrated in figure 4.7, instead of comparing with a baseline, here the rate of decline in accuracy (by removing the mean to show only the rate of change) with subsequent perturbation steps is used to identify the model with the steepest drop based on each model returned 100 classification accuracies, as manifested in figure 4.8. This prediction drop takes place at a faster rate in models where the classification uses data patterns in more compact regions within the gait cycle sequence. This allows more straightforward identification of the gait cycle events corresponding to such regions, with reference to the standard cycle presented in figure 3.1. Among the three proposed models, the Parallel DCNN (see figure 4.2.b) experiences the steepest decrease in accuracy with perturbation, as figure 4.9 shows, therefore is the preferred candidate to attempt the identification of gait events most vulnerable to gait deterioration due to PD. Table 4.3 summarises the performance of the three DCNN models.

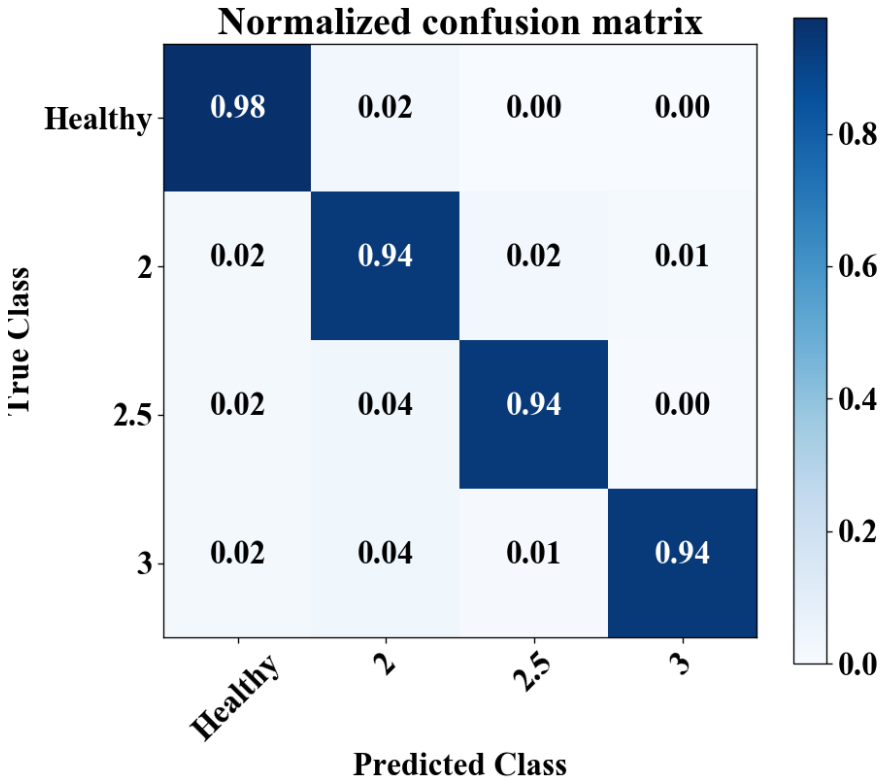



Figure 4.6. Quadruplet 2D-CNN predictions on 1281 sample is shown as confusion matrix.

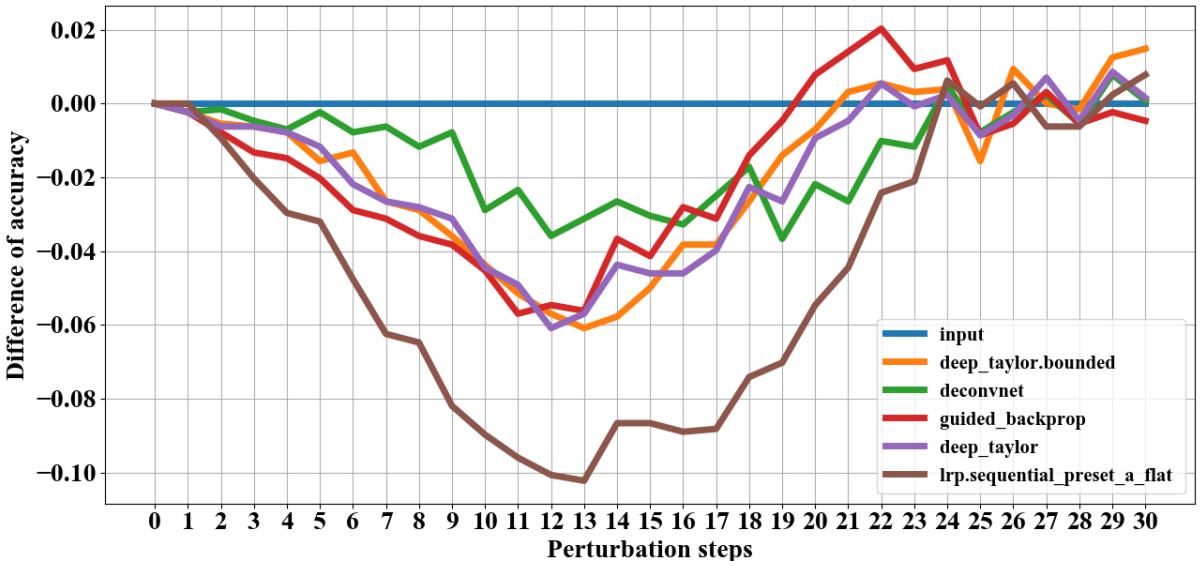



Figure 4.7. LRP method selection by perturbation steps progressively removing information with the highest relevance scores. Steeper initial decrease indicates better identification of gait events with most weight in the classifications.

### 4.3.5 Gait Event Assignment Using LRP

Gait GRF data take the form of periodic sequences which are characterized as repetitive cycles for each foot. We note that the normal gait cycle is initiated by the heel strike of one foot, followed by other gait events described in figure 3.1, in strict order. Therefore, the LRP-generated heat map of the temporal variations in the GRF signal can reveal which events in the gait cycle are most relevant for the classifications. Consequently, gait event assignment is best performed on the data sequences in figure 4.1 after spatial averaging and standardization. A representative spatially averaged sensor signal sequence is shown in figure 4.10 (a) for a healthy subject. The highlighted gray area corresponds to one gait cycle, while the plotted signal is given by the Spatial Average (SA) metric, computed as:

$$SA[n] = \frac{1}{18}\sum_{i=1}^{18}(x_i[n] \tag{4.5}$$

Here $x_i$ are the readings from individual sensors and $n$ enumerates the frames in each sample. Recall that each foot has 8 sensors attached (16 total) and the two sums one for each 8 sensors for each foot is available giving 18 signals in total. figure 4.10 (b) shows the expanded gait cycle from figure 4.10(a) with the gait events color-coded and labelled as per figure 3.1. These labels are used as references in section 4.4. below.

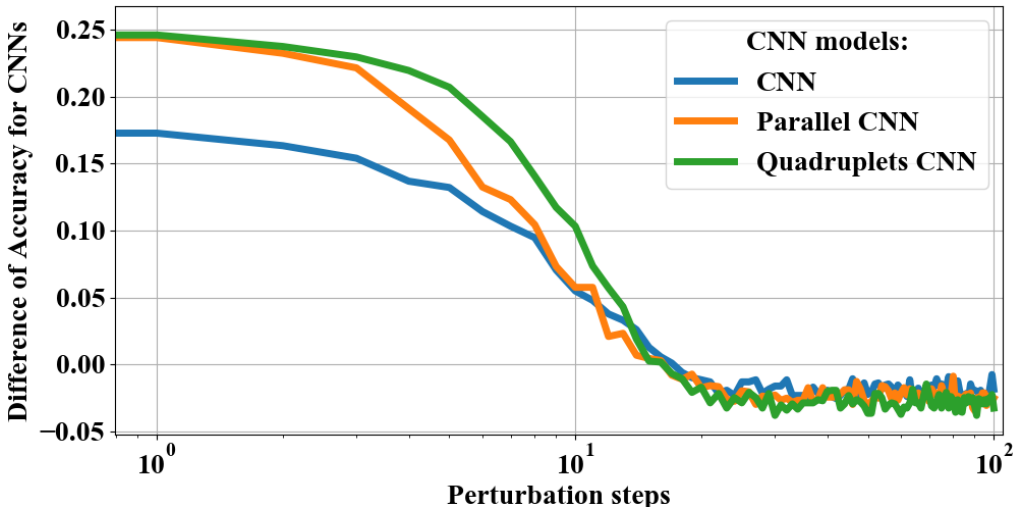



Figure 4.8. Semi-log plot of the perturbation effect on the proposed DCNNs architectures. The decline in accuracy results from progressively removing information from the input data based on LRP-SPF and re-predict, at each step, 100 steps total.

The random data sample in figure 4.11 is used to illustrate the choice of the LRP method by the perturbation approach, as well as the LRP relevance scores obtained for the Parallel DCNN classifications. It shows the processed original GRF signals in the top plot; the middle plot shows the regions replaced with Gaussian noise, in view of the relevance scores shown in the bottom plot. While temporal data patterns yield classifications, the temporal maps of LRP scores highlight data intervals most significant for a given class. The plot of LRP scores consists of sharp peaks, well defined in the temporal domain, thus attributable to time-stamped gait events. Figure 4.11 displays the spatially averaged data signals for the four classes with their respective LRP score maps. The most prominent peaks are attributed to observable gait events, labelled in consistence with the gait cycle in figure 3.1. These are further discussed in section 4.4.

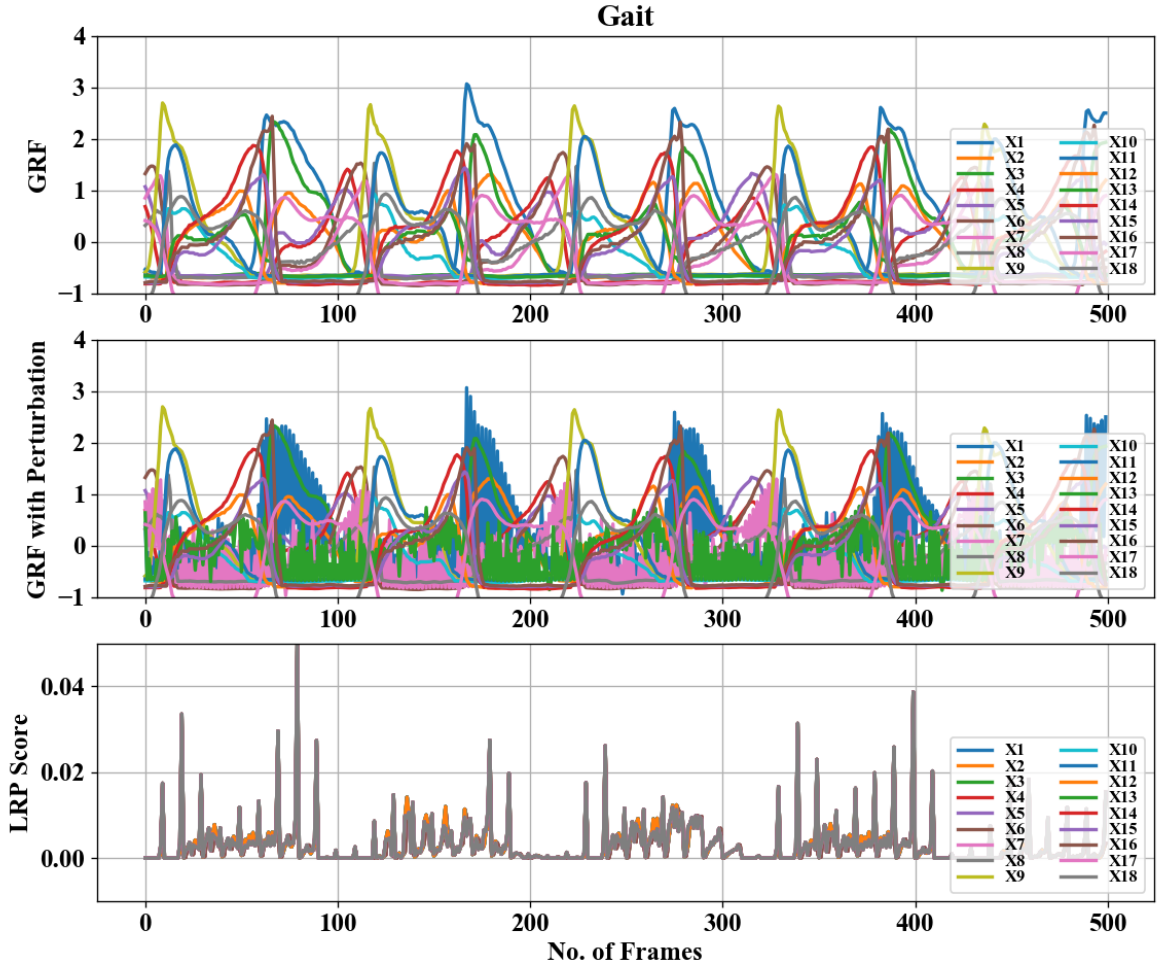
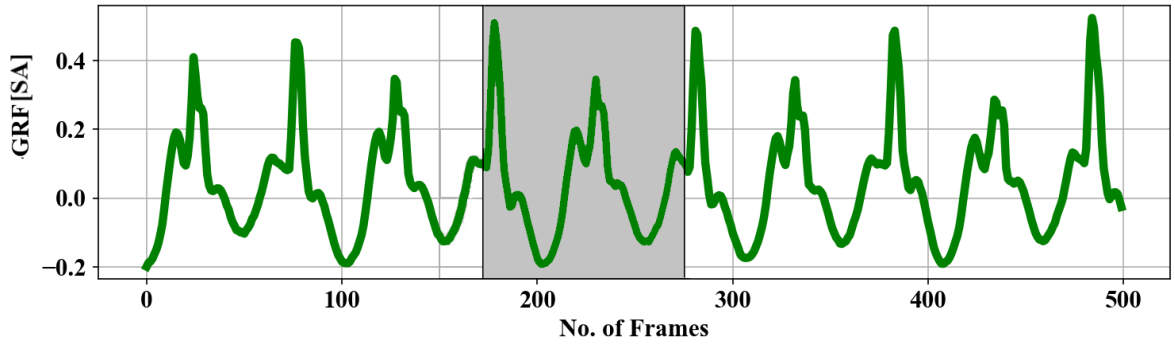
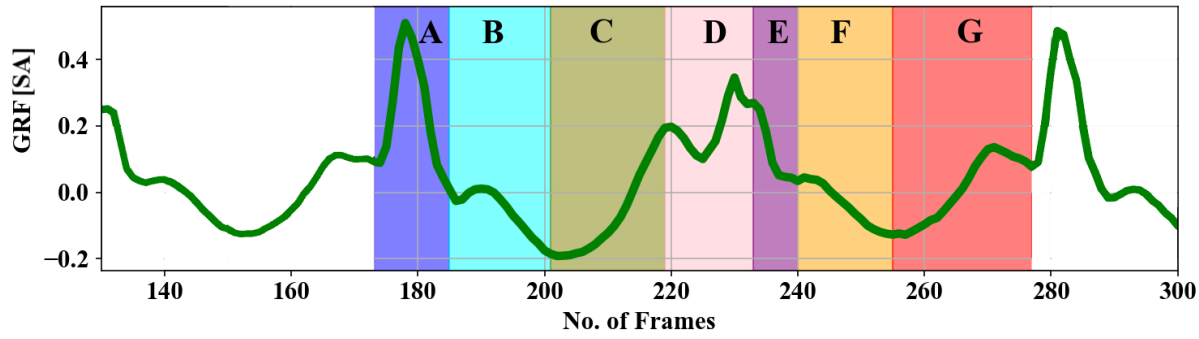



Figure 4.9. Perturbation with Gaussian noise based on LRP relevance scores, used for LRP and DCNN selection. Top plot: healthy gait processed GRF signals; Middle plot: GRF with Perturbation noise; Bottom plot: LRP relevance scores for the selected LRP Sequential

Preset a Flat (LRP-SPF) and Parallel DCNN. $x_i = [x_1 \quad \dots \quad x_{18}]$ represent the 18 signals after data standardization (see figure 4.2 and section 4.3.1). The LRP plot is dominated by $x_{18}$ because all the signals are plotted on top of each other in the temporal domain. LRP relevance score are highly dependent on the temporal changes, whereas spatial variation does not affect the model prediction.




(a)




(b)

Figure 4.10. Gait events processed SA (see equation 4.5) signal top. The highlighted gray area in (a) is explained in (b) based on gait events for one foot from figure 4.2.1 as: A- Heel strike, B- Loading response or flat foot, C- Mid-stance or single support, D- Terminal stance or heel rising, E- Pre-swing or double-limb support, F- Initial swing and Mid-swing or toe-off, G-Terminal swing.

## 4.4 Discussion

To the best of our knowledge, there is no literature on the deterioration of gait based on the foot placement analysis (for PD patients with different severity levels) based on

machine learning models' classification. The spatiotemporal signal in figure 4.1 implies that gait have normal events. Abnormal gait, otherwise difficult to detect visually, can be detected by machine learning, in alignment with the knowledge of the ground truth labels. However, the magnitude of GRF in newton shows a decrease attributable to the severity of PD. The main objective of this work is to find the best deep learning model for PD severity rating and relate the model predictions to the gait cycle events shown in figure 3.1 and figure 4.10.

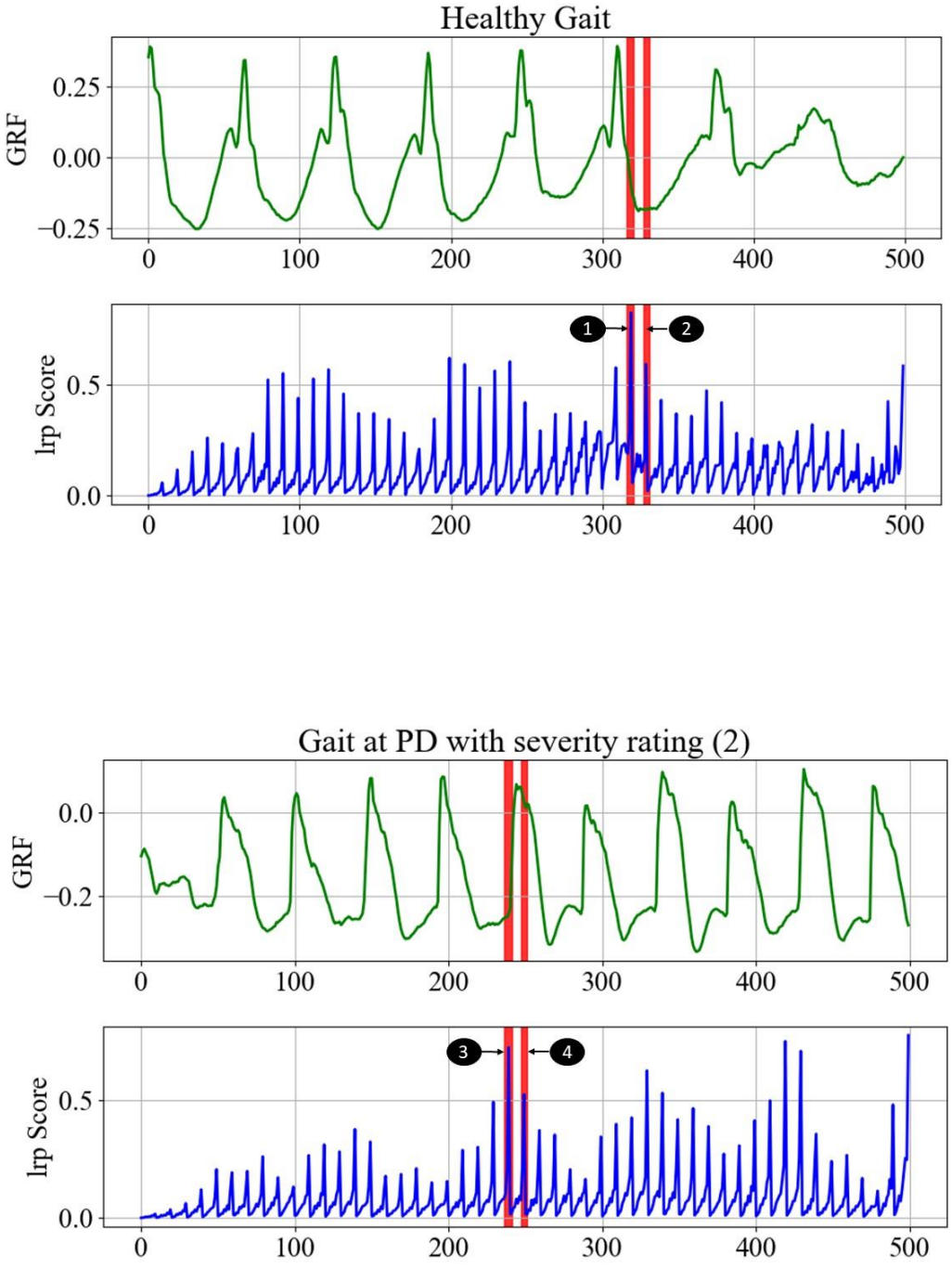
Table 4.4. PD classification results on PhysioNet three datasets.

| Reference | Methods | Accuracy [%] |
|---|---|---|
| E. Abdulhay et al. [35] | SVM | 92.7 |
| Y. N. Jane [160] | Q-BTDNN | 91.5 |
| Ertuğrul et al. [161] | 1D-LBP+MLP | 88.89 |
| Medeiros et al. [137] | PCA | 81.00 |
| Wu et al. [138] | SVM | 84.48 |
| This work | Parallel 2D-DCNN | **95.5±0.28** |

SVM: Support Vector Machine; 1D-LBP+MLP: Shifted 1D-Local Binary Patterns + Multi-Layer Perceptron; PCA: Principal Component Analysis; Q-BTDNN: Q-back propagated time delay ANN.

Research towards machine learning classifications from PD data specifically PhysioNet data is based on the use of manual feature extraction methods with the classical machine learning methods as shown in table 4.4. The best classification results from manual extraction are reported in [35] using SVM classifier (92.7%). Our previous work on PD severity classification [145], reported that the 2DCNN outperformed the 1DCNN, SVM, decision tree algorithm, logistic regression algorithms, multi-layer protection and LSTM. In this article, we explore three architectures for automatic extraction and LRP analysis. The proposed DCNNs identified PD, as well as rated the severity of the deviation from healthy gait, achieving better classification performance with F1-score of 98% for each dataset and for the datasets combined with different random state (see table 4.3). The best classification accuracy is achieved with the parallel 2D-DCNNs, with mean performance and standard errors of 95.5% and 0.28%, respectively. Additionally, the parallel 2D-DCNNs exhibit robustness at perturbation with Gaussian noise as shown in

figure 4.8. This suggests that the model is adequate for detecting gait deterioration from the spatiotemporal GRF signal. As an additional substantial enhancement, our LRP approach allows classification results to be related to visual observations similar to those established in medical practice to diagnose PD.



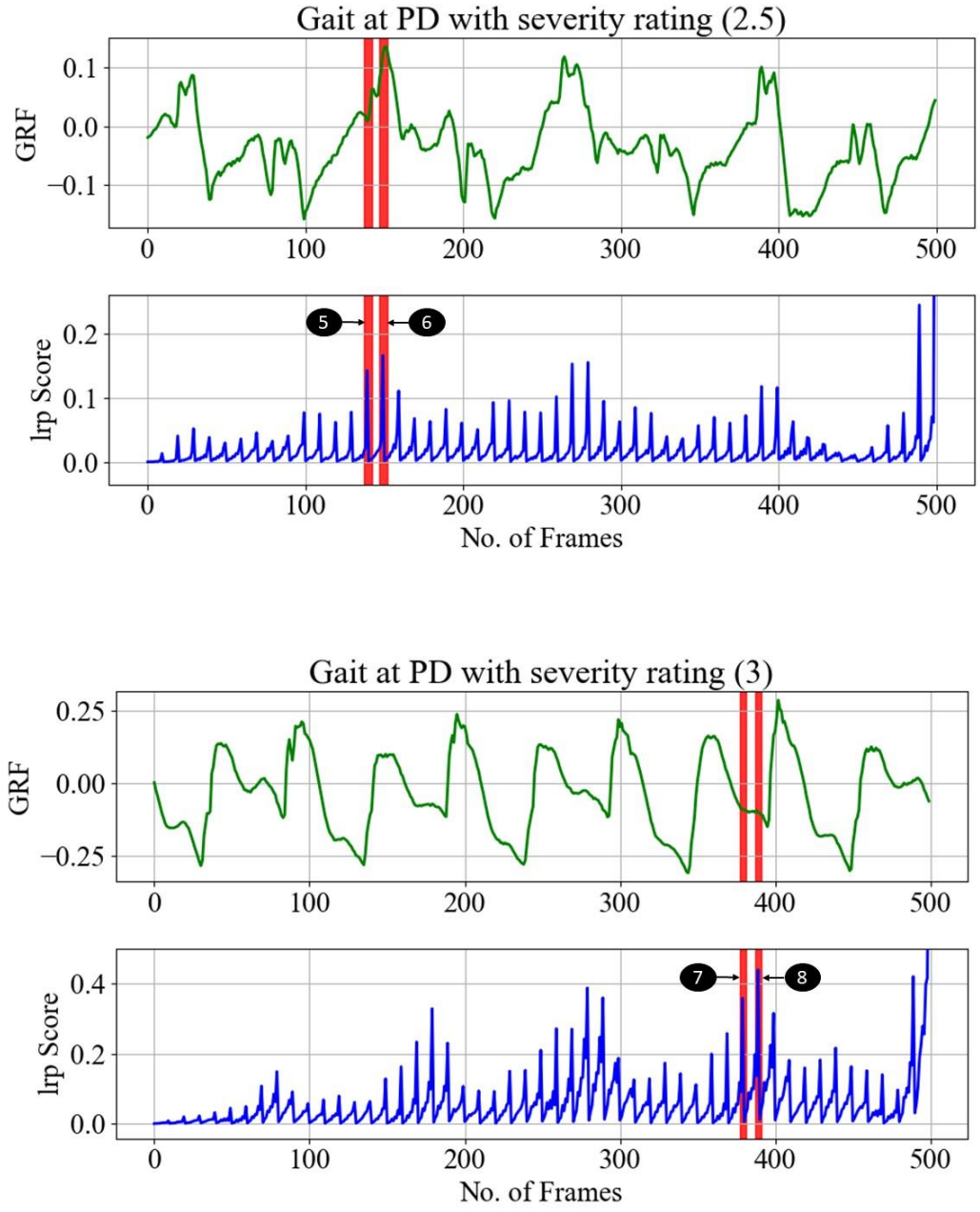



Figure 4.11. LRP method applied on randomly selected samples for the four PD severity ratings. SA of gait spatiotemporal signals: green; SA for LRP relevance scores over the same temporal period: blue. Vertical red bars with number labels display consistency with gait events listed below with capital letters as per figs. 3.1 and 4.10: 1, 3 and 6 - Heal strike and foot flattening (A); 2- Mid-stance and single support (C); 4- Loading response after the double support interval (B), 5 and 8 - Terminal swing and ready for the heel strike (G), 7- (F) Initial swing and Mid-swing or toe-o

The DCNN classifies the raw spatiotemporal signals as healthy or within three severity ratings as shown in the confusion matrix (figure 4.4, 4.5, 4.6). The best LRP method is selected by applying a perturbation technique, which detects the highest sensitivity to removal of information from the input data sequence (figure 4.9). The selected LRP -SPF [141] was found to be superior to well-known methods such as deconvolution [158] and guided backpropagation [159].

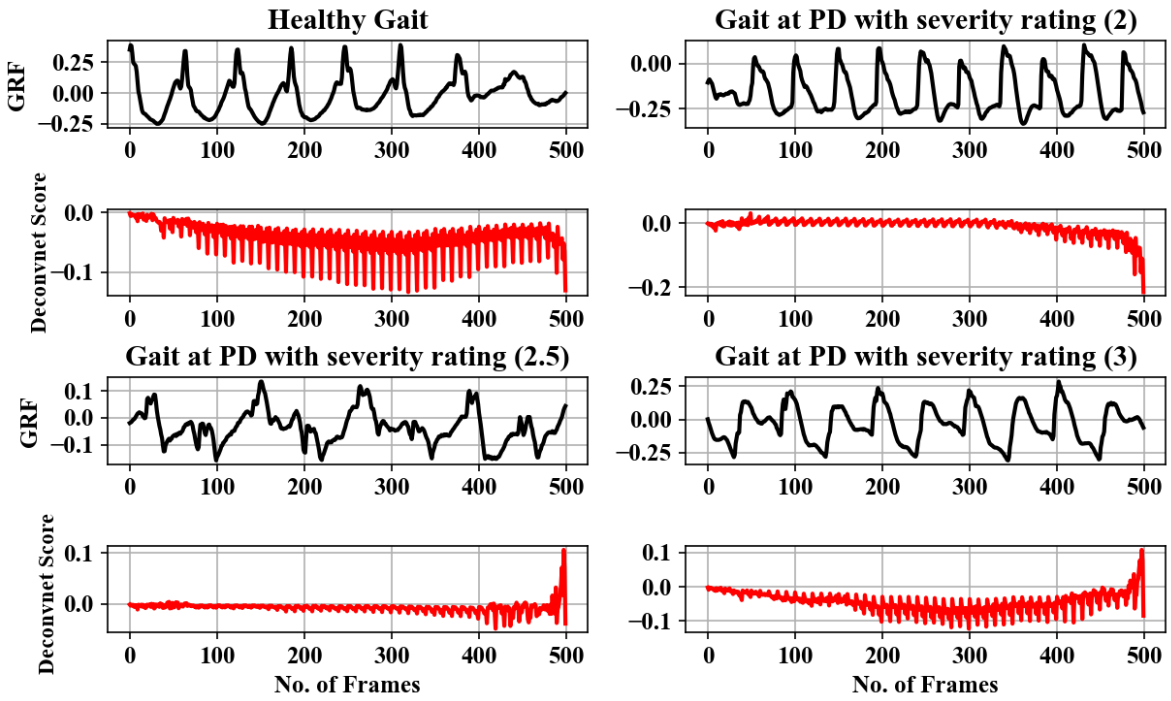



Figure 4.12. A deconvolution decomposition method applied to explain the parallel DCNN model prediction for PD severity ratings. SA of gait spatiotemporal signals is in black plot and model decomposition using deconvolution for each class in red plot. The deconvolution plot spikes every 10 frames.

Among the DCNN architectures (figure 4.8), the Parallel DCNNs model shows the steepest decrease in the perturbation procedure. Therefore, that model is learnt and used for to generate the heat map or relevance randomly selected samples (figure 4.11), without distinction between left and right foot signal is for two feet. The gait cycle events identified as key at each level of PD severity are listed below:

1) PD Severity Level 0 (Healthy Gait): 1- Heel strike and foot flattening (A), 2- Mid-stance and single support (C). This indicates that the healthy person's ability to maintain balance is stronger than the PD patients', with strong balance suggesting that the

forces are applied rhythmically to achieve the lower limbs' synchronized movement with stable posture.

2) PD Severity Level 2: 3- Heel strike (A), 4- Loading response after the double support interval (B). The heatmap shows that the subjects affected with PD level 2 have a weaker heal strike followed by a weaker balance in double support, where this feature is marked by the model by 96% f1-score.
3) PD Severity Level 2.5: 5- Terminal swing (G), 6- Heel strike (A). This shows that the subject has weaker foot landing or flat foot landing after the balance is compromised by the single support.
4) PD Severity Level 3: 7- Initial swing and mid-swing or toe-off (F), 8- Terminal swing and ready for the heel strike (G).

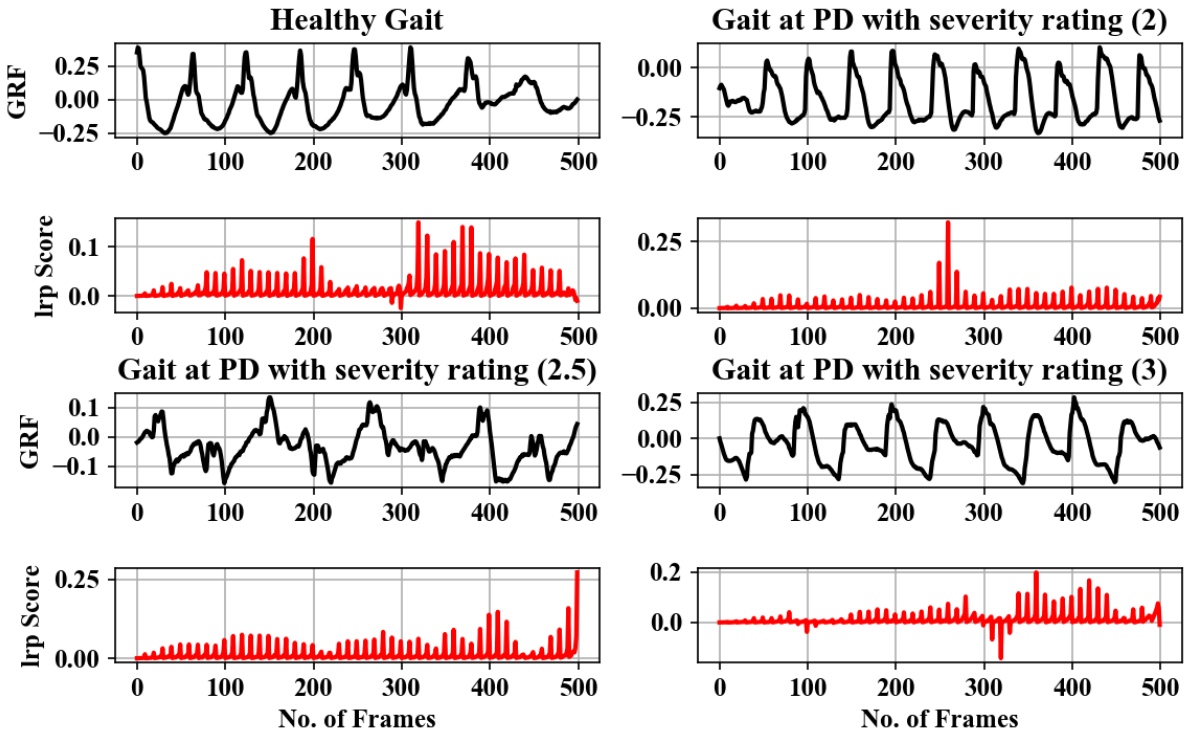



Figure 4.13. A Deep Taylor decomposition method applied to explain the parallel DCNN model prediction for PD severity ratings. Similar to figure 4.11 and 4.12 the Deep Taylor decomposition plot spikes every 10 frames.

Here the balance is compromised by weak GRF resulting from unstable body posture and implies high risk of falling. This conclusion is based on linking the stages of PD in [162]

(description of how the stage of PD is affecting the body poster during gait using visual observation) to the events that are highlighted by the model for a certain PD severity.

The above markers for classification align with the observations in literature that PD-induced gait GRF deterioration affects the body balance and posture. The latter are with the closest relevance to gait events identified by the heat maps in figure 4.11 as the highest LRP scores, while the other gait events are less significant to the classifications. It is worth mentioning that these markers are identical by 95.5% in 1281 samples, such that the removal of these regions in the 95.5% of samples resulted in strong decay in the model prediction. The interpretation given above is in very good agreement with the description of the Hoehn and Yahr Scale staging criteria, as follows: “Stage 0 - No signs of disease, Stage 2 - Symptoms on both sides but no impairment of balance, Stage 2.5 - Mild symptoms on both sides, with recovery when the ‘pull’ test is given (the doctor stands behind the person and asks them to maintain their balance when physically pulled backwards), Stage 3 - Balance impairment, mild to moderate disease, physically independent” [162]. However, the staging criteria do not refer to the gait events adversely influencing the body’s postural balance, due to the advancement of disease.

The analysis of LRP score in figure 4.9 and 4.11 reveals a consistent spikes in the LRP plot for every 10 frames, thus further investigation has been carried out by plotting the model decomposition using deconvolution, Deep Taylor and LRP-SPF for a single stream DCNN model. As shown in figures 4.12, 4.13, and 4.14, the spikes are also consistent in all the plots. Further, the sensors used to record PD and healthy gait in [146],[147],[148] are pressure-sensitive sensors to measure the forces underneath the foot as a function of time at a rate of 100 Hz.

These spikes can be an artefacts generated from the data processing, either in the forward pass (classification) or the backward pass (LRP decomposition). At this stage, we are not able to find a satisfactory explanation of these spikes; however, the data is considered reliable, based on its noise resilience as demonstrated in our perturbation analysis.

## 4.5 Conclusion

In this article, deep learning models are designed, tested and proven their ability to classify gait deterioration in PD patients. The models outperform previous manual

feature extraction methods and exhibit robustness to perturbation noise, overcoming the variability of human gait from person to person. The LRP analysis confirms that body balance in [162] is a critical aspect for diagnosing PD, where higher levels of the disease affect patient's ability to walk without being at risk of falling. Furthermore, the identification of the most relevant gait cycle events can aid clinical practitioners in diagnostics based on visual observation criteria, or based on handcrafted features from other GRF sensors. As for future directions, the methodology proposed in this article can contribute to developing a strategy towards personalized longitudinal monitoring the progression of PD severity.

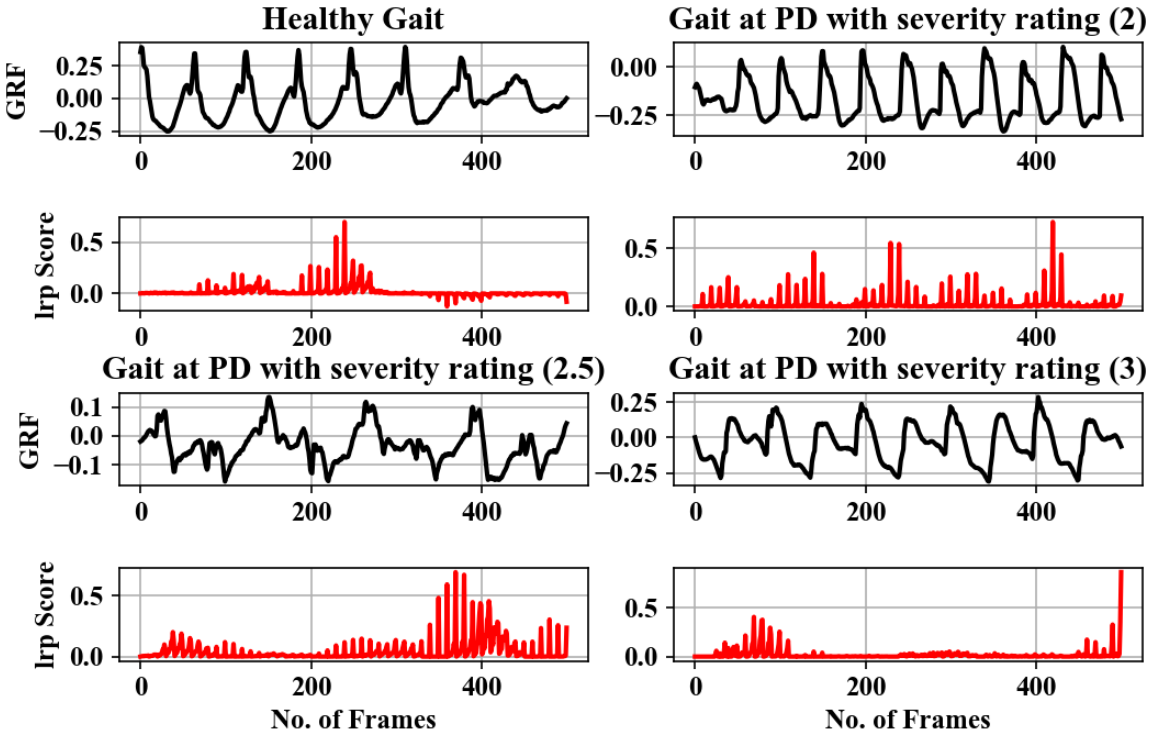



Figure 4.14. A LRP sequential preset a flat (LRP-SPF) applied to explain the single DCNN model prediction for PD severity ratings. Similar to figure 4.11, 4.12 and 4.13 the single DCNN model decomposition plot spikes every 10 frames.

# Chapter.5 Gait Under Cognitive Load Part (1)

# Sensor Fusion for Analysis of Gait under Cognitive Load: Deep Learning Approach

**Abdullah S. Alharthi, Syed U. Yunas, and Krikor B. Ozanyan**

## Abstract

Human mobility requires substantial cognitive resources, thus elevated complexity in the navigated environment instigates gait deterioration due to naturally limited cognitive load capacity. This work uses deep learning methods for 116 sensors fusion to study the effects of cognitive load on human gait of healthy subjects. We demonstrate classifications, achieving 86% precision with Convolutional Neural Networks (CNN), of normal gait as well as 15 subjects' gait under two types of cognitive demanding tasks. Floor sensors capturing multiples of up to 4 uninterrupted steps were utilized to harvest the raw gait spatiotemporal signals, based on the ground reaction force (GRF). A Layer-Wise Relevance Propagation (LRP) technique is proposed to interpret the CNN prediction in terms of relevance to standard events in the gait cycle. LRP projects the model predictions back to the input gait spatiotemporal signal, to generate a "heat map" over the original training set, or an unknown sample classified by the model. This allows valuable insight into which parts of the gait spatiotemporal signal have the heaviest influence on the gait classification and consequently, which gate events are mostly affected by cognitive load.

Article published at 2020 IEEE Sensors Applications Symposium (SAS). Date of Conference: 9-11 March 2020, 978-1-7281-4842-7/20/$31.00 ©2020 IEEE. Abdullah S. Alharthi, Syed U. Yunas, and Krikor B. Ozanyan are with the department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, United Kingdom, (e-mail: abdullah.alharthi@manchester.ac.uk; syed.yunas@manchester.ac.uk; k.ozanyan@manchester.ac.uk). DOI: 10.1109/SAS48726.2020.9220046

## 5.1 Introduction

Gait patterns have been intensively studied in recent years to understand the relationship between cognitive load and gait disturbance, with understandable interest on changes with age. For many years, gait of healthy individuals was considered to be separate from the cognitive system, but now it is accepted that such systems are inter-related at the level of the cerebrum. Furthermore, it has been asserted that gait is not an automatic motor sequence independent from cognition but can be affected by cognitive demanding tasks [163].

Yogev et al. [164], based on the impact of executive function and attention in gait, suggested that the latter is no longer regarded as an automated motor activity that receives minimal cognitive input. Indeed, the limb movements that produce gait with the ability to navigate within often complex surrounding to successfully reach a desired destination, influences gait. The multifaceted neuropsychological influence on humans to appropriately control the mobility and related behaviors, is intensively investigated in recent years for healthcare applications, leading to the hypothesis that the changes in individual gait under dual task (performing a demanding cognitive task while walking), are determined by the capacity decline of cognitive executive function in the normal ageing process, when the overall efficiency of the brain is reduced [163], [164], [165].

Despite the step in research on the influence of cognitive load on gait in the past few decades, the systems providing sensor data have been limited to the following: instruments attached to the body non-invasively [144], [166], force plate where the user must watch where to step [144], or monitoring of gait while walking on a treadmill [167], [168]. These methods may influence the fidelity of recording the natural gait pathology under cognitive demanding tasks, as imposed conditions are quite different from the natural environment in which an individual would walk.

The interaction of the human body with the walking surface is the point of contact with the environment, which cannot be avoided or modified at will. This interaction is typically described in terms of the ground reaction force (GRF). Thus, large area, unobtrusive floor sensor systems are called for, to mimic walking under cognitive load in a real-world. This allows investigating the spatiotemporal dynamics of the forces placed on the ground

during the gait cycle, by recording the signal of the cyclic behavior in varying spatial regions and time periods. Multi-sensors data analysis and fusion methodology would be required for adequate data analysis aiming to compare natural gait with that influenced by cognitive load. An obvious motivation is to study the consistency of gait patterns which can be associated with fall risk [169], [170] and mental decline, e.g. Parkinson's disease [171], [172].

Cognitive load influence studies have targeted to quantify gait deterioration using established criteria by visual monitoring, or using statistical analysis methods for classification. It is difficult to claim that any of these methods, or their combination, would be able to access the maximum variability of gait deterioration and thus provide the best achievable classifications.

Therefore, the goal of this work is twofold: firstly, to categorize gait parameters in healthy adults while performing cognitive demanding tasks, using automatic extraction of optimal gait features by deep CNN; secondly, to interpret the model performance on unseen spatiotemporal signals by applying the technique of Layer-Wise Relevance Propagation (LRP) [21] to attempt linking key known events in the gait cycle to cognitive deterioration. As a main data acquisition methodology, the subject's unique gait signatures based on GRF signals were recorded using a set of multiple floor sensors based on plastic optical fiber (POF) technology, in a system specially designed for optimal spatiotemporal sampling [106],[131]. The main sensor fusion and data processing methodology is deep CNNs with LRP applied on the classifications to allow interpretation of the behavioral changes due to a demanding task while walking.

## 5.2 Methodology

### 5.2.1 iMAGiMAT System

The gait cycle recordings reflect the recurrent contact of the feet with the surface. The experimental work utilizes a photonic guided-path tomography sensor head, as part of an original iMAGiMAT footstep imaging system [106], for its ability to record unobtrusively temporal samples from each sensor under natural walking environment (a

general retail floor carpet). The sensors constitute a carefully designed set to allow collaborative sensor fusion and deliver spatiotemporal sampling adequate for gait events.

The 1m x 2m area system (see figure 5.1) contains 116 POF sensors, arranged in three plies sandwiched between the carpet top pile and the carpet underlay - a lengthwise ply with 22 POF at 0° and two 47 POF plies arranged diagonally at 60° and 120°. The system electronics is concealed in a closed hard shell located at the rectangular boundaries at carpet surface level. The operational principle of the system is based on capturing the deformation caused by the GRF variations as the POF sensor's transmitted light intensity decreases because of surface bending. This captures the specifics of foot contact and generates robust data without constraints of speed or positioning anywhere on the active surface.

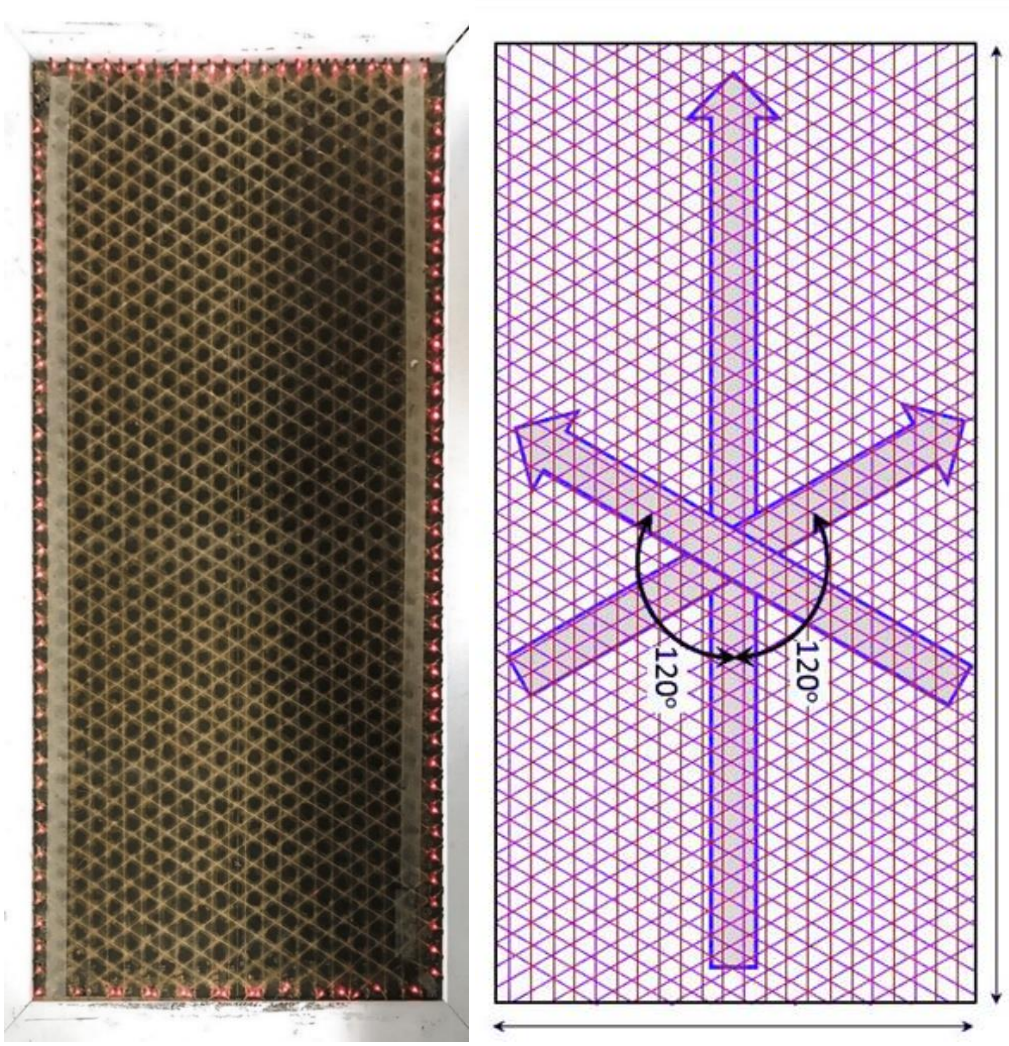



Figure 5.1. iMAGiMAT design on the right and actual system on the left.

### 5.2.2 Data Acquisition

Ethical approval was obtained through the Manchester University Research Ethics Committee (MUREC); each participant's written consent was obtained prior to experiments and research was performed in accordance with the ethics board general guidelines. Healthy participants were invited to walk along the 2 m length direction of the iMAGiMAT sensor head normally, or while performing cognitive demanding tasks. Since the walking routine involved steps before and after contacting the active sensor area, the recorded gait spatiotemporal signals were able to capture around 4 uninterrupted footsteps at each pass.

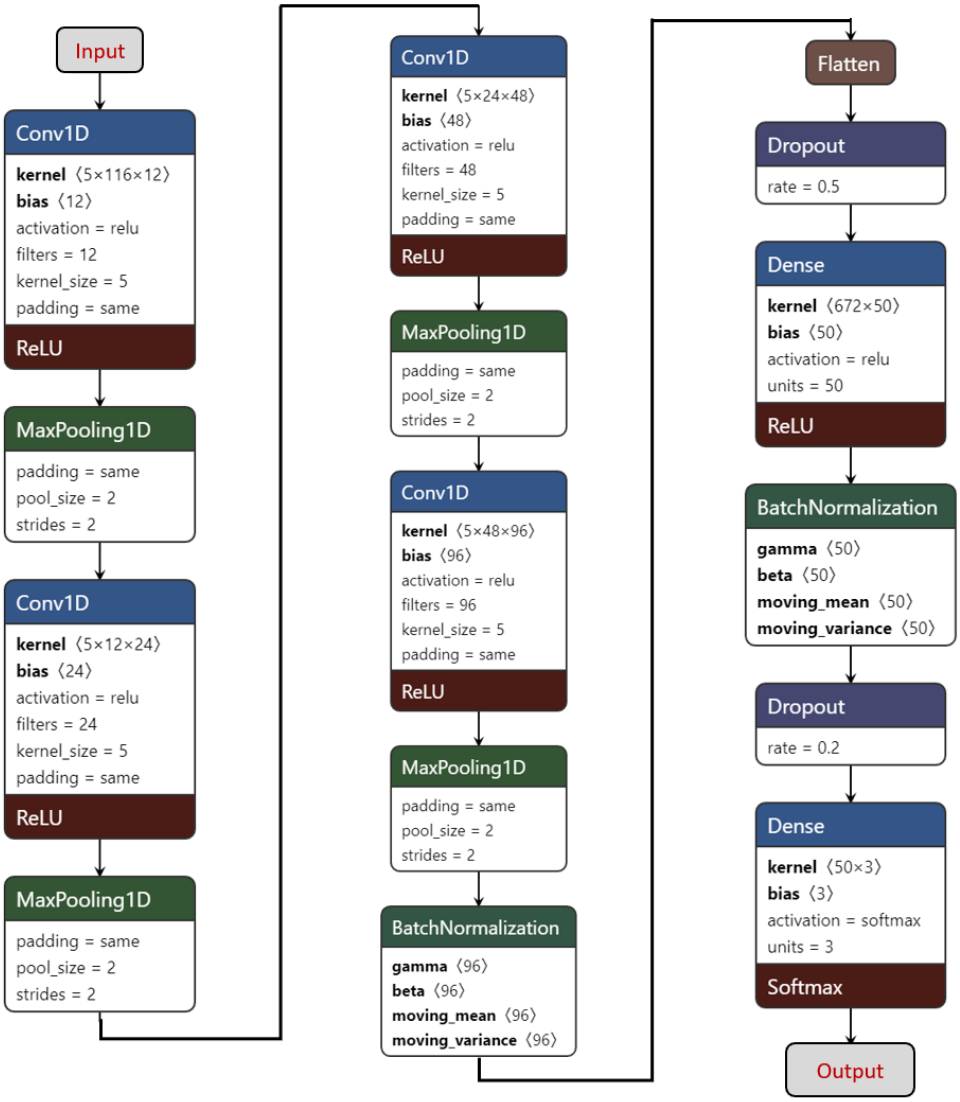



Figure 5.2. CNN network architecture for gait classification, and LRP.

15 healthy subjects aged 20 to 38 years, 13 male and 2 females, participated in this experiment. The number of subjects in this experiment is impacted by the COVID-19 pandemic, in which we were unable to record more data. Each subject was asked to walk several gait cycles before and after stepping on the iMAGiMAT system for each experiment, so the captured gait data is unaffected by start and stop. With a capture rate of 20 frames/s (each frame comprising the readings of all 116 sensors), experiments yielded 5 s long time sequences each containing 100 frames.

Three manners of walking were defined as normal gait, as well as two different dual tasks, and experiments were recorded for each subject, with 10 gait trials for each manner of walking in a single assessment session, total number of samples is 450. Each manner of walking was identified as class, as follows:

- Class 0, Normal Gait: walking at normal self-selected speed.
- Class 1, Gait with serial 7 subtractions: normal walking speed attempted, while simultaneously performing serial 7 subtractions (count backward from a given random number by sevens).
- Class 2, Gait while texting: normal walking speed attempted, while simultaneously typing text on a mobile device keyboard.

### 5.2.3 Convolutional Neural Network (CNN)

State-of-the-art CNNs have established themselves in a variety of classification tasks, providing new insights into complex spatiotemporal signals. The models learn a high level of abstraction and pattern by applying convolution operations to extract features and perform classification. Commonly, the model consists of convolution layers, pooling layers and normalization layers, with a set of shared filters and weights.

The architecture of the CNN model engineered for this study is shown in figure 5.2. The model is proposed based on extensive research on gait in our previous work [173],[131],[145],[150]. The model consists of 12 stacked layers to categorize all 15 subjects' gait, normal and dual task. The model is trained and validated using a batch size of 100 samples for each iteration, 200 epochs being optimal for training. The training,

validation and testing sizes are set to be 70%, 10% and 20% respectively. To improve the model performance a regularization method is utilized as follows:

i. A batch normalization followed by a dropout of size 0.5, after the last MaxPooling layer was flattened.
ii. A dropout of size 0.2 before the output layer.

Decisively, *softmax* classifier is placed at the final layer to classify gait spatiotemporal signals. In figure 5.2, each convolution layer has (kernel × feature maps × filter) labels, and they are known with the number of layers as hyperparameters. These hyperparameters are optimized based on extensive experimenting. For this study, experimentation and algorithm implementation are done using Keras libraries [174]. This toolbox provides variety of commands tested until the best model is reached. This is based on testing the model on unseen data.

### 5.2.4 LRP

LRP [20], [21] is an emerging machine learning method for explaining the predictions of nonlinear deep neural networks, and it has shown notable success in image classification [21], [143] as well as subject identification based on their gait [144].

The aim is to quantify the contribution of a single component of an input $x_i$ (in our case, an iMAGiMAT sensor signal at a specific time frame) to the prediction of $f_c(x)$ (as gait class $c$) made by the CNN classifier $f$. Gait class prediction is redistributed to each intermediate node via backpropagation until the input layer. The LRP outputs a heat map over the original signal to highlight the signal sections of highest contributions to the model prediction. To understand this methodology, we first note that a CNN network consists of multiple learning parameters as follows [144]:

$$f_c(x)=\omega^T x_i + b_c = \sum_i \omega_{ic}\, x_i + b_c \tag{5.1}$$

Here, $\omega_{ic}$ is the weight of a neuron $i$ in a specific layer $l$ for class $c$ and $b_c$ is the bias term for class $c$; these parameters are learned in the CNN during supervisory training by fusing the POF sensor signals. The output $f_c(x)$ is evaluated in a forward pass and the

parameters ($\omega_{ic}, b_c$) are updated by back-propagating using model error or the loss. For the latter, we base the computations on categorical cross entropy.

The LRP decomposes the CNN output for a given prediction function of gait class $c$ as $f_c$ for an input $x$ and generates a "relevance score" $R_i$ for all neurons $j$ that receive relevance from neuron $k$ for the prediction of $f_c(x)$, such that $LRP = \sum_i R_{i\leftarrow j} = \sum_j R_{j\leftarrow k} = f_c(x)$. The LRP starts at the CNN output layer after removing the *Softmax* layer. In this process, a gait class $c$ (e.g. normal gait, subtraction 7 gait, or walking while texting) is selected as input to LRP and the other classes are eliminated.

Assuming we have a single output neuron $j$ in one of the model layers, this neuron receives a relevance score $R_j$ from an upper layer neuron $k$ or the output of the model (class $c$). The scores are redistributed between the connected neurons throughout the network layers, based on the contribution of the inputs $i$ in signal $x$ using the activation function (computed in the forward pass and updated by back-propagating during training) of neuron $j$.

The neuron $k$ will hold a certain relevance score based on the activation function of that neuron, and it passes its influence to consecutive neurons ($j$) in reverse direction. Finally, the method outputs relevance scores for each sensor signal at a specific time frame. These scores represent a heat map, where the strong values at specific time frames highlight the areas that contributed most to the model classifications. The entire set of model relevance scores $R_j$ can be represented as follows [20],[22]:

$$R_j^{(l)} = \sum_k \frac{a_j^{(l)}\omega_{jk}^{(l+1,l)}}{\sum_j a_j^{(l)}\omega_{jk}^{(l+1,l)}} R_k^{(l+1)} \qquad (5.2)$$

Here $j$ indexes neurons in layer $l$, and $k$ is computed over all neurons joined to neuron $j$. Equation 5.2 satisfies the relevance conservation property $\sum_j R_{j\leftarrow k} = \sum_k R_k^l = f(x)$.

## 5.2.5 LRP Heatmapping Validation

The LRP produces heatmaps over the test samples for a certain class of gait prediction. Understandably assessing the generalization of these heatmaps over a gait class is essential for their effectiveness, where gait events highlighted by neural network need to be consistent on all the test samples for a gait class. Impractical way to examine the

consistency of heatmaps of a gait class is by visually examining all the test samples of that gait class. As reported in [24], a reliable and practical technique to assess LRP heatmaps is the "pixel-flipping procedure". Similar to the perturbation introduced in section 4.2.3, in this chapter we apply the pixel-flipping procedure by testing whether the removal of features highlighted by the explanation (as most relevant (MoRF) approach) leads to a strong decay of the network prediction abilities [24]. This strong decay only occurs if the heatmaps are consistent. For comparison, a random removal of information from each sample is applied in each step with MoRF regions based on LRP heatmaps, where a region of size 7×7 is considered.

## 5.3 Results

The iMAGiMAT system captures a sequence of periodic events characterized as repetitive cycles for each foot. Further explanation of gait cycle events can be found in [131] or see chapter 3. Figure 5.3 shows the spatial average *SA* of the spatiotemporal gait signal based on the GRF:

$$SA[n] = \frac{1}{116}\sum_{n=1}^{116}(x_i\,[n]) \tag{5.3}$$

to capture gait events due to the stepping behavior. Here $\boldsymbol{x_i}$ are the readings from individual sensors and $n$ enumerates the frames in each sample.

### 5.3.1 Gait spatiotemporal Classifications

The proposed model is trained and validated on 360 spatiotemporal signals samples of size 100x116, which is 80% of the 15 subject's gait trials. The remaining 90 samples of the same size 100x116 are used for testing the model prediction and computing the LRP relevance map. Three types of gait including cognitive demanding task patterns are learned. This is based on fusing 116 POF sensors in the deep layers of the CNN model, to extract gait patterns automatically. The model learns gait events for different subjects while they perform cognitive demanding tasks. The classifications result showing in figure 5.4, indicates that the classification accuracy is 85% for normal gait, 90% for walking with 7s subtraction and 79% for walking while texting. The highest confusion was between the two dual tasks. These accuracies were obtained after extensive

Optimization of the CNN model shown in figure 5.2, with the goal to achieve best prerequisite for performing LRP analysis.

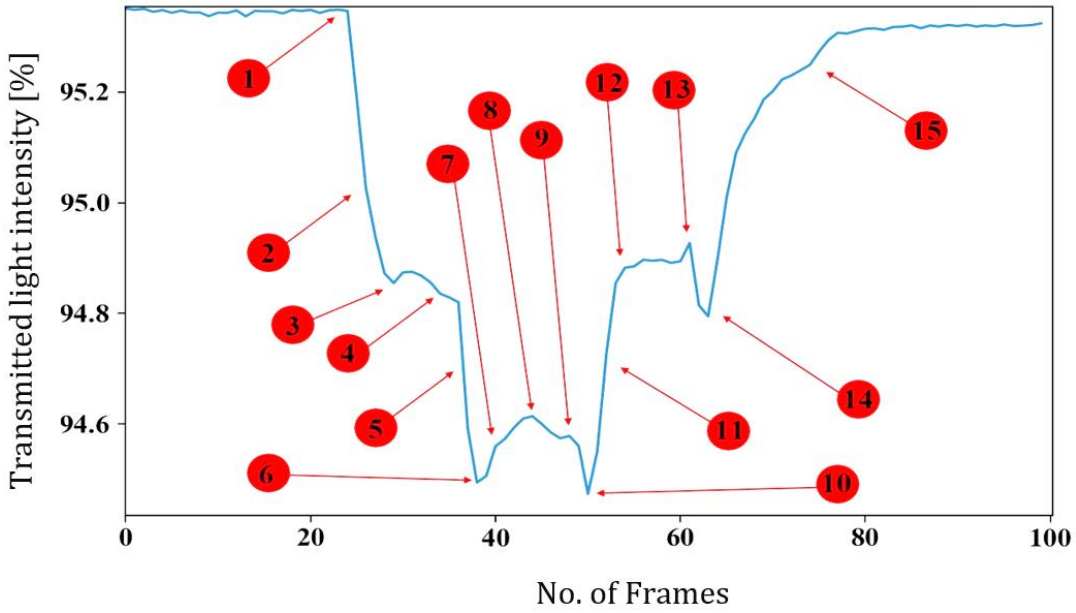



Figure 5.3. Gait spatial average of gait spatiotemporal signals (see equation 5.3), y axis: transmitted light intensity [%]; x axis: frames. Gait events recorded by the sensors in a typical gait cycle[131]: 1- Heel strike, 2- Foot-flattening, 3- Single support, 4- Opposite Heel strike, 5- Opposite Foot-flattening, 6- Double support, 7-Toe-off, 8- Foot swing, 9- Heel strike, 10- Double support, 11- Toe-off, 12- Foot swing, 13- Opposite Heel strike, 14- Single support, 15-Toe-off.

### 5.3.2 Gait Spatiotemporal Classification Decomposition

The CNN model in figure 5.2 parameters were frozen after completing the training for further use to analysis the classification results. Deep Taylor decomposition [21] based on the LRP (see 5.2.4) was utilized for this work using the iNNvestigate library [153]. The LRP method decomposes the CNN prediction, to time-resolved input relevance scores $R_i$, for each spatiotemporal input $x_i$. This highlights which gait cycle events are relevant towards informing the model's final classification outcome.

Using LRP heatmapping validation technique, the heatmaps are assessed by removing region of size 7×7 (based on both LRP Deep Taylor decomposition MoRF region and random region) in each step and replacing them with Gaussian noise and re-predict the

model on the test data for 1000 steps while the model performance is monitored. As shown in figure 5.5 (a), the model strongly decays after few steps compared to the removal of random regions. Figure 5.5 (b) shows that the model reaches the lowest prediction performance after approximately 200 steps.

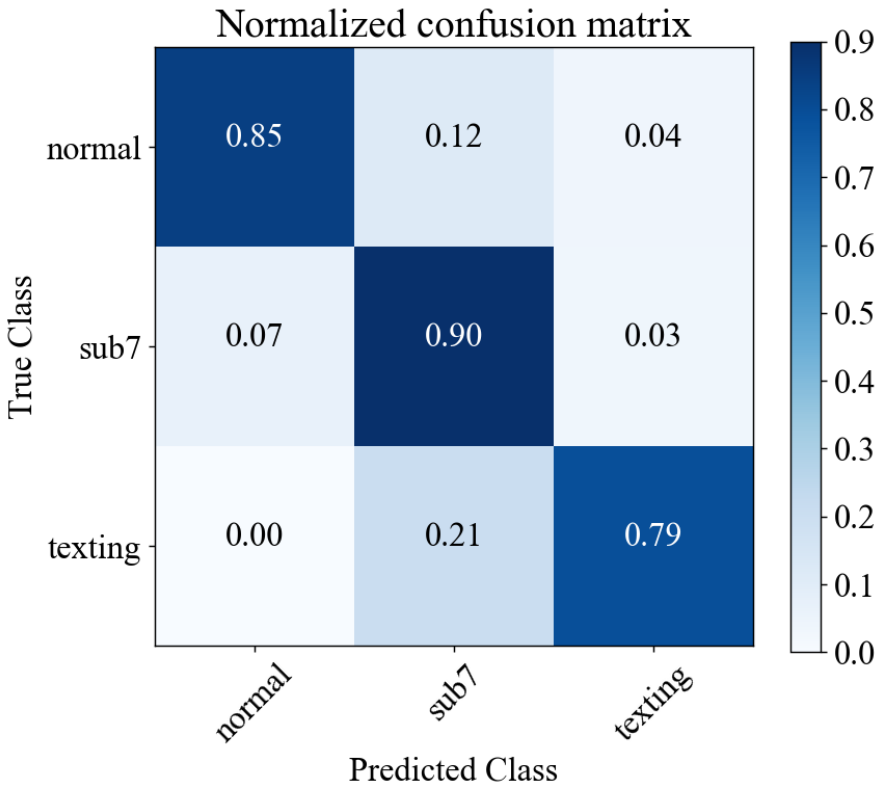
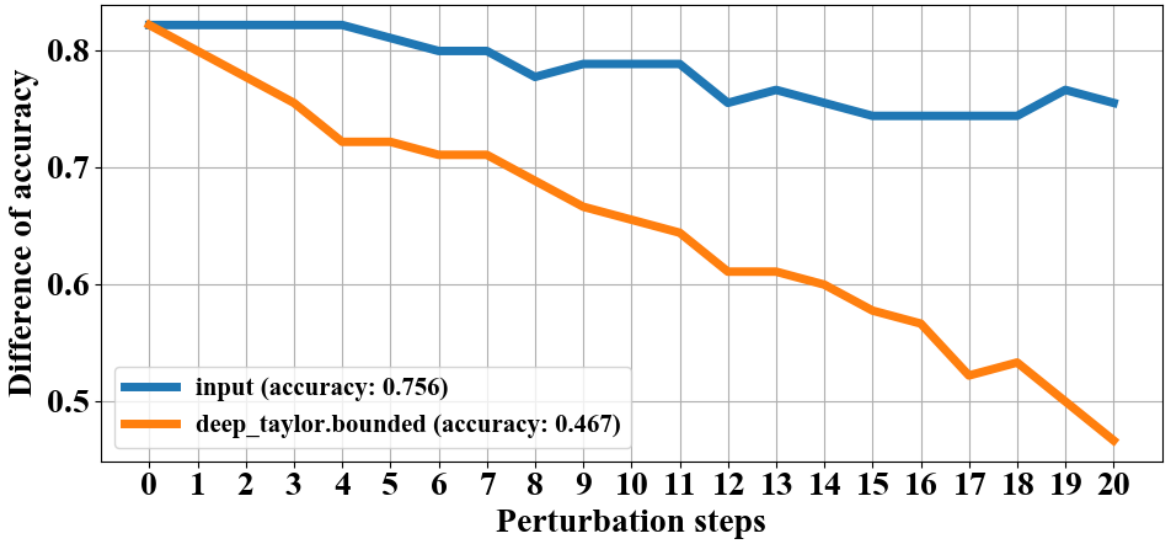
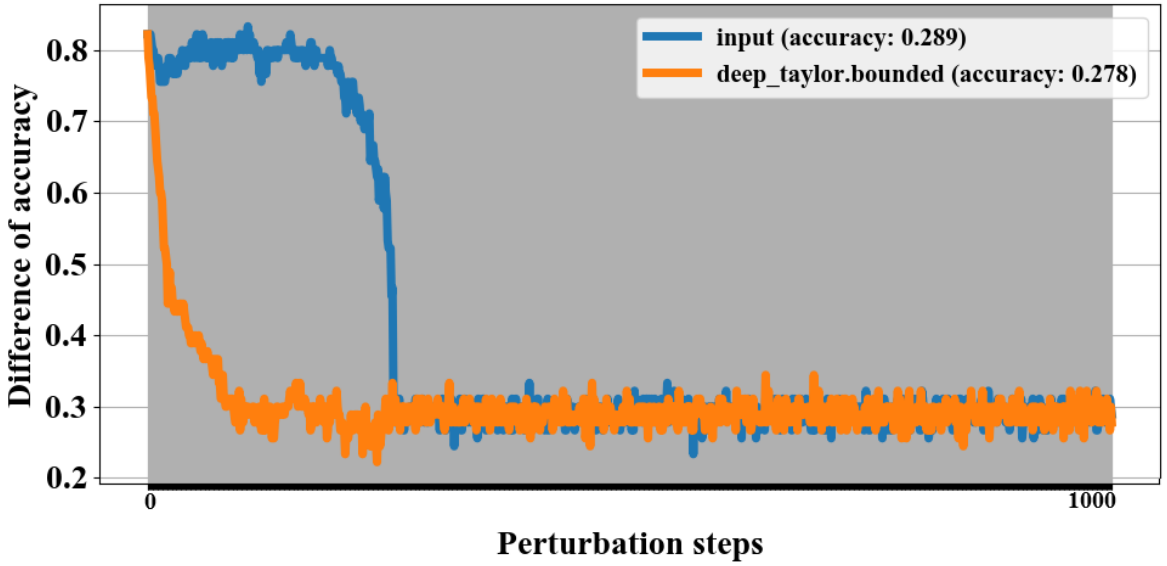



Figure 5.4. CNN model confusion matrix classification of 15 subjects.

Figure 5.6 displays an example overlay of a 1-D "heat map" for dual task samples classified with 100% true positives. For more in-depth analysis, we compare in figure 5.7 the *SA* of the raw IMAGiMAT POF signals from (equation 5.3) with a relevance temporal quasi-signal generated from LRP calculations (1-D heat maps), as follows (further explanation in 5.4): the LRP generates relevance scores (see figure 5.6) over the input space of the original spatiotemporal signal; these scores are then used to outline the input variable most contributing to the prediction of a certain class. Figure 5.7 shows the temporal sequences with identified gait events for three subjects (raw sensor signal and computed LRP quasi-signal row pairs for each subject).




(a)




(b)

Figure 5. 5. LRP heatmaps validation by perturbation technique. Information with the highest relevance scores is progressively removed and the test samples are re-predicted. Steeper initial decrease indicates better identification of gait events with most weight in the classifications. a) shows the model predictions in 20 steps based on removing relevance scores using Deep Taylor decomposition and random removal of information. b) shows the model performance after 1000 steps of information removal.

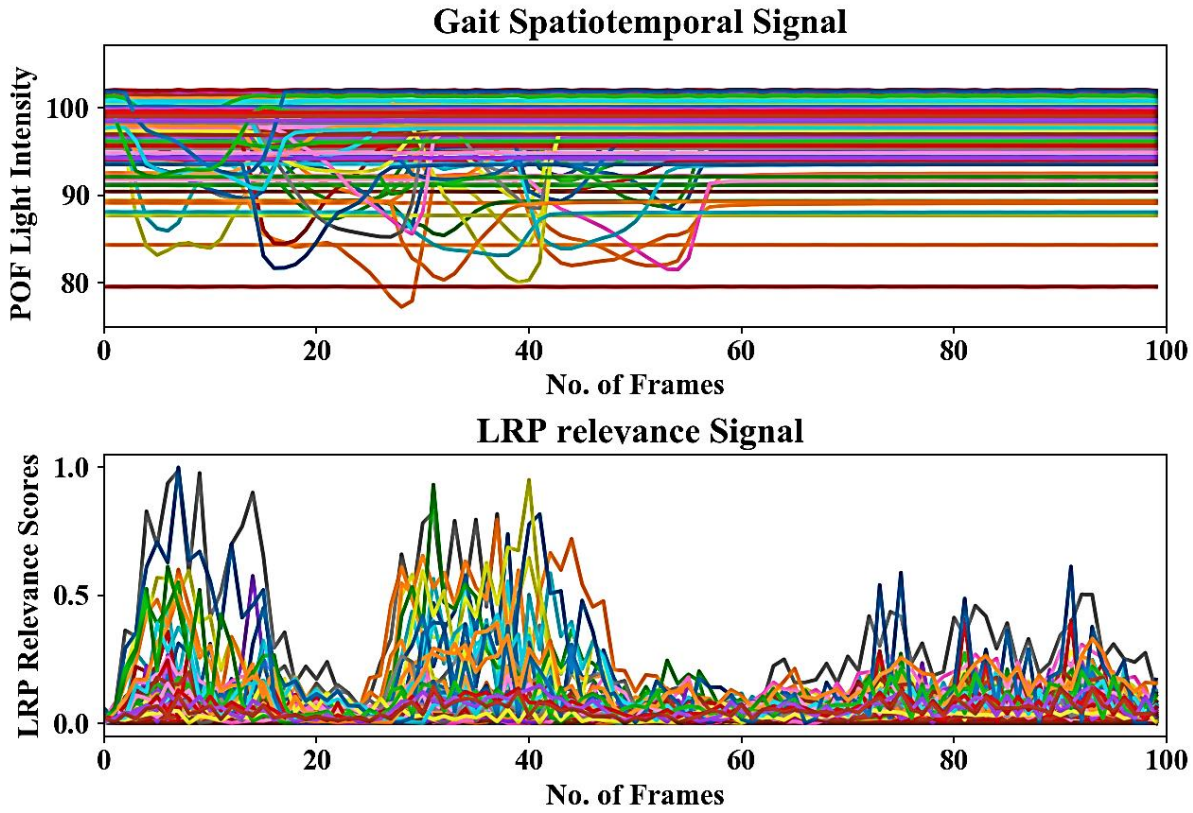



Figure 5.6. Raw IMAGiMAT POF signals top, corresponding LRP scores signals bottom.

## 5.4 Discussion and Conclusion

In this chapter, GRF-derived spatiotemporal signals from healthy control subjects are recorded with the floor sensor iMAGiMAT system, under the conditions of normal walking, as well as walking while performing cognitive demanding dual tasks. Deep learning methods are applied to extract gait features automatically end-to-end from raw data. Gait under cognitive load is classified with 86% weighted precision. The LRP relevance scores point out which parts of the gait spatiotemporal signal were most relevant for classification, as shown in figure 5.7 for three subjects. To understand this approach, we note that on multiple repetitive occasions each subject will initiate a gait cycle (explained in chapter 3) by heel strike, strictly followed by other gait events described in figure 5.3.

Figure 5.4 details a subject's recorded gait temporal signal, where it starts by heel strike and ends by toe off when the user steps out of the iMAGiMAT floor sensor. Consistent with this, each gait spatiotemporal sample (classified by the CNN network) is assigned a corresponding spatiotemporal LRP computed quasi-signal as shown in figure 5.7. The

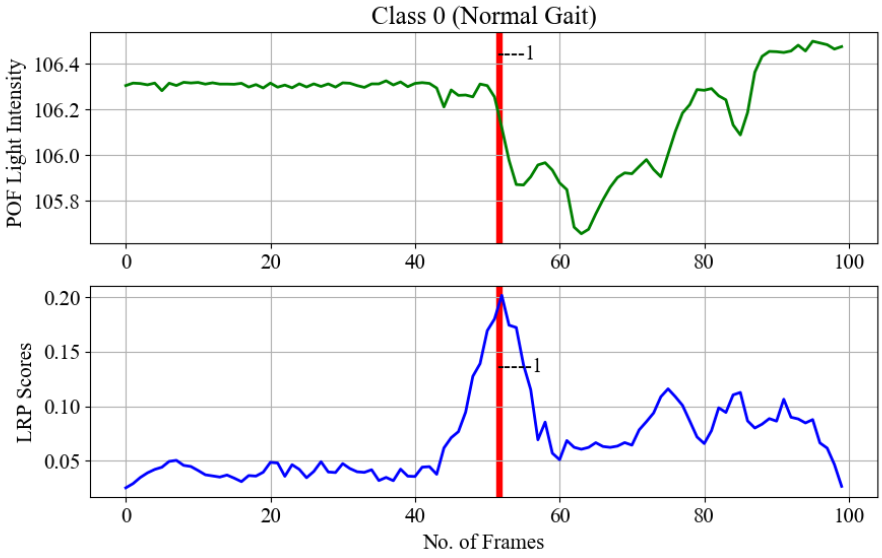
inherent noise in the latter suggests a focus on the gait events on the surface of the carpet, rather than particular computed numbers.

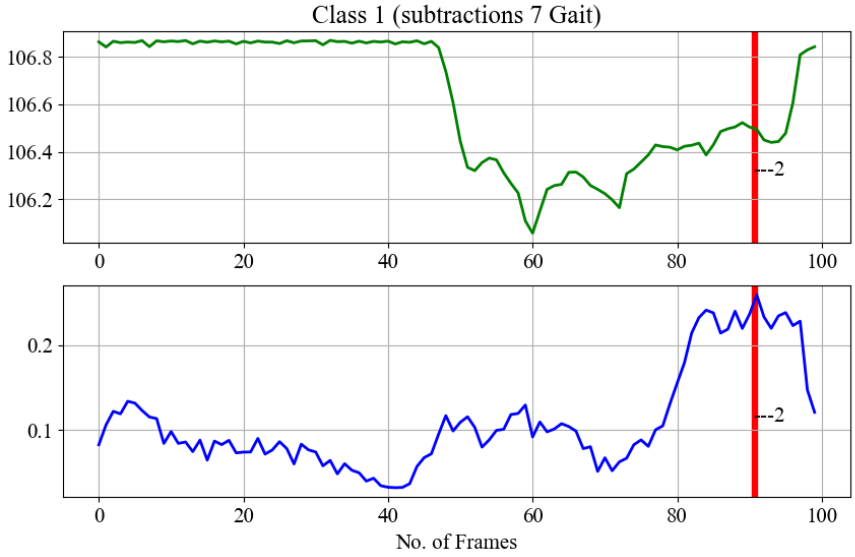
Therefore, we use (equation 5.3) on the LRP output to generate a heat map, which is possible to overlap with the temporal activity on the carpet. The overlap with events numbered from 1 to 9 on figure 5.7 implies that the effect of cognitive load on gait can be summarized as follows.

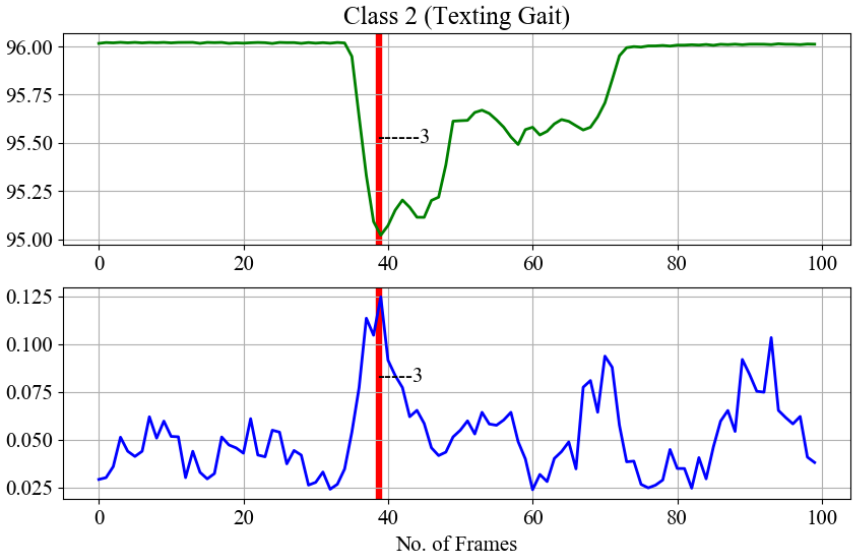
i. Normal gait: the signals most influential for the CNN to identify this class are after the heel strike and before the foot-flattening (figure 5.6, event numbers 1, 4, 7).
ii. Walking while performing 7s subtraction: the CNN classification of this gait is based strongly on the transition between foot-flattening and opposite toe-off (figure 5.6, event numbers 2, 5, 8).
iii. Walking while writing a text in smartphone: LRP scores for this event, are based on the transition between foot-flattening and double or single support (figure 5.6, event numbers 3, 6, 9).

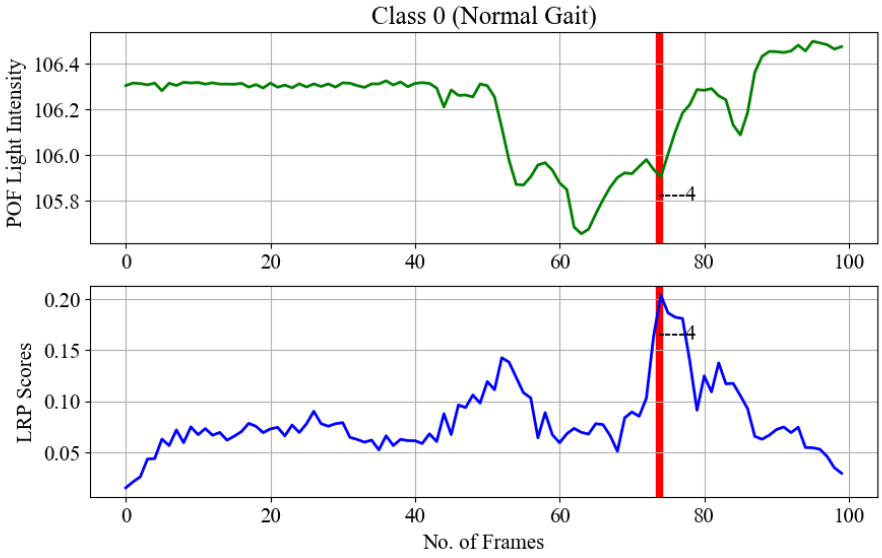
To assess the generalization of these gait events findings using LRP on all the test samples for each class, we apply the removal of region based on both LRP Deep Taylor decomposition MoRF and random region and re-predicting gait class. As shown in figure 5.5 (b) the model reaches lowest performance at 20% accuracy where the gait classes have to take a random prediction. Furthermore, it can be inferred from figure 5.5 that the LRP is effective in finding the most relevant region. The strong decay of the network prediction performance suggests that LRP maps are identical for each class and no further visual examination of the heatmaps is needed.

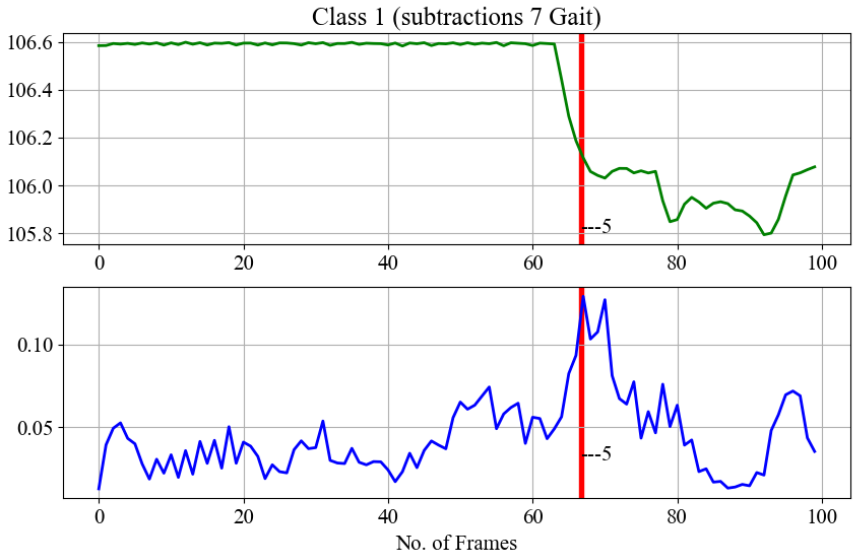
Overall care should be taken while inferring from these early results because human interpretability is hard to define in general [24], and “different users may have different capabilities at reading explanations and making sense of the features that support them” [24]. The LRP analysis indicates that subjects’ normal gait is characterised by an emphatic heel strike (high GRF due to the heel landing); gait while performing 7s subtraction highlights flat-foot (shift of emphasis away from hill strike due to less GRF resulting from weakened heel landing), gait while texting is characterised by weak support (compromised balance while walking).

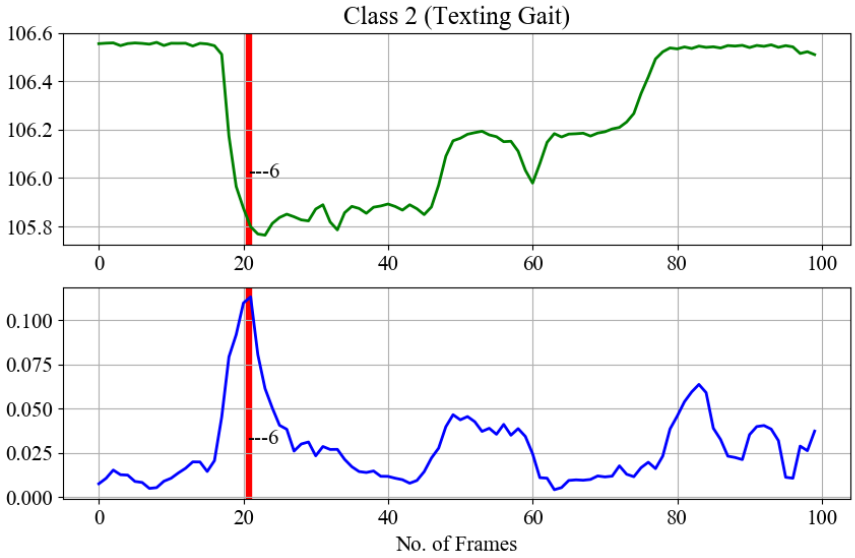

Class 0 (Normal Gait)
POF Light Intensity
106.4
106.2
106.0
105.8
---1
0
20
40
60
80
100
LRP Scores
0.20
0.15
0.10
0.05
---1
No. of Frames

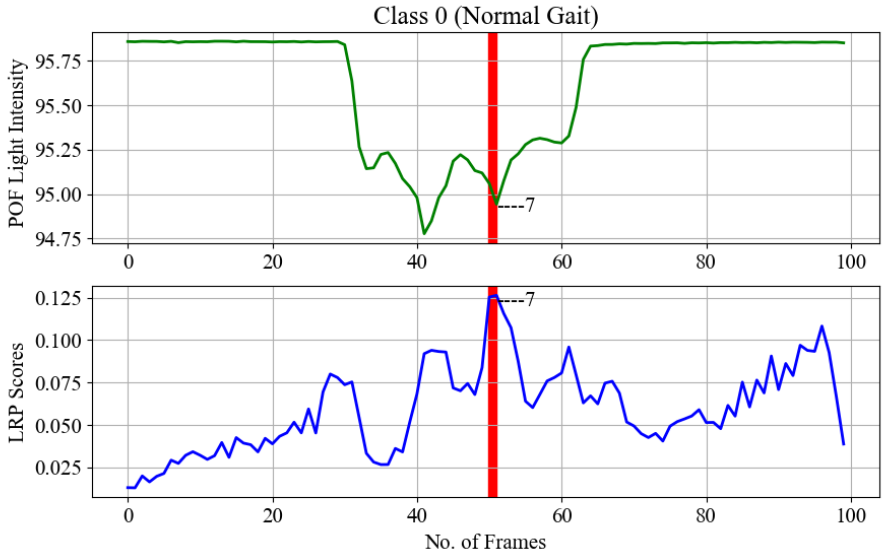

Class 1 (subtractions 7 Gait)
106.8
106.6
106.4
106.2
---2
0.2
0.1
---2
No. of Frames

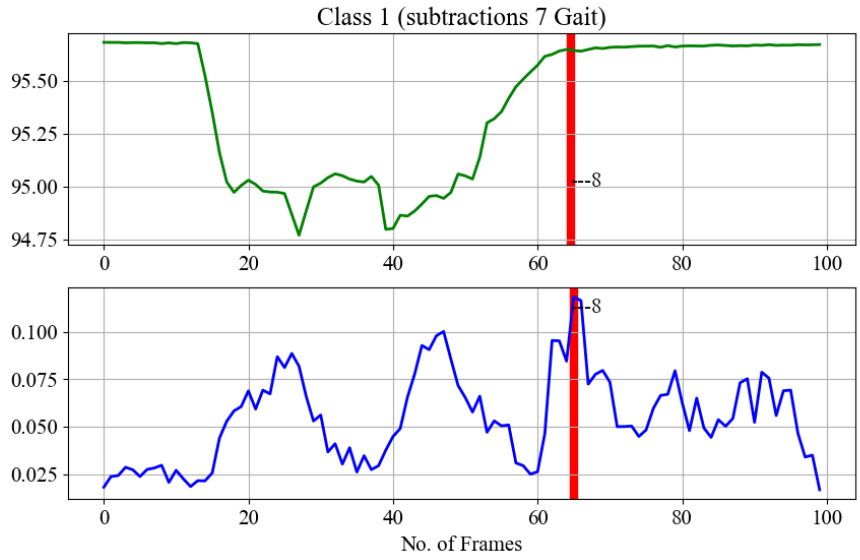

Class 2 (Texting Gait)
96.00
95.75
95.50
95.25
95.00
------3
0.125
0.100
0.075
0.050
0.025
---3
No. of Frames


Class 0 (Normal Gait)
POF Light Intensity
106.4
106.2
106.0
105.8
----4
0
20
40
60
80
100
LRP Scores
0.20
0.15
0.10
0.05
----4
No. of Frames


Class 1 (subtractions 7 Gait)
106.6
106.4
106.2
106.0
105.8
---5
0.10
0.05
No. of Frames


Class 2 (Texting Gait)
106.6
106.4
106.2
106.0
105.8
---6
0.100
0.075
0.050
0.025
0.000
No. of Frames


Class 0 (Normal Gait)
POF Light Intensity
95.75
95.50
95.25
95.00
94.75
----7
0
20
40
60
80
100
LRP Scores
0.125
0.100
0.075
0.050
0.025
----7
No. of Frames


Class 1 (subtractions 7 Gait)
95.50
95.25
95.00
94.75
---8
0.100
0.075
0.050
0.025
---8
No. of Frames

Class 2 (Texting Gait)
95.50
95.25
95.00
94.75
-----9
0.10
0.05
-----9
No. of Frames

Figure 5. 7. LRP methods are applied on three subjects' testing data (three pairs of graph rows), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. SA of gait spatiotemporal signals: green; SA for LRP relevance signals over gait temporal period: blue. Vertical red bars with numbers display consistency with gait events as per figure 5.4: 1- Initial foot-flattening, 2- Heel strike and Foot-flattening, 3- Single support, 4- Between opposite heel strike and double support, 5- Foot-flattening, 6- Between foot-flattening and single support, 7- Between heel strike and double support, 8- Toe-off, 9- Between foot-flattening and single support.

In this study we examine gait events influenced by cognitive load. This gives insight into which parts of gait is influenced rather than step time using smartphone [166], or monitoring each participants response while walking on a treadmill [181]. The results are promising, but apart from assessing the particular classification accuracy numbers, the main weakness in the conclusions comes from the limited number of subjects yet included in the currently reported starting phase of this work. A larger number of subjects, grouped in clusters of 10-15 (each group data is processed separately and combined with other groups), is expected to yield better results in studying the effect of cognitive load on gait using LRP. Immediate future work will involve a larger number of subjects, better coverage of age variations while maintaing gender balance.

It is worth mentioning hypotheticaly that the results reported at this point, and the LRP decomposition of gait classifications in particular, indicate a possible bridge between the output of artificial intelligence systems for processing of high quality gait data and decisions based on visual observations by humans, or quantitative parameters derived from such observations, (c.f. figure 3.1 in chapter 3) which are tested and routinely implemented in current practice. In healthcare alone, these approaches may address some of the significant interest in the detection of the onset of Parkinson disease [176] and Alzheimer's [177], freezing of gait [178], and fall risks [169]. Other potential spin-offs are in the areas of biometrics and security.

# Chapter. 6 Gait Under Cognitive Load Part (2)

# Spatiotemporal Analysis by Deep Learning of Gait Signatures from Floor Sensors

**Abdullah S. Alharthi, Alexander J. Casson, and Krikor B. Ozanyan**

## Abstract

The recognition of gait pattern variation is of high importance to various industrial and commercial applications, including security, sport, virtual reality, gaming, robotics, medical rehabilitation, mental illness diagnosis, space exploration, and others. The purpose of this paper is to study the nature of gait variability in more detail, by identifying gait intervals responsible for gait pattern variations in individuals, as well as between individuals, using cognitive demanding tasks. This work uses deep learning methods for sensor fusion of 116 plastic optical fiber (POF) distributed sensors for gait recognition. The floor sensor system captures spatiotemporal samples due to varying ground reaction force (GRF) in multiples of up to 4 uninterrupted steps on a continuous 2x1 m area. We demonstrate classifications of gait signatures, achieving up to 100% F1-score with Convolutional Neural Networks (CNN), in the context of gait recognition of 21 subjects, with imposters and clients. Classifications under cognitive load, induced by 4 different dual tasks, manifested lower F1-scores. Layer-Wise Relevance Propagation (LRP) methods are employed to decompose a trained neural network prediction to relevant standard events in the gait cycle, by generating a “heat map” over the input used for classification. This allows valuable insight into which parts of the gait spatiotemporal signal have the heaviest influence on the gait classification and consequently, which gait events, such as heel strike or toe-off, are mostly affected by cognitive load.

Manuscript published at the IEEE Sensors Journal. Received March 26, 2019; revised July 10, 2019; accepted July 11, 2019. Date of publication July 15, 2019; date of current version October 4, 2019. The work of A. S. Alharthi was supported by the Government of Saudi Arabia. The work of S. U. Yunas was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC DTP) under Grant EPN509565/1. The associate editor coordinating the review of this paper and approving it for publication was Dr. Marco J. Da Silva. (Corresponding author: Abdullah S. Alharthi.) The authors are with the department of Electrical and Electronic Engineering, The University of Manchester, Manchester M1 3BU, U.K. (e-mail: abdullah.alharthi@manchester.ac.uk; syed.yunas@manchester.ac.uk; k.ozanyan@manchester.ac.uk). Digital Object Identifier DOI: 10.1109/JSEN.2021.3078336

# 6.1 Introduction

Gait recognition has been intensively studied in recent years for the best achievable accuracy in distinguishing a certain target case in a plethora of applications, e.g. in healthcare, biometrics [131] (or chapter 3) and authentication for surveillance and forensics [10]. Conventional gait sensing technologies are borrowed from well populated research areas, such as image and video recognition, wearable sensors [131], and speech recognition [179]. Previous studies proposed cameras to record video sequence of the body motion [180], wearable sensors to acquire the limbs' trajectories and body posture [144], or microphones to capture the sound of footsteps [179]. These methods generally perform well, however the quality of data embodying the spatial and temporal aspects of gait is affected by clothing's conditions, angles in video sequences, the limitation in the format of the captured signals in wearable sensors and motion capturing systems, or the presence of other environment acoustic noise in the captured footstep sound.

An alternative approach is to use force plates to acquire the ground reaction force (GRF) [110], considering that the interaction with the walking surface is at the point of contact with the environment, which cannot be avoided or modified at will. However, force plates cover small areas, and the user may often need to artificially modify their gait aiming to step at a certain location. A preferable choice for GFR sensing of gait would be a large area unobtrusive floor sensor system, delivering high resolution spatiotemporal signals, which requires minimal (if any) cooperation or special attention by the user and allows long period, continuous data capture [131].

Due to the ambulatory nature of human gait, the environment varies, directly or indirectly rendering an individual's gait inconsistent. A growing body of studies demonstrates that attention demanding tasks alters the walking pattern in all healthy individuals in response to dual tasking, and gait should not be regarded as an automated motor activity that receives minimal higher-level cognitive input [163], [164]. Certainly, deliberate gait is essential for humans to navigate within often complex environment's while concurrently execute cognitive demanding tasks while walking to reach a desired destination. The progressive adjustment of gait is regarded as a way for individuals to minimize the energy cost during the gait cycle, while walking based on environment's

burdens [181], [182]. Humans continuously optimize energy cost in real-time, even if the energy savings while walking are small [183]. Thus, the energy conservation drives changes in gait dynamics, based on cognitive processing of information.

Furthermore, person's emotional state, such as pride, happiness, fear and anger, are found to affect gait [44]. However, the inconsistency of gait can be advantageous for gait biometric verification since gait under cognitive load would carry a unique signature due to individual variations in the cost of processing cognitive information. Another possible scenario is to detect and characterize cognitive decline with age e.g. facilitated by floor-embedded inexpensive sensor systems. Such a scenario is relevant to clinical diagnosis of mobility decline and impairments in gait and postural control due to natural aging process [184]; fall risk [169], [170] as well as mental decline, e.g. Alzheimer's disease, dementia with Lewy bodies, vascular dementia, Parkinson's disease [172].

This work utilizes floor sensors and implements an original protocol for recording gait under cognitive load ("dual tasks"). For the purpose, a healthy subject's natural gait is used as a control recording for the acquisition and processing of gait data under a cognitively demanding task. This work addresses in a novel way a number of challenges, identified in literature and by own observations, pertaining to the physical sensing layer, as well as the data processing. Signals are recorded with an original floor sensor system, specially designed, and built for spatiotemporal sampling with multiple plastic optical fiber (POF) distributed sensors [106],[185],[186],[187]. Multiple sensor fusion is achieved by deep learning with convolutional neural networks (CNN) used to classify subjects' gait. Explainable [188] CNN methodology is applied to relate achieved classifications to the input data, allowing visually observable gait events to be identified as most relevant for a certain class. The explainable methodology is based on the existing well-defined cyclic locomotion patterns in healthy human gait, allowing to query what part of the gait cycle is essential for recognition and what are the parts acting as a background (irrelevant for the classification) by CNN processing. Layer-Wise Relevance Propagation (LRP) techniques are applied in this work to interpret the CNNs prediction and identifying the gait events with largest weight for gait recognition.

The rest of the chapter is organized as follows. After introducing the necessary background in section 6.2 and methodology in section 6.3, the experiments and results

are presented in section 6.4. Substantial discussion of the results is included in section 6.5, followed by conclusions in section 6.6.

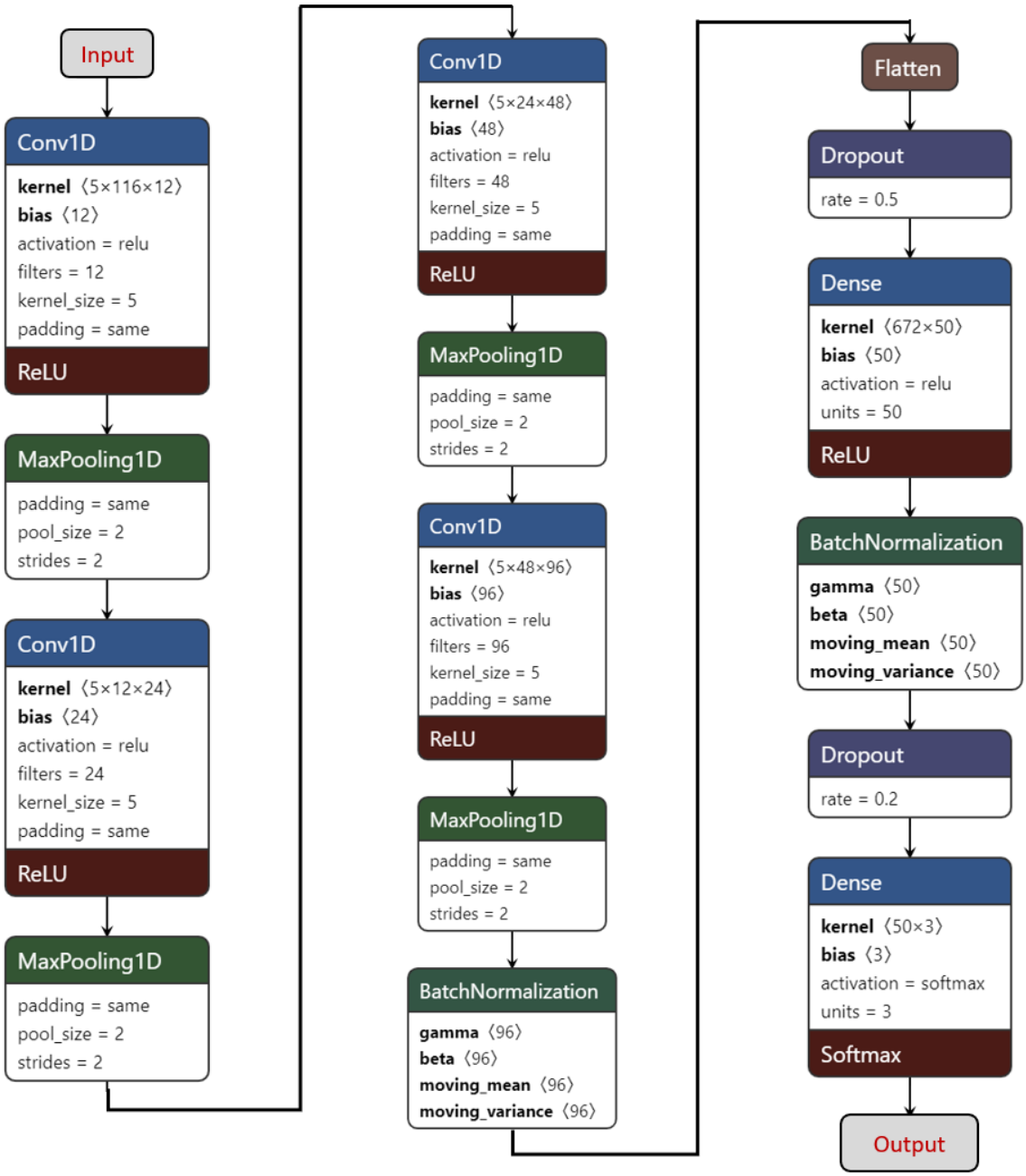



Figure 6.1. Proposed CNN network architecture for gait classification. Color-coding of boxes: convolution layers and fully connected layer (blue); pooling layers (green); flattening layer (brown); dropout layers (navy); input at the top and softmax output layers (gray). The diagrams are generated using Neutron [192] from GitHub repository based on the models' weights and biases.

## 6.2 Background

### 6.2.1 Deep CNNs and LRP

The deep CNNs core building block is the convolution operator, with the term 'deep' referring to the number of layers. A convolution layer learns a high level of abstraction and pattern by applying convolution operations, with the aim to extract features' representations automatically.

The deep CNNs utility is based on the ability to model complex relationships between inputs and outputs and find patterns in divergent data with high background levels. Widely researched applications, presenting weak features on high background, are face [189], text [190] and speech recognition [191]. Since gait is characterized by events occurring naturally in all healthy humans, a comparatively small number of weakly manifested features can be picked up to make a classification in biometrics or indicate a health condition by detecting spatiotemporal deviations from gait previously categorized as "normal" for a specific individual.

For most data classification cases a nonlinear deep ANN model acts as a black box, since the reasons as to why the models reached such decisions are not straightforward to trace to physical events. LRP [188] is a backward propagation method applied in interpretable deep learning, which identifies which parts of the ANN input vector carry most weight in the model prediction. This method has proven to be successful in a number of applications [189],[190],[191] to identify the divergent features for a model to make prediction, thus LRP is promising for unravelling the spatiotemporal signal and scrutinizing gait events, to the benefit of developing a certain application, e.g. aid to clinical assessment or to biometric verification.

### 6.2.2 iMAGiMAT System

The iMAGiMAT footstep imaging system is an original Photonic Guided-Path Tomography floor sensor head [106],[185],[186],[1187]). It can record unobtrusively temporal samples from a number of strategically placed distributed POF sensors on top of a deformable underlay of a commercial retail floor carpet. Each sensor comprises of low

cost POF (step-index PMMA core with fluorinated polymer cladding and polyethylene jacket, total diameter 1mm, NA=0.46) terminated with a LED (Multicomp OVL-3328 625nm) at one end and a photodiode (Vishay TEFD4300) at the other. The sensors constitute a carefully designed set to allow collaborative sensor fusion and deliver spatiotemporal sampling adequate for discerning gait events. The 1m x 2m area system is managed by 116 POF sensors, arranged in three parallel plies, sandwiched between the carpet top pile and the carpet underlay: a lengthwise ply with 22 POF sensors at $0^{o}$ angle to the walking direction and two independent plies, each consisting of 47 POF sensors, arranged diagonally at $60^{o}$ and $-60^{o}$ respectively (see [106], figure 6 therein).

The electronics is contained in a closed hard-shell periphery at carpet surface level, and is organised in 8-channel modules: LED Driver boards as well as input transimpedance amplifier boards to receive the data and send it to a CPLD (Complex programmable logic device) to reformat the data for processing by a Raspberry pi single board computer for export via Ethernet/WiFi.

The operational principle of the system is based on recording the deformation caused by the GRF variations, as bending affects the POF sensors transmitted light intensity is affected by surface bending. This captures the specifics of foot contact and generates robust data without constraints of speed or positioning anywhere on the active surface.

### 6.2.3 Gait

The details of the GRF dynamics follow the gait cycle, as the 7 intervals defined by the contact of one or both feet with the walking surface (see chapter 3, figure 3.1 therein). The full cycle gait intervals (events) for a single foot are: A- Heel strike or Initial contact; B- Loading response or foot flat and double support; C- Mid-stance; D- Terminal stance or Heel-off; E- Pre-swing; F- Initial swing and Mid-swing. In a healthy gait cycle these 7 intervals are grouped in Stance Phase (A, B, C, D - 60% of the gait cycle, when the foot is in contact with the ground) and Swing Phase (E, F, G - 40% of the gait cycle, foot swinging and not in contact with the ground).

Further to our recent topical review (chapter 3) on gait recognition, gait recognition literature in the past two years has been focused on solving the view- and clothing invariant problem for video sequences with more advanced machine learning methods,

such as generative adversarial network [120],[121]. Recent work on wearable and floor sensors has been applied for medical applications such as the impact of muscle fatigue on gait characteristics [124], health monitoring [125] and age-related differences [126].

## 6.3 Methodology

### 6.3.1 Data Acquisition and Pre-Processing

Twenty-one physically active subjects aged 20 to 40 years, 17 male and 4 females, without gait pathology or cognitive impairment, participated in this experiment. The study was carried out under the University of Manchester Research Ethics Committee (MUREC), ethical approval number 2018-4881-6782. All participants were informed about the data recording protocol in accordance with the ethics board general guidelines and each subject written consent was obtained prior to experiments. Each participant was asked to walk normally, or while performing cognitively demanding tasks, along the 2 m length direction of the iMAGiMAT sensor head. The captured gait data is unaffected by start and stop, as it is padded on both ends with unrecorded several gait cycles before the first footfall on the sensor. With a capture rate of 20 timeframes/s (each timeframe comprising the readings of all 116 sensors), experiments yielded 5s long adjacent time sequences, each containing 100 frames. The recorded gait spatiotemporal signals were able to capture around 4 to 5 uninterrupted footsteps at each pass.

Five manners of walking were defined as normal gait plus four different dual tasks, and experiments were recorded for each subject, with 10 gait trials for each manner of walking in a single assessment session; thus the total number of samples is $10 \times 5 = 50$ per-subject. The five manners of walking are defined as follows:

- Manner 1, Normal Gait: walking at normal self-selected speed.
- Manner 2, Gait while listing to a story: audio input through headphones, then answer questions after gait recording is completed.
- Manner 3, Gait with serial 7 subtractions: normal walking speed attempted, while simultaneously performing serial 7 subtractions (count backward in sevens from a given random 3-digit number).

- Manner 4, Gait while texting: normal walking speed attempted, while simultaneously typing text on a mobile device keyboard.
- Manner 5, Gait while talking: walking at normal self-selected speed while talking or answering questions.

A set of measured data as $x_{n,s} = [x_{n,1} \quad \cdots \quad x_{n,116}] \in \mathbb{R}^{n\times 116}$ is harvested from the iMAGiMAT system, where $n$ is the number of the data block (100 frames) and $s$ enumerates the POF sensors. A total number of 1050 samples are recorded for 21 subjects and placed in a 3D matrix of dimensions $1050 \times 100 \times 116$. The recorded amplitude of data varies due to the weight of each subject, therefore, data standardization is implemented as a pre-processing step, to ensure that the data is internally consistent, such that the estimated activations, weights, and biases update similarly, rather than at different rates, during the training process and testing stage. The standardization involves rescaling the distribution of values with a zero mean unity standard deviation, using the following equation:

$$\widehat{x_{n,s}} = \frac{x_{n,s} - \mu(x_{n,s})}{œ(x_{n,s})} \tag{6.1}$$

Here $\widehat{x_{n,s}}$ is gait data rescaled so that $\mu$ is the mean and œ is the standard deviation.

## 6.3.2 CNN Structure

Gait analysis in this study is handled as a supervised learning process. Here, we propose a CNN model, based on extensive experimentation with different algorithms and architectures in our previous work [173],[131],[145],[150],[193] as automatic feature extractor and classifier. The model shown in figure 6.1 maps the gait spatiotemporal signal $\widehat{x_{n,s}}$ to an output label $y$ by learning an approximation function $y = f\left(\widehat{x_{n,s}}\right)$. The network consists of an input layer, 4 convolution layers, 4 pooling layers, 2 fully connected layers, 1 batch normalization, and an output layer with a softmax classifier. The set of 12 stacked layers in figure 6.1 utilizes Conv1D kernels (filter size × number of feature maps × number of filters), MaxPooling strides of 2 and pool size of 2.

The network total parameters are 71,669, with 71,477 trainable and 192 non-trainable parameters. The parameters can be calculated from figure 6.1 as 71,477= ∑kernels +

bias. A stride of 1 and the same padding is used in the convolutional layers to output the same length of the original input shape. To improve the model performance, a regularization method is utilized as follows: (1) Batch normalization (to normalize the activations of the previous layer at each batch, by maintaining the mean activation close to 0 and the activation standard deviation close to 1 [194]). (2) The Batch normalization followed by a dropout [195] of size 0.5, after the last MaxPooling layer was flattened, by transforming a matrix to one single column vector. (3) Dropout of size 0.2 before the output layer.

An Adam (adaptive moment estimation) [157] is utilized to train the model. The used optimizer parameters are $\alpha$ =0.002, $\beta$1 =0.9, $\beta$ 2 =0.999, $\varepsilon$=1e-08. Here $\alpha$ is the learning rate or the fraction of weights updated; $\beta$ 1 and $\beta$ 2 are the exponential decay rates for the first and second moment estimates, respectively; $\varepsilon$ is a small number to avoid division by zero. The loss is computed using categorical cross-entropy in every iteration to minimize the network error [196]. The convolutional layers weight parameters are initiated with a Glorot uniform [197] with zero bias. The model is trained and validated (for several experiments) using a batch size of 100 samples for each iteration; 200 epochs are found optimal to train the model. The training and validation sizes are set to be 70% and 10% respectively, where 20% is reserved for testing the model accuracy. The model is trained, validated, tested for several runs with data split using different random state parameter with different seeds. The mean performance and standard error are used to report the accuracy as follows:

$$SE = \frac{\sqrt{\frac{\sum(F1-\mu)^2}{q}}}{\sqrt{q}} \tag{6.2}$$

Here F1 is the model F1 score, $\mu$ is the F1 scores mean and $q$ is the number of F1 scores accuracies. The numerator is the standard deviation.

### 6.3.3 LRP

In this paper the POF sensor spatiotemporal signal $\widehat{x_{n,s}}$ is learned by the CNN model to make a prediction of gait class $c$ using a classification function $f_c(x)$. This function can be interpreted as a "heat map", or relevance scores, to quantify certain signal changes associated with gait intervals. The interpretation of this function is achieved by

decomposing the classification decision into regions of relevance contributed to the model classification scores. This approach is known as LRP [21], which is based on backpropagated quantity preserved between intermediate nodes of adjacent layers. The LRP produce a heat map identical to the original spatiotemporal signal $\widehat{x_{n,s}}$, which can be overlaid on the original signal to highlight the contributions of gait intervals to the CNN model prediction of gait class. As the CNN consists of stacked layers, each layer has feature maps and can be treated as neurons, where neurons are activated according to [21]:

$$a_j^{1+l}=\sigma\left(\sum_i g_{ij} + b_j^{(l+1)}\right) \tag{6.3}$$

$$\text{with} \qquad g_{ij=}a_i^l\,\omega_{ij}^{(l,l+1)}$$

The sum is computed over all previous layer neurons that are connected to neuron *j*, where $a_i^l$ denote the activation of a neuron *i* in the previous layer in forward direction, and $g_{ij}$ denote the contribution of neuron *i* in layer *l* to the activation of the neuron j in layer *l* + 1. The neuron weight $\omega_{ij}^{(l,l+1)}$ is received in forward direction by neuron *i* from neuron *j* in the previous layer and $b_j^{(l+1)}$ is a bias term. The function $\sigma$ is a nonlinear monotonously increasing activation function to define the output of that node given an input, ReLU=max(0,x) passing only positive signals. During CNN supervisory training, these parameters are learned by computing the output $f_c(x)$ based on evaluating $a_j^{1+l}$ in a forward pass and updated by back-propagation using the model error. For the latter, computations are based on categorical cross entropy [180]. The LRP approach utilizes the CNN output $f_c$ for input $x_{n,s}$ and generates a "relevance score" $R_i^l$ for the $i^{\text{th}}$ neuron of layer $l$ and $R_j^{l+1}$ from the $j^{\text{th}}$ neuron in layer $l$+1, where the relevance conservation principle is satisfied as [20]:

$$\sum_i R_{i\leftarrow j} = \sum_j R_j = f_c(x) \tag{6.5}$$

# 6.4 Experiments and Results

In this section the experiments and results are illustrated in the following sections.

### 6.4.1 Overview

All algorithms for LRP computation are implemented in Python 3.7.3 programming language using Keras 2.2.4 [183], TensorFlow 1.14.0 and iNNvestigate GitHub repository [153]. All codes are run using a desktop with intel core i7 6700 CPU @3.4 GHz. After data standardization, the deep CNN model is applied on the dataset in order to test the validity of the algorithms for identifying gait signatures. We compared the CNN predictions to manually labelled ground truth in several experiments, based on individuals' identity and the changes to normal gait incurred by cognitive load. The models' classification performance is evaluated using confusion matrices. The performance of the LRP methods is examined in detail in subsection 6.4.4.

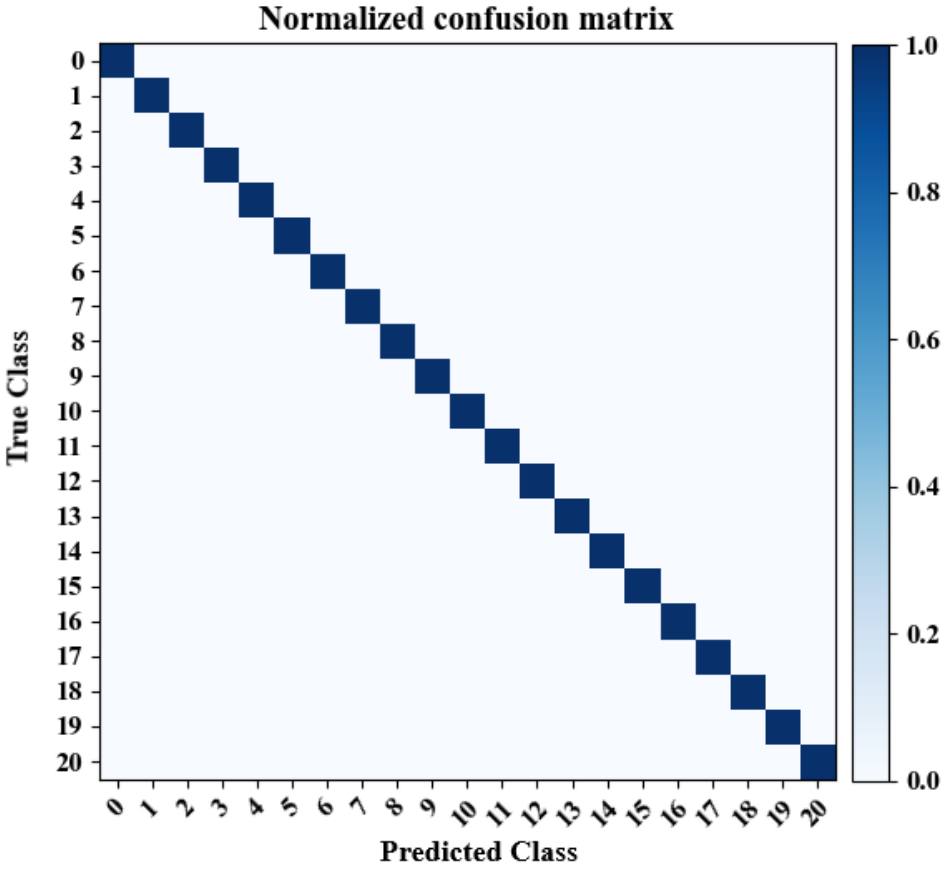
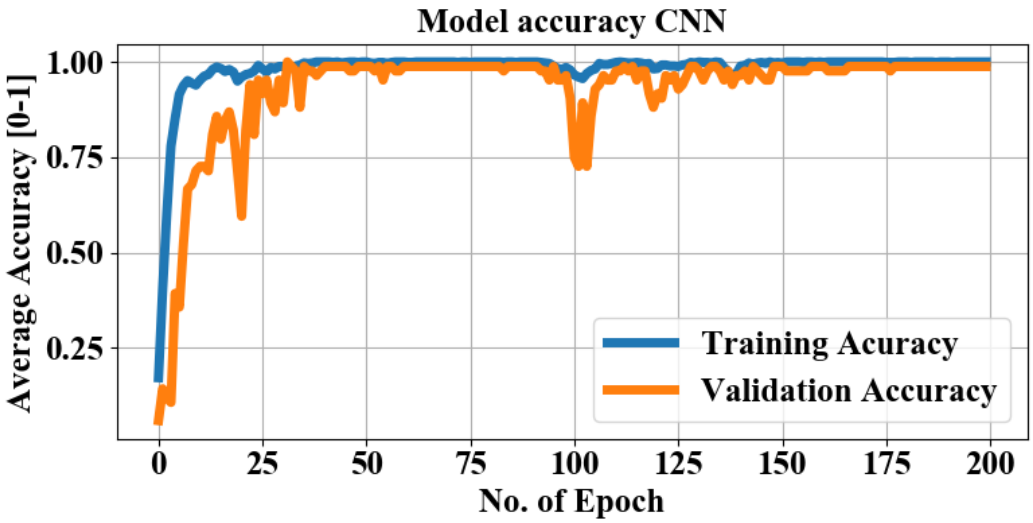
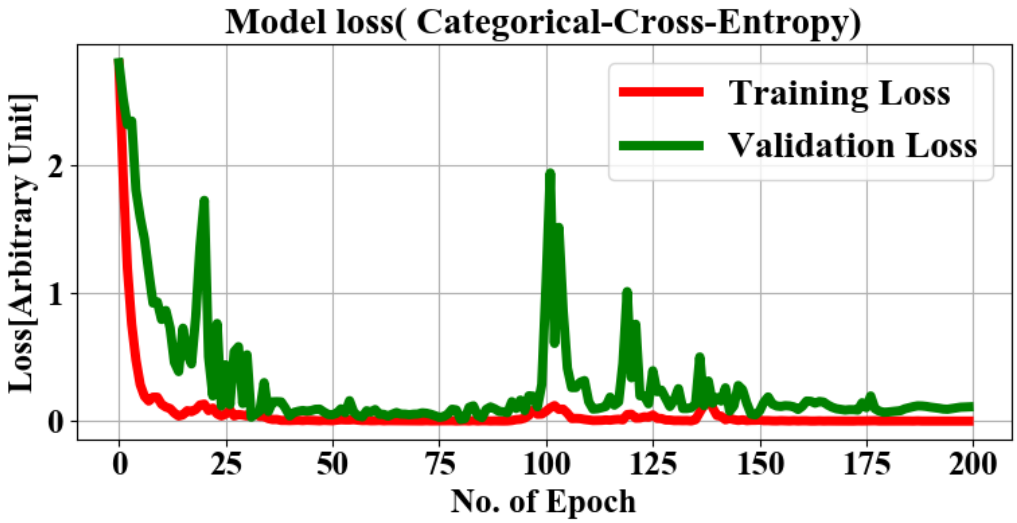



Figure 6.2. Gait signature classification confusion matrix for 21 subjects. The diagonal squares are the true positives, in this case 100% of classification and elsewhere are the false positives (0%).




(a)




(b)

Figure 6. 3. Model training and validation loss in (a) and model accuracy in (b) for 21 subjects in 5 classes.

## 6.4.2 Gait signature classification

The proposed model is trained, validated, and tested on m×n×s (m=number of samples, n=number of frames, S=number of POF sensors) spatiotemporal samples as K×100×116 for several runs using different *random state* parameters. m is chosen on the basis of

experimental protocols and the mean performance and standard error are used to calculate the accuracy.

Experiments are conducted to investigate the ability of the deep CNN to identify gait signature patterns by fusing 116 POF sensors in the model's deep layers, to extract gait patterns automatically in the following experiments.

Table 6.1. Models Classification Accuracy for Each Subject.

| Subject number | F1- score | Subject number | F1- score | Subject number | F1- score |
|---|---|---|---|---|---|
| 0 | 95% | 7 | 87% | 14 | 100% |
| 1 | 65% | 8 | 90% | 15 | 75% |
| 2 | 93% | 9 | 90% | 16 | 80% |
| 3 | 90% | 10 | 77% | 17 | 100% |
| 4 | 87% | 11 | 91% | 18 | 100% |
| 5 | 91% | 12 | 90% | 19 | 80% |
| 6 | 73% | 13 | 100% | 20 | 69% |

#### 6.4.2.1 Experiment 1: 21 Subject gait signature verification

To demonstrate the model's ability to verify the identity of a subject based on their gait signature, we assigned each subject's data a label numbered from 0 to 20, containing 50 samples of normal and cognitive load as explained in section 3.3.1. The model is trained, validated and tested on m=1050 samples with different *random state* parameters, and the mean performance and standard error are used to calculate the accuracy. The median classification confusion matrix is shown in figure 6.2, where the model achieved F1-score of 100% prediction and mean performance and standard error of $99.5 \pm 0.28$ %.

Figure 6.3 demonstrates the learning curve performance of the CNN over the iterations while training. We generate the training loss for each of the training sets and validation loss for each of the validation sets over the epochs. Figure 6.3a shows the average training and validation losses.

Table 6.2. F1-score predictions for binary classification.

| Data Group for Classification | 1 testing subject | 2 testing subjects | 4 testing subjects | Test with all subjects |
|---|---|---|---|---|
| Class 0 vs class 1 | 100% | 85% | 81% | 79% |
| Class 0 vs class 2 | 95% | 87% | 58% | 69% |
| Class 0 vs class 3 | 60% | 68% | 63% | 79% |
| Class 0 vs class 4 | 100% | 85% | 74% | 81% |

The training loss starts from 3 and gradually reduces to 0. The validation loss generally follows the training loss, with a few spikes, and stablizes after 150 epochs. As expected, accuracy increases with decreasing loss, demonstrated in figure 6.3b. The average training and validation accuracy stablizes after 150 epochs after a few spikes.

For additional testing of the model performance in real-life scenarios, we evaluate the model on imposter and client classification. The client's data are used for the model training and validation and only 20% of that data is used for testing, while the imposters data are only used at the testing stage. The model at the testing stage predicts the client's gait sample identity with an F1 score of 100% and unable to predict the imposter which returns 0% F1 score. This is achieved by taking 17 subjects as clients (m=850 samples), and 4 subjects as imposters (m=200 samples). Clients were split 70%-10%-20% for training validation and testing respectively. In the testing stage, the model was able to correctly distinguish imposters and clients in 100% of cases.

#### 6.4.2.2 Experiment 2: Gender Classification

To demonstrate the model ability to recognize gait signatures, we perform a two-class classification based on the gender of the subject using the normal gait and cognitive load samples. The model is trained and validated with 6 subjects (m=300 samples) including 3 males and 3 females. Model testing is by predicting the gait class of two new subjects (m=100 samples, never seen by the model), the selection of males and females are done

randomly. In this experiment the deep CNN prediction achieved F1-score of 95%, with 96% true positive prediction for the male samples, and 94% true positive prediction for the female samples.

Table 6.3. . F1-Score Predictions for Comparison of CNN with Classical Classifiers.

| Classifier | Experiment 1 | Experiment 3 |
|---|---|---|
| SGD | 77% | 42%, N=47% |
| KNN | 87% | 51%, N=81% |
| GPC | 5% | 22%, N=0% |
| CNN | 100% | 50%, N=92% |

N: True positive prediction of normal gait.

#### 6.4.2.3 Experiment 3: 21 Subjects Cognitive Load Classification

The aim of this experiment is to show that in healthy subjects the influence of cognitive load on gait varies from subject to subject and the normal gait can be predicted with higher true positive rates than predictions under cognitive load. Five types of gait signatures, normal and four cognitively demanding task patterns, are learned for 21 subjects. The performance observed for the 5 classes is shown in figure 6.4, as the median confusion matrix based on several runs with F1-score of 50% and mean performance and standard error of $48.25 \pm 1.03\%$.

The results show that normal gait is predicted by a true positive incidence of 92%, while there is notable confusion between the dual tasks performed by the 21 subjects. The different *random state* parameters return the same result, where the normal gait true positive prediction is higher than 90% and substantial confusion between the dual task cases. Figure 6.5 shows the CNN learning curves over the training iterations, where the training loss declines from 1.8 to 0 while the validation loss rises from 1.8 to around 4 for 200 epochs, resulting in low validation accuracy as per figure 6.5b to evidence severe overfitting.

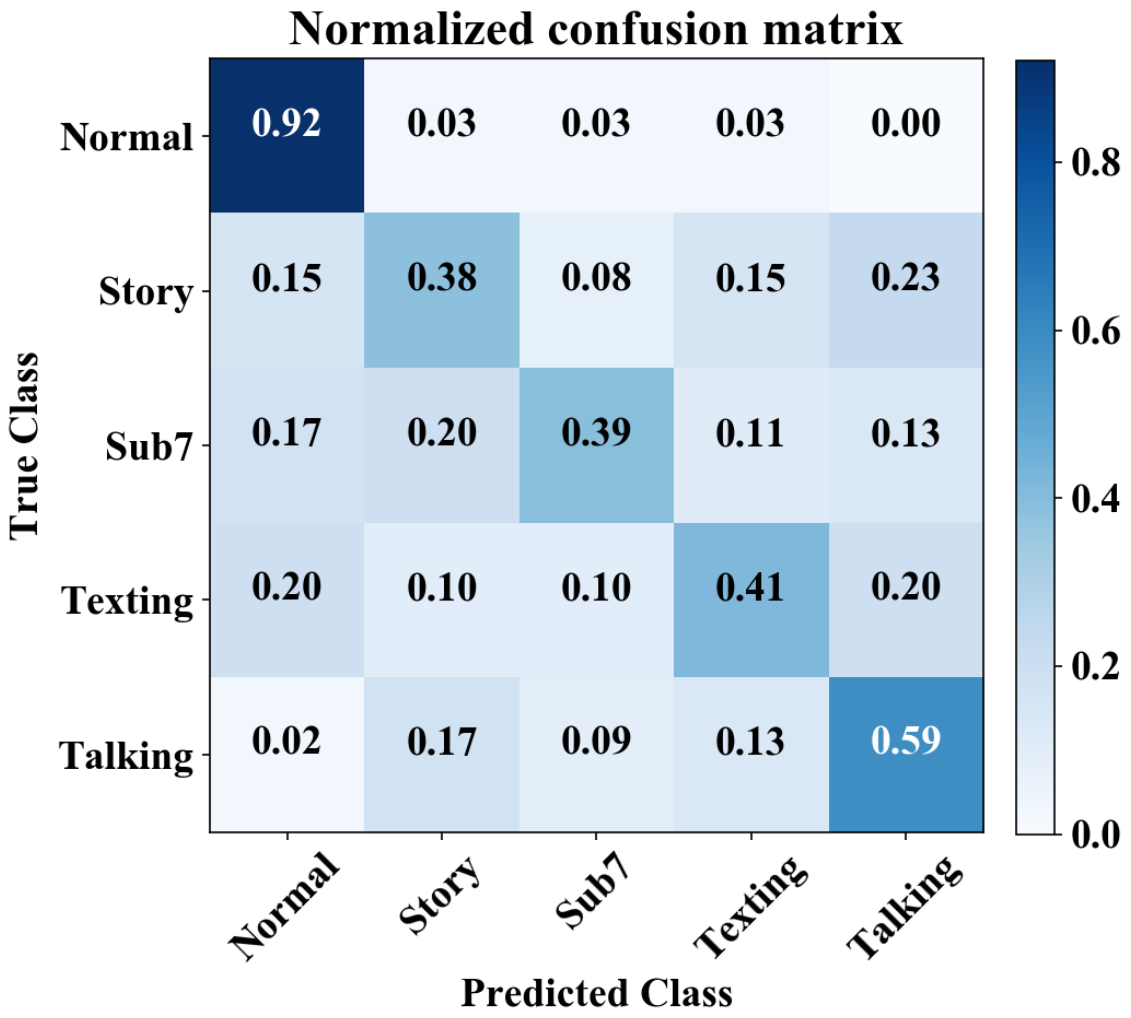
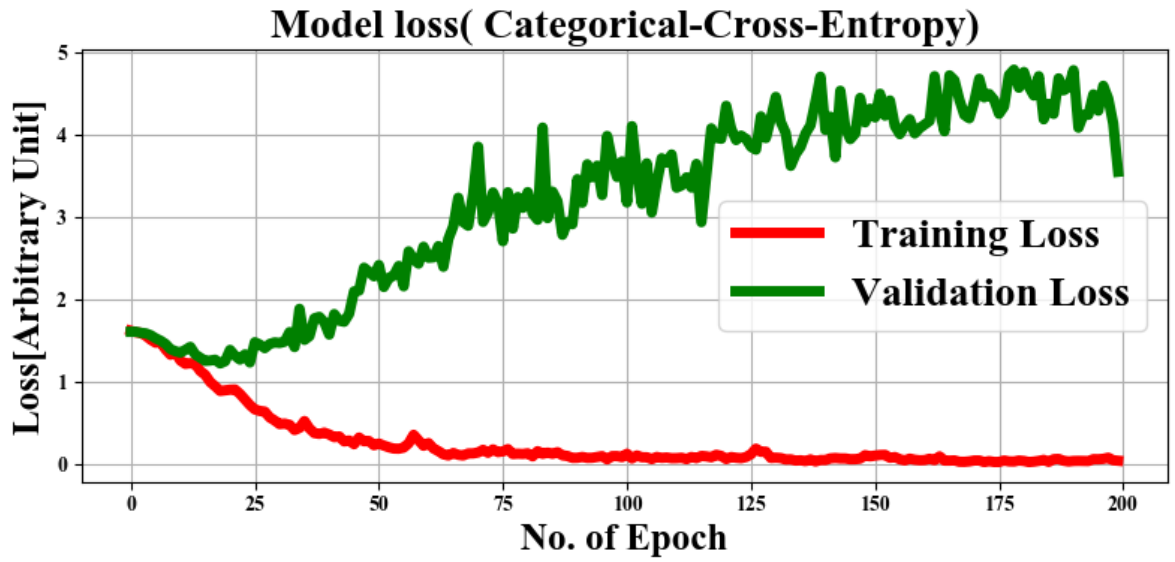


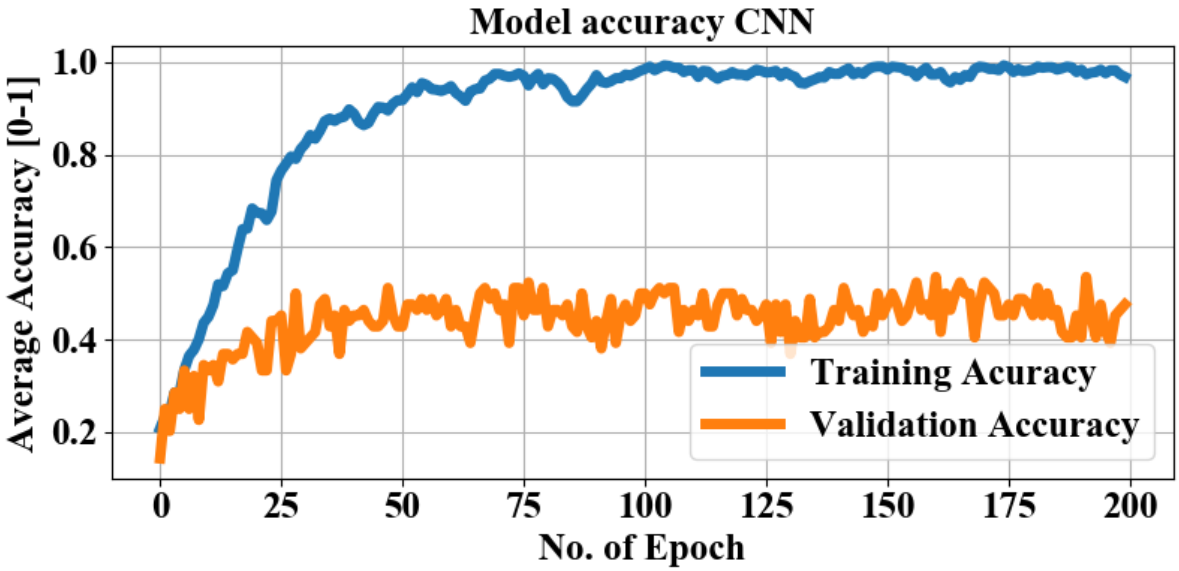

Figure 6.4. Confusion matrix for classification under cognitive load: 21 subjects, 5 classes.

#### 6.4.2.4 Experiment 4: Single Subjects Cognitive Load Classification

In this experiment gait patterns are investigated within each subject, to show that each subject gait under cognitive load can be learned and predicted. This is achieved by training, validating, and testing the CNN to classify each subject gait pattern using the normal gait and cognitive load. Each subject data is split using *random state* to cover all the 5 classes for testing with m=50 samples. The model evaluation using the F1- score is detailed for each subject in table 6.1. Gait data is predicted with more than 85% F1-score for 16 subjects, and for 6 subjects F1-scores are between 65% and 77%.




(a)




(b)

Figure 6.5. Model loss (a) and training and validation accuracy (b) under cognitive load, 21 subjects and 5 classes.

#### 6.4.2.5 Experiment 5: Binary Classification under Cognitive Load

To study patterns for each of the 4 dual tasks (M2-M5) representing variants of cognitive load, we organize the data into four groups so that binary classification performance to distinguish between gait under normal (class 0) and cognitive load (one of classes 1, 2, 3

or 4, depending on the particular data group) conditions can be studied separately for each dual task. The CNN is trained 16 times, implementing 4 runs with each of the 4 data groups. The F1-scores for each run are shown in table 6.2. The first run in each data group is based on training and validating the CNN on 20 subjects and test the model on 1 subject, to see if we can predict gait of one person from 20 people. In the second run, the numbers are 19 and 2, respectively; in the third – 17 and 4 respectively. The last run is based on splitting the data into 70% for training, 10% for validation and 20% for testing, using m=420 samples with *random state* of 200 seed parameters (since the accuracy doesn't change with the *random state seed*). As shown in table 6.2, the highest classification performance is achieved in the first runs (except for the group containing class 3). This is used essentially in the implementation of LRP to analyse the gait classes for that subject in the first run, as reported further comparison with Statistical Classifiers.

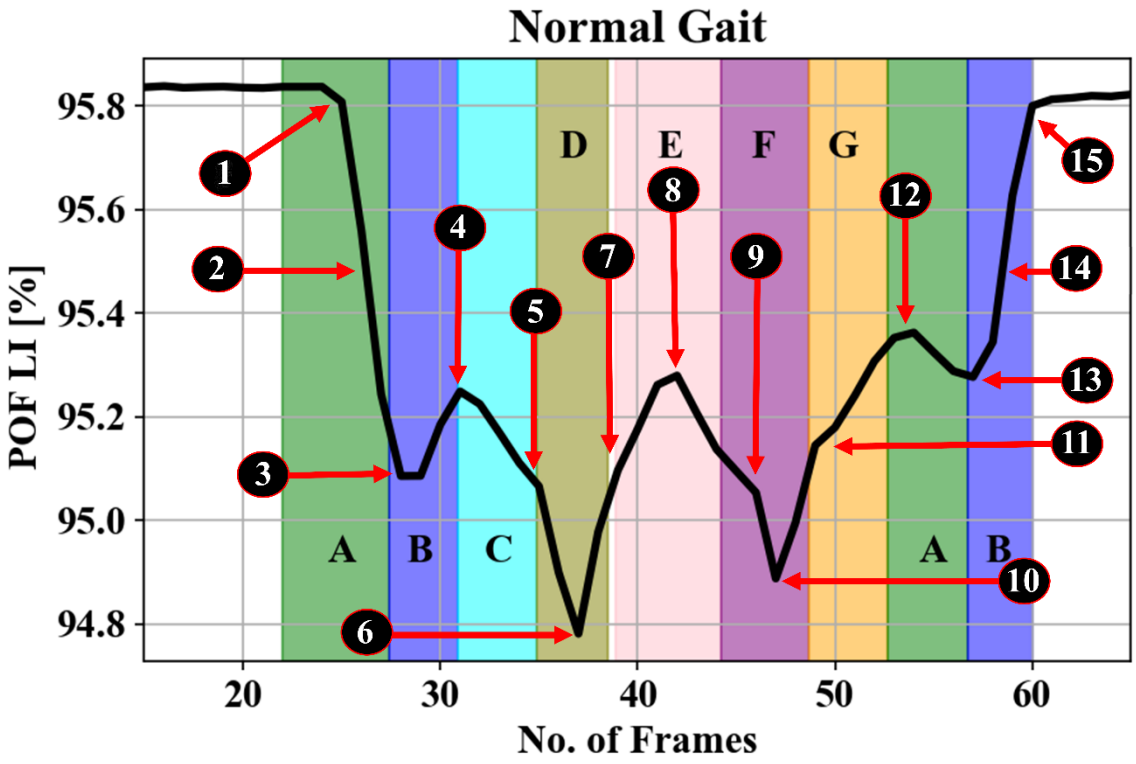



Figure 6.6. Representative gait cycle spatial average of a spatiotemporal signals (see equation 6.5). Gait events recorded by the sensors in a typical full gait cycle of two steps ([1] as A,B,C,D,E,F,G): 1- Heel strike, 2- Foot-flattening, 3- Single support, 4- opposite Heel strike, 5- opposite Foot-flattening, 6- Double support, 7-Toe-off, 8- Foot swing, 9-

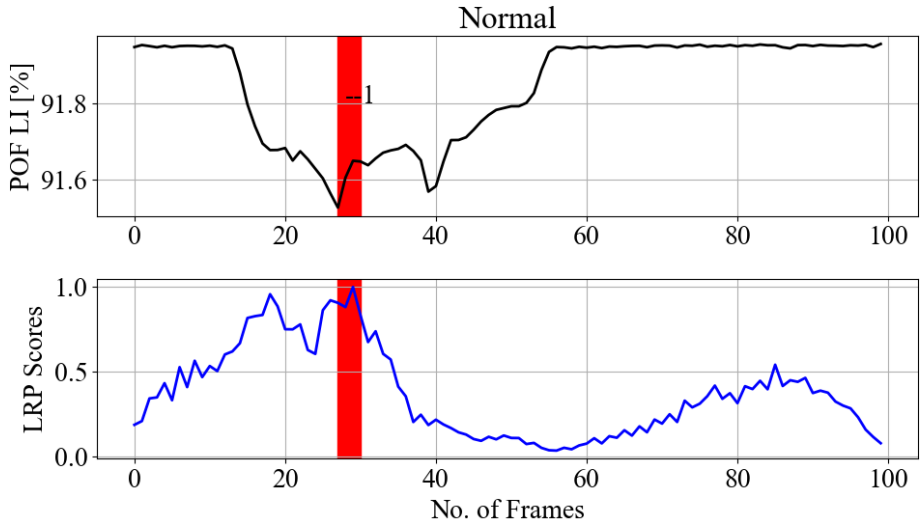
Heel strike, 10- Double support, 11- Toe-off, 12- Foot swing, 13- opposite Heel strike, 14- Single support, 15-Toe-off.

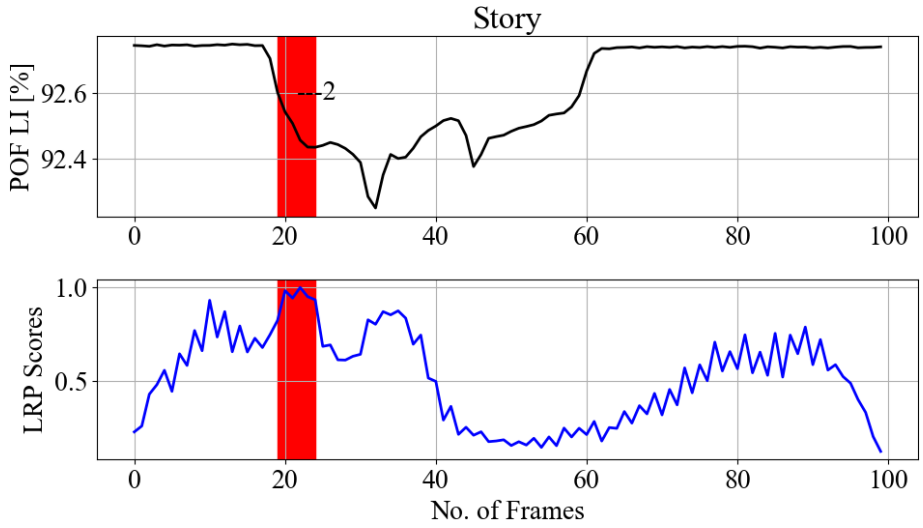
Gait signature recognition achieved high accuracy compered to Cognitive Load Classification; therefore, the validity of these achieved classification is verified with statistical classifier. Here we compare the classification results achieved by CNN in experiment 1 and 3 with statistical classifier algorithms, such as Stochastic Gradient Descent (SGD) [198], K-Nearest Neighbors (KNN) [199], and Gaussian Process Classifier (GPC) [200]. To change the format of the statistical classifiers input, the data are flattened to length 11600=100×116, with and samples of m =1050 for experiment 1 and 3. The classification F1-scores on experiment 1 and 3 are shown in table 6.3. GPC fails in the true positive prediction of normal gait, while KNN achieves the best classification results. However, the CNN outperforms the statistical classifiers for both gait signature recognition and normal gait prediction.

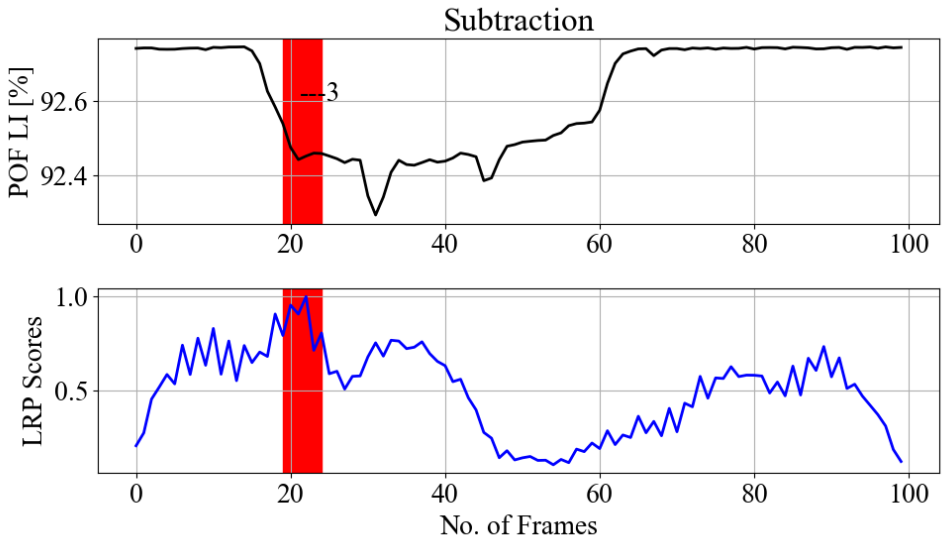
### 6.4.3 LRP Analysis of Gait Spatiotemporal Classifications

The focus of this section is to identify the features picked up by the model to classify gait under cognitive load. To obtain accurate LRP relevance scores $Ri$, the model true positive prediction should be high. Therefore, the gait class with high positive rate is considered for LPR analysis. The learned CNN model parameters in experiment 1, 3 and 5 were frozen for LRP analysis. Experiment 4 is to check if there is a variation in gait within a subjects; therefore it is not considered for LRP analysis. LRP Sequential Preset a Flat (LRP-SPF) [141] based on the LRP (see 6.3.3) was utilized for this work, as it has shown sensitivity to gait inconsistency using perturbation [193].

The iMAGiMAT system captures a sequence of periodic events as distinct, but similar cycles for each foot. This spatiotemporal sequence is generated by the change of light transmission intensity in the POF sensors: $x_i = [x_1 \quad \cdots \quad x_{116}] \in \mathbb{R}^{n\times 116}$. However, a typical interpretation of the gait cycle, based on visual observation, is derived much less from the spatial component than the temporal one.


Normal
POF LI [%]
91.8
91.6
1
0
20
40
60
80
100
LRP Scores
1.0
0.5
0.0
No. of Frames


Story
POF LI [%]
92.6
92.4
2
0
20
40
60
80
100
LRP Scores
1.0
0.5
No. of Frames


Subtraction
POF LI [%]
92.6
92.4
3
0
20
40
60
80
100
LRP Scores
1.0
0.5
No. of Frames

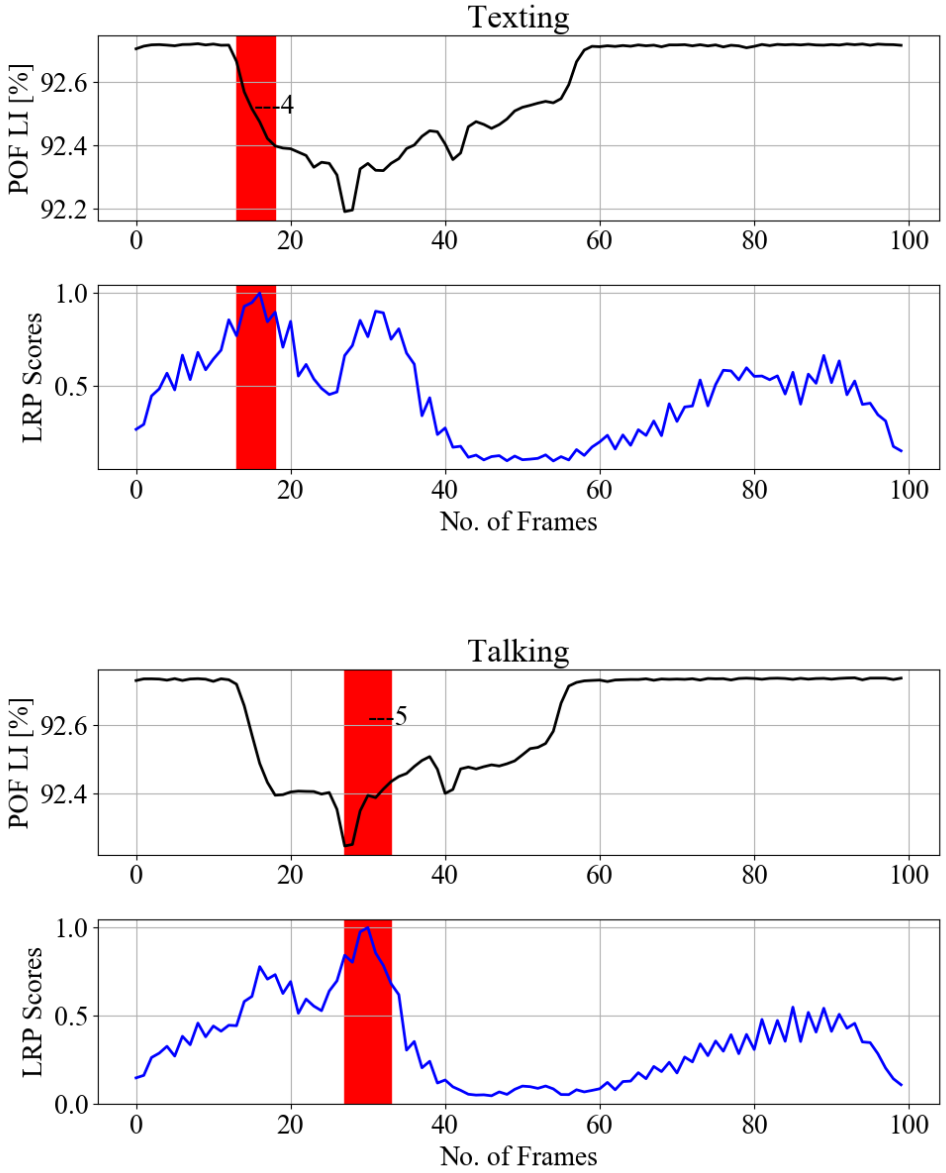



Figure 6.7. LRP methods applied on a single subject from experiment 1 testing data (each column is one pair), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. SA of gait spatiotemporal signals: black; SA for LRP relevance signals over gait temporal period: blue; POF LI (Plastic Optical Fiber Light Intensity). Vertical red bars with numbers display correspondence to gait events as per figure 6.6: 1,5- Loading response or Foot flat and Double support, 2,3,4 - Loading response or Foot flat and Single support.

Thus to progress towards interpreting the CNN classifications in terms of observable gait

events, we average over the spatial domain according to:

$$SA[n] = \frac{1}{s}\sum_{s=1}^{116}( x_{n,s}) \quad (6.5)$$

Here $x_{n,s}$ are the readings from individual sensors $s$ at a specific frame $n$ within each sample and $SA$ is the frame $n$ spatial average calculated as the arithmetic mean over all sensors. Figure 6.6 displays a typical $SA$ of the spatiotemporal gait signal, labelling the main gait events over a two-step gait cycle.

Figures 6.7 and 6.8 display randomly chosen samples of single subjects, returning 100% true positives prediction for gait signature verification in experiment 1; figure 6.9 displays randomly selected samples of normal gait classified with 100% true positives in experiment 3; figure 6.10 shows predicted gait samples in experiment 5 for a subject never seen by the model when the training set is 20 subjects.

The top panels in figures 6.7, 6.8, 6.9 and 6.10 display calculated $SA$ aligned against the relevance "heat map", generated from the calculated LRP scores and displayed in the bottom panels (to be discussed further in section 6.5). The SA temporal sequences have different values on the y axis due to the nature of the captured gait signal, which is influenced by the individual anthropometry of subjects

## 6.5 Discussion

### 6.5.1 Classification of Gait Signatures under Cognitive Load

The present study investigates the importance of cognitive load influence for gait inconsistency. We present a comparison of classification performance between 5 types of gait: normal and under cognitive load in 4 different tasks. Deep CNNs not only outperform, unsurprisingly, the classical classifier methods but also achieve an F1-score of 100% (see figure 6.2 and table 6.3) for gait signature verification in experiment 1 with 21 healthy adult's data, and 100% prediction of 4 imposters and 17 clients. The learning curve in figure 6.3 demonstrates the good match of the CNN methodology for gait verification tasks. The network parameters are updated via backpropagation to map gait

during training to 21 classes are correctly optimized at the validation stage, which is important for the testing stage to make prediction for gait verification.

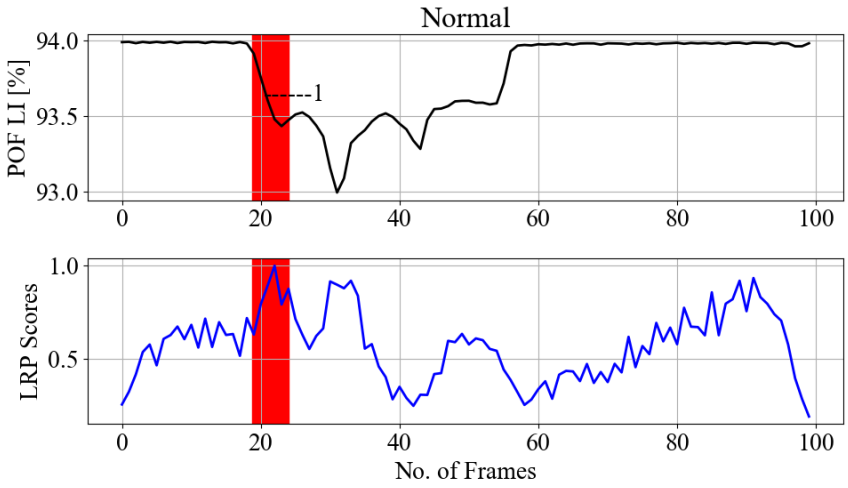
Experiment 2 is in essence an extra validation of the adequacy of the spatiotemporal sampling of GRF by the 116 sensors and their fusion, as well as that the classification performance of the trained models. An F1-score of 95% is achieved for test data from an unseen male as well as an unseen female. Although experiment 2 has the character of a sanity check, the results lend support to the value of floor sensor gait data as a biometric.

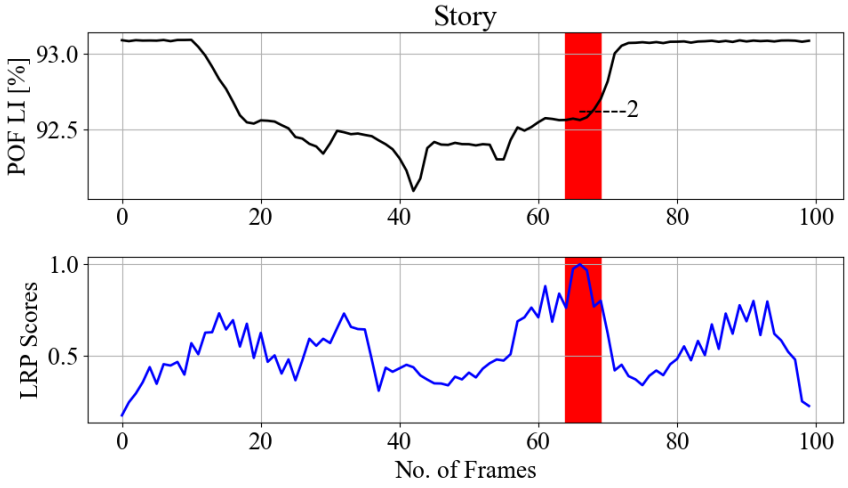
Experiment 3 is conducted to study the possibility to classify cognitive load on healthy subjects. It has shown that normal gait is classified with higher true positive rate compared to any of the classes of gait under cognitive load. This experiment also indicates that the achieved true positive rates in predicting normal gait are higher for the CNN model compared to the classical classifiers (see figure 6.4 and table 6.3). The learning curve in figure 6.5 indicates overfitting [201], to imply that the gait patterns under cognitive load diverge among the 21 subjects. Samples obtained under cognitive load samples are hard to fit due to the inconsistency of gait pattern changes among the subjects.

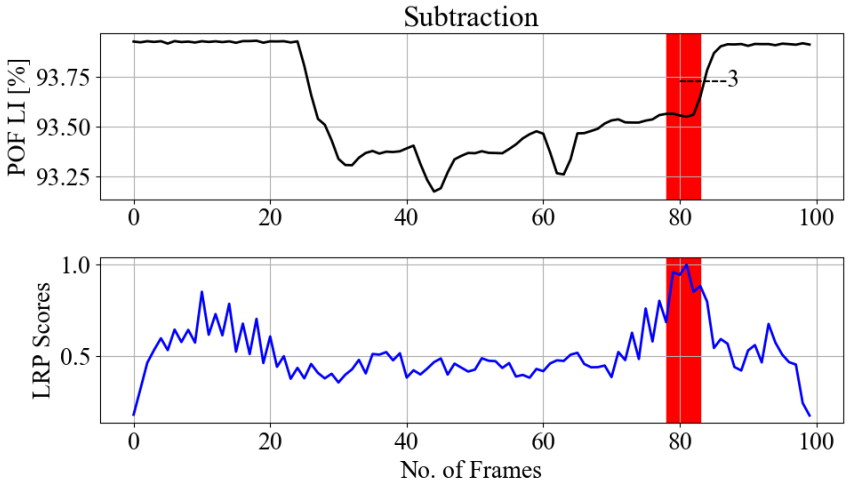
The results from the first three experiments suggest that while the dual-task data obviously contributes to the high F1-scores in experiments 1 and 2, it results in substantially degraded true positive rates in experiment 3. However, experiment 4 shows that when classifications are within a single subject the performance is notably better: for 16 subjects (out of 21) the gait under cognitive load the F1 score ranges between 80 to 100%, with the remaining 5 subjects the range being between 69% to 77%.

These observations can be discussed in the light of humans having a natural gait pattern evolved over millions of years; however, changes in gait when experiencing cognitive load at any particular instance are specific to the individual, expressing their response to the impaired ability to process cognitive information. In experiment 5, we use binary classifications (see table 6.2) to distinguish normal gait from gait under the 4 variants of cognitive load. The best classification results are obtained when the model learns normal or dual-task gait features for a single subject. This implies that although learned gait features under cognitive load may not be readily portable across subjects, they are

consistent for each individual and can contribute substantially for correct subject classifications; however, the accuracy drops if more subjects are involved.



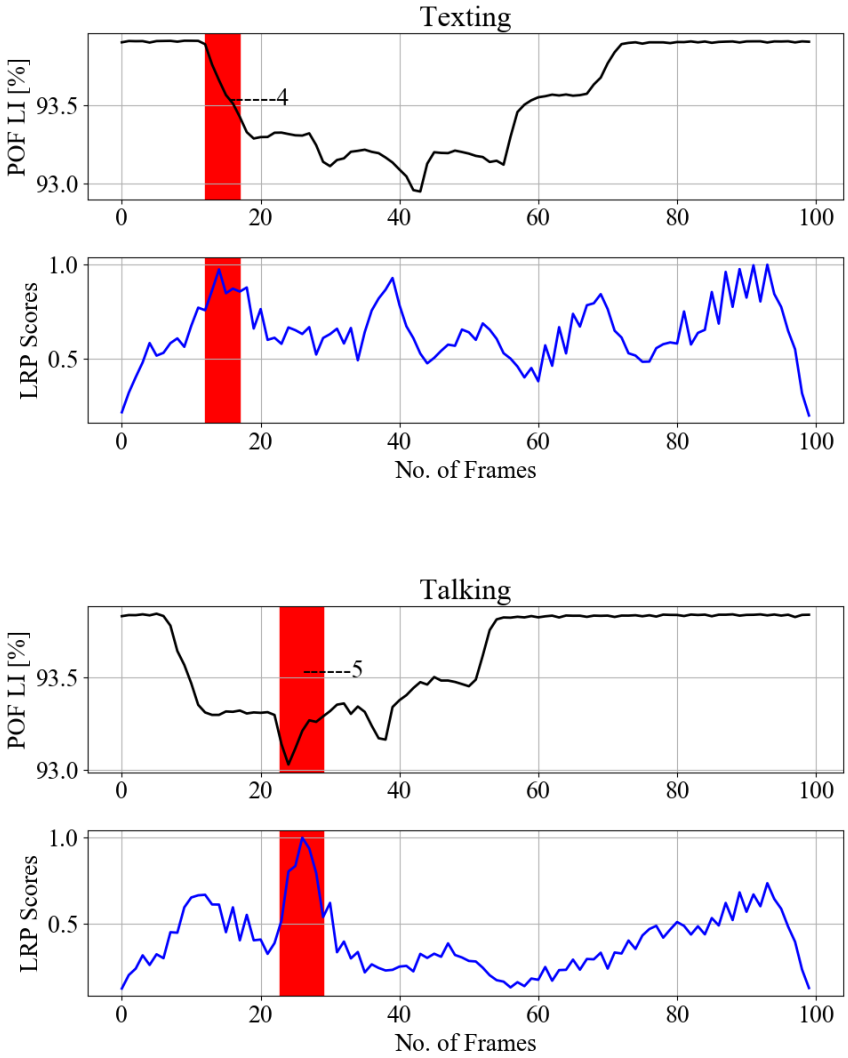



Figure 6.8. Consistent with identifying gait events relevant for the CNN prediction, random subject from experiment 1 gait events are: 1,4- Loading response or Foot flat and Single support, 2,3- Foot swing and opposite Heel strike, 5- Loading response or Foot flat.

## 6.5.2 Interpretation of Classifications

Figures 6.7, 6.8, 6.9 and 6.10 provide the link between the LRP relevance scores ("heat map") and the time sequence of the calculated *SA* signal in a single gait cycle window. The LRP score maxima are suitable pointers to the parts of the gait cycle which are most relevant for the classifications.

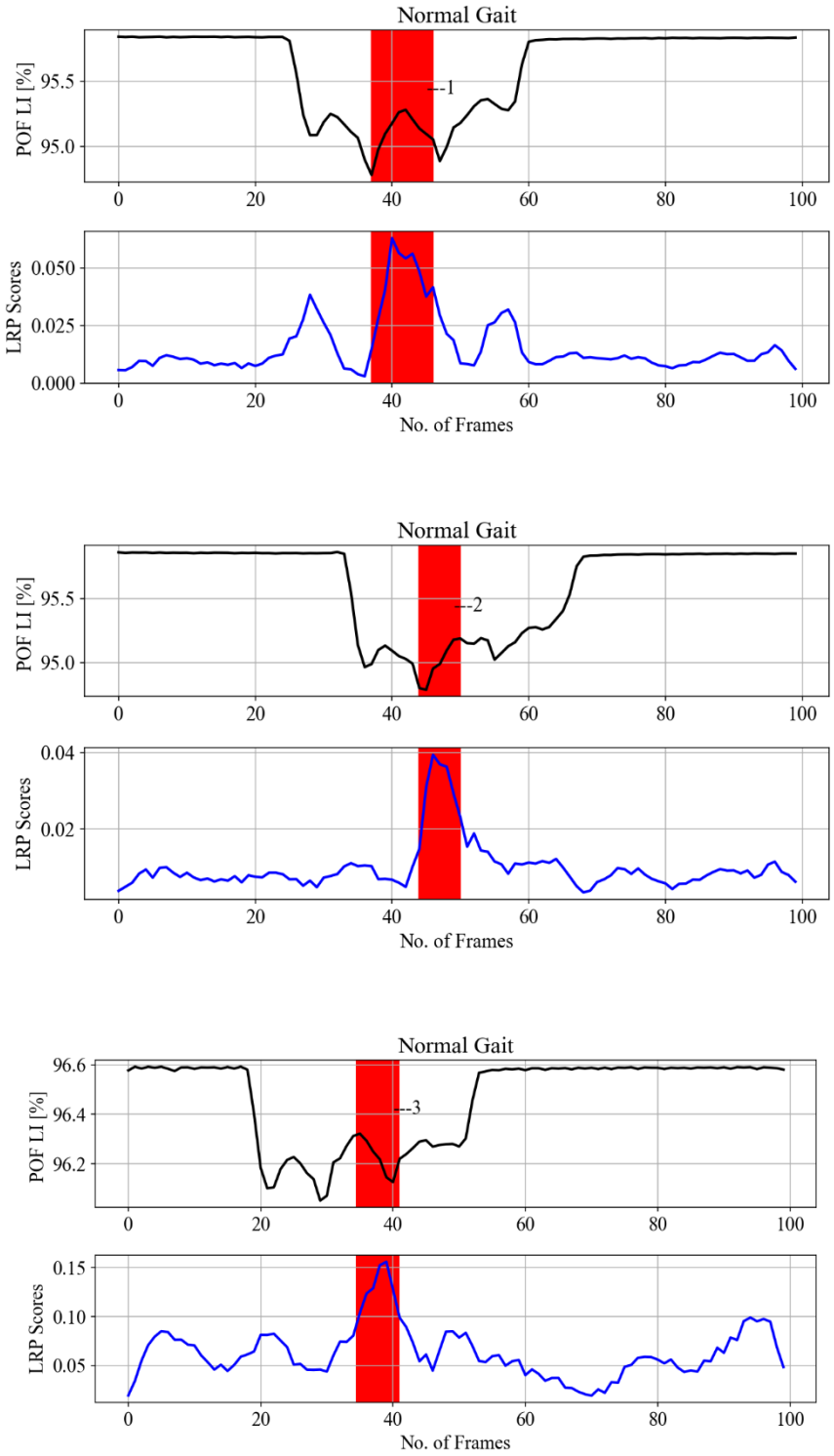



Figure 6.9. LRP methods applied on normal gait samples (from different subjects) from experiment 3 testing data, to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. Gait events are: 1,2,3- loading response or Foot flat and Double support.

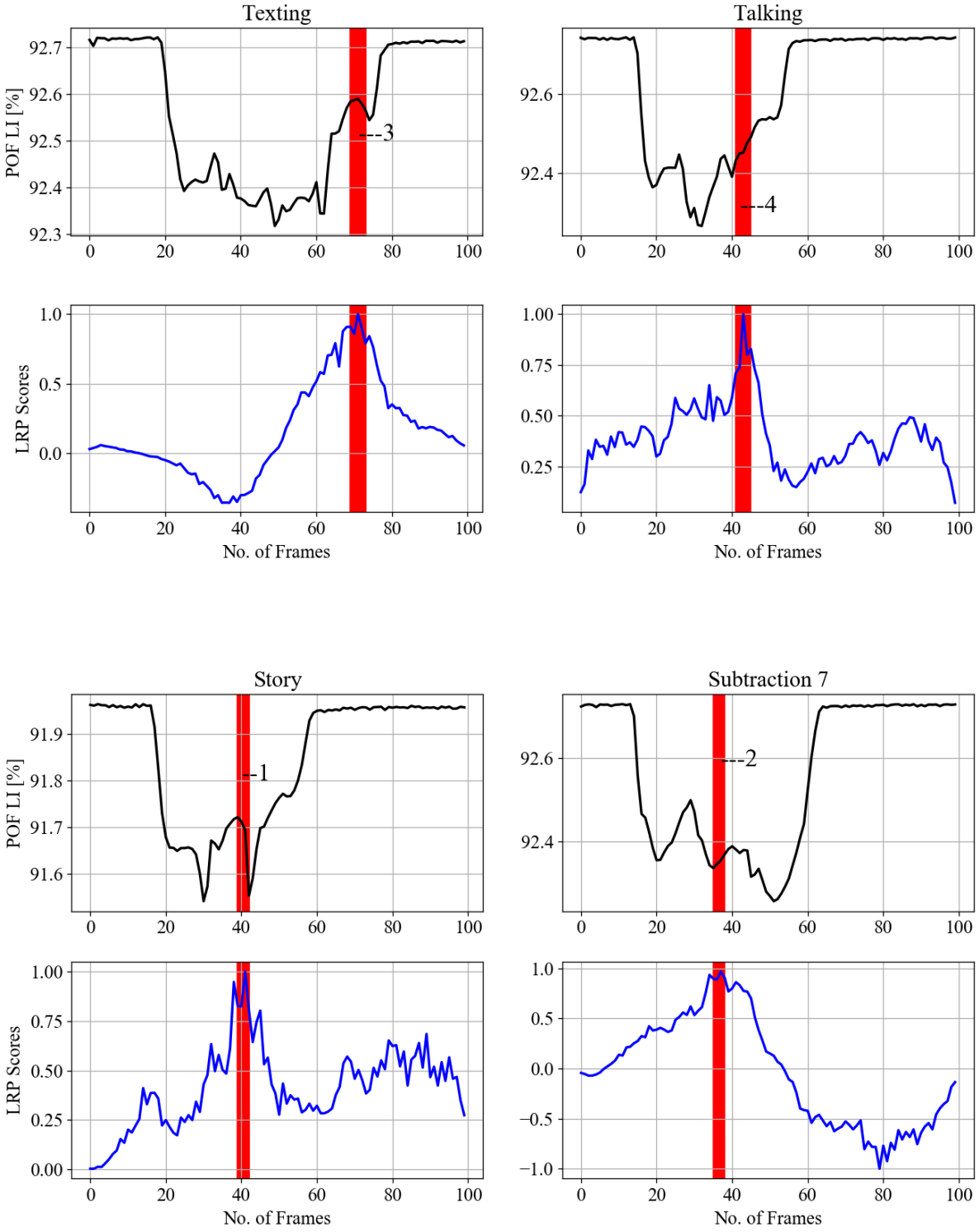



Figure 6.10. LRP methods applied on a single subject from experiment 5 testing data (each column is one pair), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. Gait events are: 1- Heel strike, 2- Toe-off, 3- between Foot swing and opposite Heel strike, 4- between Double support and Toe-off.

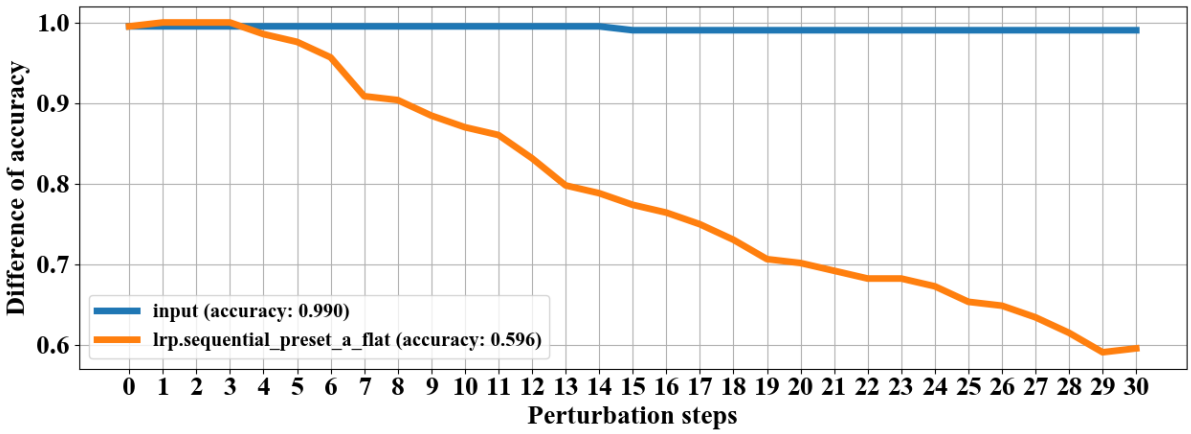
For accurate heat maps of a specific gait class the model's true positive prediction in the confusion matrix must be close to 100% for most of the testing samples, which points to the results from experiment 1 (figures 6.7 and 6.8), experiment 3 for normal gait heat maps - in figure 6.9 and experiment 5 for a single subject predicted gait under the 4 variants of cognitive load - in figure 6.10. Focusing just on one complete gait period (two steps) is justified by the fact that on multiple repetitive occasions each subject will initiate a gait cycle (see full description of the gait cycle events in chapter 3) by performing a heel strike, strictly followed by other gait events described in figure 6.6 and ending in a toe off.

Figure 6.7 indicates that loading response has high relevance to assigning a gait signature to one out of the 21 subjects gait samples, notably even under cognitive load, as indicated by with gait events numbered from 1 to 5. Figure 6.8 shows another subject randomly selected out of the 21. The indicated gait events are loading response for normal gait (1), gait while texting (4) and gait while talking (5); Foot swing and opposite Heel strike for gait while listening to a story (2) and while performing subtraction (3). The indication of events numbered 1, 2, 3 on figure 6.9 implies that normal gait identified by loading response or Foot flat and double support for 21 subjects. This gait event is marked by the model by 92% true positive (see figure 6.4) to distinguish normal gait from 4 cognitive load classes. Figure 6.10 shows cognitive load gait samples for one subject as per experiment 5 (the one subject never seen by the model) summarized as follows:

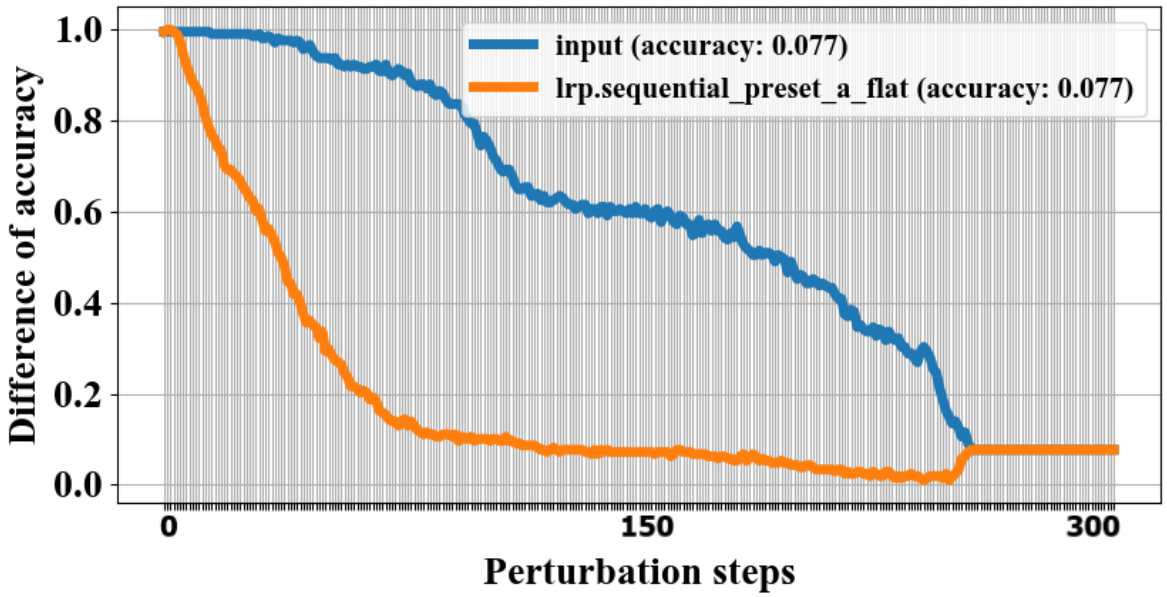
i. Gait while listening to story: Heel strike is significant for distinguishing listing to story from normal walking.
ii. Gait while performing serial 7 subtraction: Toe-off is significant for distinguishing 7 subtraction from normal walking.
iii. Gait while texting in smart phone: the transition from foot swing to opposite Heel strike is significant for distinguishing texting from normal walking.
iv. Gait while talking: the transition from double support to Toe-off is important to distinguishing talking from normal walking.

Overall, the LRP analysis indicates that subjects' normal gait is characterised by loading response; while the other cognitive load gait classes is classified by landing or lifting the feet on/from the surface of the iMAGiMAT system. For subject verification there are many

second relevant scores are used to predict the identity of the subject based on gait signature.




(a)




(b)

Figure 6. 11. LRP heatmaps validation by perturbation technique for experiment 1. Information with the highest relevance scores is progressively removed and the test samples are re-predicted. Steeper initial decrease indicates better identification of gait events with most weight in the classifications. a) shows the model predictions in 30 steps based on removing relevance scores using LRP Sequential Preset a Flat (LRP-SPF) and random removal of information. b) shows the model performance after 300 steps of information removal.

Figure 6.11 shows the assessment of the validity of the LRP heatmaps for subjects' identification using cognitive load. Here we apply the removal of region based on both LRP Sequential Preset a Flat (LRP-SPF) MoRF and random region removal and re-predicting gait class. As shown in figure 6.11 (a) the model prediction strongly decays using the LRP for the removal of information compared to the removal of random information. Figure 6.11 (b) shows the model performance over 300 steps. It can be seen that the model reaches lowest performance accuracy where the gait classes have to take a random prediction. Furthermore, it can be inferred from figure 6.11 that the model is effective in finding the most relevant region to identify subjects and the LRP is consistent over the test samples.

### 6.5.3 Benefits and Limitations

In view of gait recognition methods, floor sensors capture high resolution spatiotemporal data of gait, which is adequate for the monitoring of gait and walking habits in a range of applications. An unobtrusive floor sensor, such as iMAGiMAT, avoids the limitations caused by clothing conditions and viewing angles in video sequences [120],[121], the format of the captured signals in wearable sensors [124] and motion capture systems[202], acoustic noise background in footstep sound capturing systems [179], or the small area and fixed position of force plates [109]. Substantially, when processed by deep learning methods, floor sensors gait data, when captured under cognitive load, allows the accurate identification of subjects. This is suitable for laboratory as well as home/dwelling use, for healthcare or security, without compromising privacy. Scaling-up and large area deployment is naturally limited by cost; however, the technology demonstrated in iMAGiMAT allows the cost of materials for manufacturing the sensor head to be kept below 150 USD/sq.m [106].

## 6.6 Conclusion

We have shown that floor sensors and gait under cognitive load can be used for subject's identifications by capturing the changes in the individual's unique gait signature due to the need to process additional cognitive information in performing additional ("dual") tasks. We have shown that, despite higher identification of subjects, gait under cognitive

load is variable among healthy subjects, thus normal gait still consistence. In particular, LRP analysis of gait classifications indicate a possible bridge between the output of artificial intelligence systems for processing of high quality gait data and decisions based on visual observations by humans, or quantitative parameters derived from such observations, typical in current practice. An obvious potential impact is in the areas of biometrics and security [109], with the usual caveats pertaining to the regular update and fine-tuning of models, e.g. similar to other areas of biometric recognition. In healthcare alone, gait data from floor sensors can contribute and enhance the already significant interest in the detection of the onset of Parkinson disease [171] and fall risks [169]. It can be speculated that a possible future direction may involve the inobtrusive sampling of subjects' gait under routine conditions over intervals spanning periods of physical and mental changes due to ageing. Analysis of such longitudinal gait data, possibly corrected for mood, emotional state, and cognitive load personalized influences on gait, could contribute to earlier detection of disease onset.

# Chapter. 7 Gait Speed from Multimodal Data

# Fusion from Multimodal Gait Spatiotemporal Data for Human Gait Speed Classifications

**Abdullah S. Alharthi and Krikor B. Ozanyan**

## Abstract

Human gait pattens remain largely undefined when relying on a single sensing modality. We report a pilot implementation of sensor fusion to classify gait spatiotemporal signals, from a publicly available dataset of 50 participants, harvested from four different type of sensors. For fusion we propose a hybrid Convolutional Neural Network and Long Short-Term Memory (hybrid CNN+LSTM) and Multi-stream CNN. The classification results are compared to single modality data using Single-stream CNN, a state-of-the-art Vision Transformer, and statistical classifiers algorithms. The fusion models outperformed the single modality methods and classified gait speed of previously unseen 10 random subjects with 97% F1-score prediction accuracy of the four gait speed classes.

Article published in the proceedings of IEEE sensor conference 2021. Abdullah S. Alharthi and Krikor B. Ozanyan are with the department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PL, United Kingdom, (e-mail: abdullah.alharthi@manchester.ac.uk; k.ozanyan@manchester.ac.uk).

# 7.1 Introduction

Gait speed analysis, among various other fields, aims to fathom and reproduce in models the diversity of movements during daily-life activities. Potentials exist in the areas of human-vehicle interaction [203], e.g. to assess pedestrian movement in proximity to self-driving cars, or human-robot interaction, e.g. on a shared and busy factory floor, to secure safety based on the prediction of human speed and movement direction. In healthcare, gait speed is an important indicator of cognitive decline in individuals [204], and is employed in tools to monitor the progression of Alzheimer's disease [205].

The examples above highlight a strong dependence on the quality and quantity of data, since useful gait speed monitoring and analysis would not be limited to visual observation, or to a single sensing modality. Multi-modality fusion has become more accessible with the recent advances in AI, but this has also brought forward the issue of access to data. It is extremely challenging to single-handedly generate an internally compatible and large enough multi-modality database, thus methods suitable for (and tested on) public databases are an important and necessary starting point for developing classification tools.

Exploiting the spatiotemporal nature of human gait, the novelty in this work is the demonstration of multi-modal data fusion for gait speed classification, using a publicly available database and introducing the Hybrid Convolutional Neural Networks Long Short-Term Memory (hybrid CNN+LSTM) architecture, as well as the Multi-stream CNN architecture. Methodology

# 7.2 Data and Data-Processing

A "multimodal dataset of human gait at different walking speeds established on injury-free adult participants" was gathered by Schreiber et al in 2019 [206],[207]. The data was recorded from 50 volunteers, including 24 women and 26 men, 37.0 ± 13.6 years, 1.74 ± 0.09 m, 71.0 ± 12.3 kg. All participants were healthy and injury-free for lower and upper extremities in the six months leading to the data collection, with no lower and upper extremity surgery in the previous two years. The data is collected while participants walked on a straight level walkway at 5 speeds during a unique session: 0–0.4 m/s, 0.4–

0.8 m/s, 0.8–1.2 m/s, as well as at two self-selected speeds: spontaneous and fast [207]. The datasets are downloaded in the standard c3d file format [207] (https://www.c3d.org) and processed using "MOKKA-Motion Kinematic & Kinetic Analyzer" [208] to their original form of raw spatiotemporal data. The latter, in three-dimensional (3D) cartesian coordinate space, consists of the trajectories of 52 reflective markers (see figure 7.1); force-plates measurements of 3D ground reaction forces, 3D ground reaction moment, centre of pressure coordinates, direction of angle coordinates; electromyographic signals collected by 8 probes, as shown in figure 7.1 (E1-E8). The final datasets for 50 participants are listed by data type in table 7.1, a total of 1142 trials without the one static as in [207]. Samples were padded with zeros as necessary, to correct for unequal size due to varying gait speed. The resulting size for each data type is shown in table 7.1, as it is a parameter of importance for the training of machine learning models.

Table 7.1. Size of Sample Frames and Dimensionality by Data Type.

| Data type | Spatial | Temporal | Dimensionality |
|---|---|---|---|
| 3D TRM | 156 | 585 | 3 |
| 3D GRF | 24 | 585 | 3 |
| Forces and Moments | 12 | 8985 | 3 |
| EMG | 8 | 8985 | 1 |

TRM: Trajectories of Reflective Markers; GRF: Ground Reaction Force EMG: Electromyography.

## 7.3 Machine Learning Methods

In this work, gait analysis is handled as a supervised classification. The models are trained and validated on 40 subjects, then tested on the remaining randomly selected 10 subjects, to predict the walking speed. To provide the required computational resources, we use the Tensor Processing Unit (TPU) (machine learning accelerator application-specific integrated circuit (ASIC) developed by Google specifically for neural network machines) in Colaboratory - Google Research. To compare the performance of several alternative

approaches for this task, we have considered the following advanced machine learning models:

### 7.3.1 Vision Transformer

The state-of-the-art Vision Transformer (ViT) was developed by Google Research, Brain Team [209] for image classification based on natural language processing (NLP) Transformer [210].

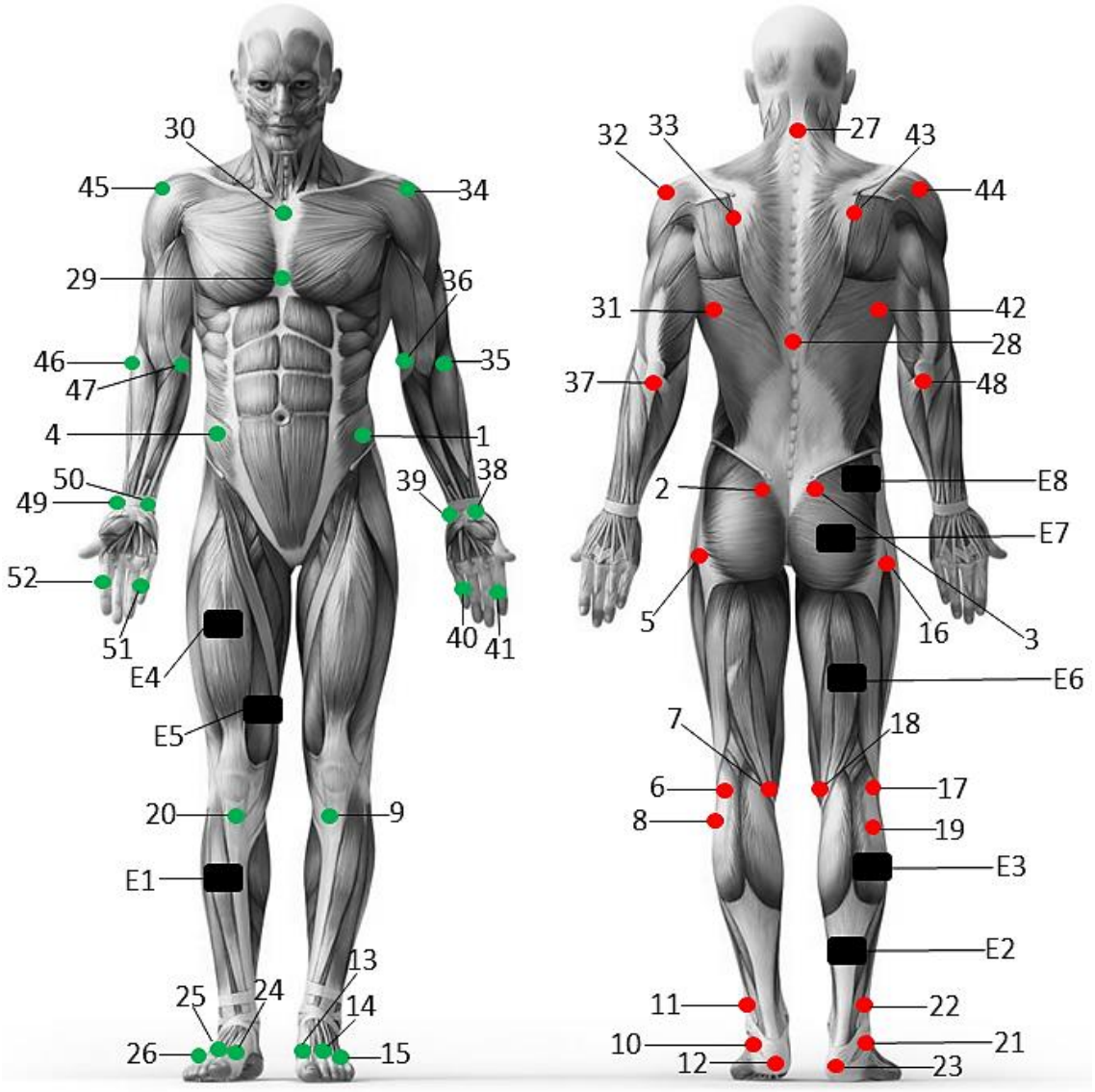



Figure 7.1. 52 reflective cutaneous markers, placed by anatomical palpation on the participants. Markers (front side – green, back side – red) are numbered by their order in the spatial domain. Electromyography signals recorded from probes of the right leg muscles are illustrated in black: E1-E8.

In our model design, we follow the original Transformer in [209], although we experimented with the model parameters to achieve the best classification result namely: learning rate, weight decay, batch size, number of epochs, input size, patch size, number batches, projection dimensions, number of heads, transformer layers, and multilayer perceptron (MLP) head units.

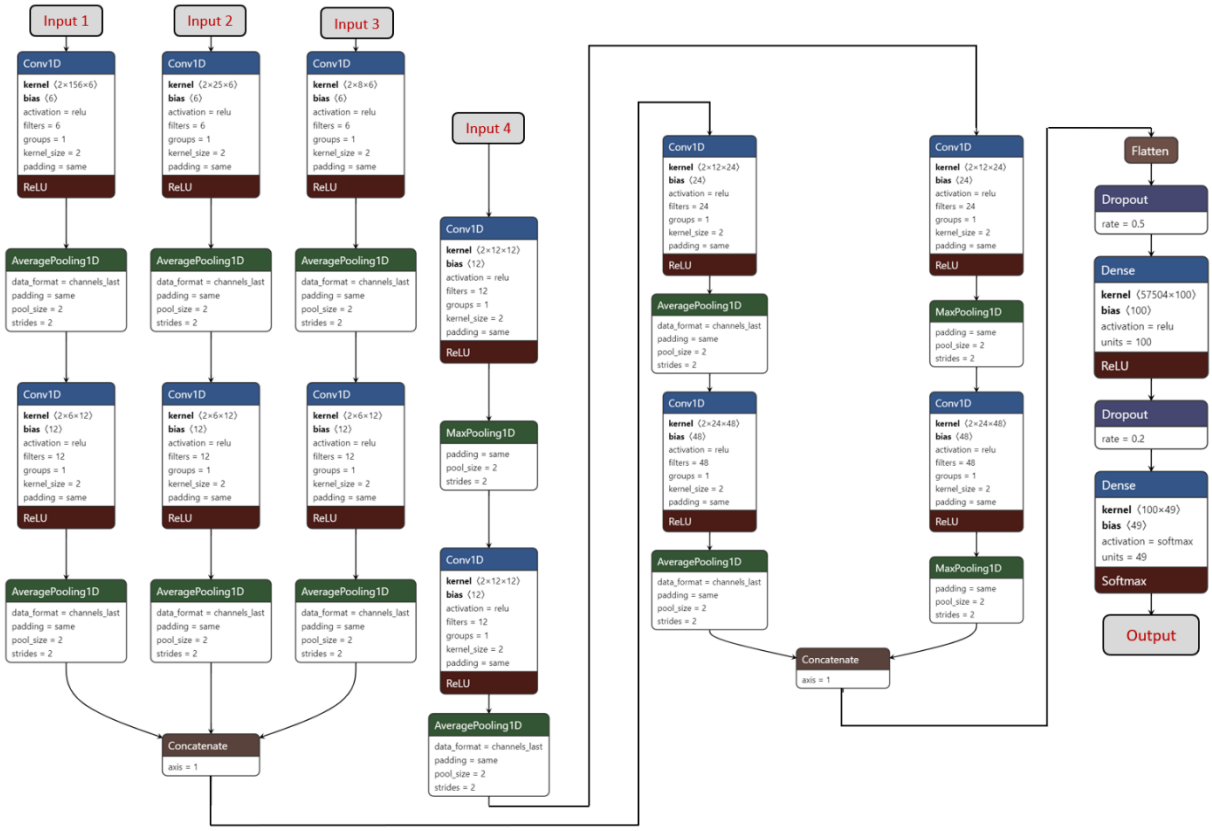



Figure 7.2. CNN+CNN network architecture, Input 1: TRM, input 2: GRF, input 3: EMG, input 4: Forces and Moments.

## 7.3.2 Multi-stream CNN and Single-stream CNN

Multi-stream CNN and Single-stream CNN networks are proposed here, based on extensive previous work (see chapter 4 to 6). The Multi-stream CNN (see figure 7.2) is designed to fuse the 4-data in the deep layers, based on weights sharing during the training and validation stage and tested on randomly selected 10 subjects. For performance comparison we propose Single-stream CNN network, (see chapter 5). This network can have only one type of data as an input.

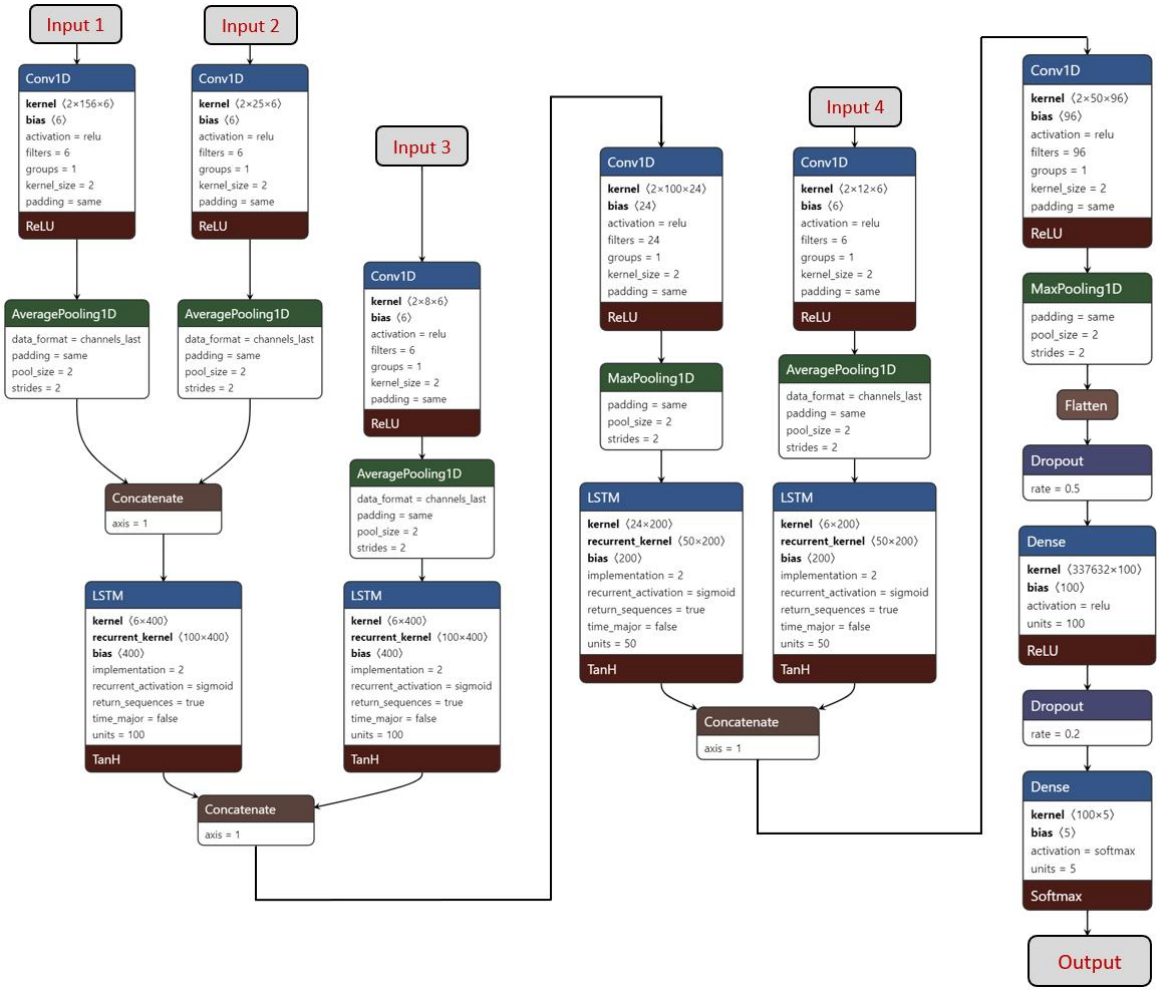



Figure 7.3. Hybrid CNN+LSTM network architecture. Input 1: TRM, input 2: GRF, input 3: EMG, input 4: Forces and Moments.

### 7.3.3 Hybrid CNN+LSTM network

The human gait data are harvested as a naturally spatiotemporal pattern. Therefore, we propose the CNN+LSTM network, such that convolution layers extract patterns followed by LSTM cells to decide which features are to be kept and passed on in the networks layer, for the 4-data type fusion. The network stacked layers are shown in figure 7.3, whereas the network is topped-up by MLP with *SoftMax,* for classification.

### 7.3.4 Statistical Classifiers Algorithms

The classical machine learning algorithms, namely Decision Tree, Support Vector Machine, and k-Nearest neighbours, are proposed as a control and reference to the classification task.

The input to these algorithms is only one data type (see table 7.1 for data type) in each classification task, hence there is no fusion. The input to these algorithms is samples flattened to form a 1D vector.

Table 7.2. F1 scores to predict gait speed for 10 random subjects.

| Model type | Single Modality F1 % | Fusion F1% |
|---|---|---|
| Transformer | 58% ± 0.3% | - |
| Decision Tree | 82% ± 0.8% | - |
| Support Vector Machine | 84% ± 1.1% | - |
| k-Nearest neighbors | 83% ± 0.7% | - |
| Single-Stream CNN | 91% ± 0.6% | - |
| Multi-Stream CNN | - | **97%** |
| Hybrid CNN+LSTM | - | **96%** |

The single modality classification results in ± F1% for each data type.

## 7.4 Results and Discussion

The proposed models are trained and tested on the data from 50 participants (40 for training and 10 testing) to predict gait speed. The reported result are the predictions of the gait speed of 10 randomly selected subjects, never seen during the training stage (consistent with a real-world scenario). The proposed fusion methods outperformed the single modality methods, namely: ViT, Single-Stream CNN, Decision Tree, Support Vector Machine and k-Nearest neighbours, as shown in table 7.2. Figure 7.4 shows a prediction confusion between gait speed classes C1 and C2 (0.8–1.2 m/s and self-selected spontaneous), which is understandable because the subjects' self-selected spontaneous speed is around 0.8–1.2 m/s. This is confirmed when combining these two classes into one class (C3) as shown by the confusion matrix in figure 7.5 (a) and Receiver Operating Characteristic (ROC) in figure 7.5 (b). The Multi-stream CNN and the Hybrid CNN+LSTM returned similar results. However, the training time for the former was less than 5 minutes, while the latter required a much longer time, about 4 hours, to train.

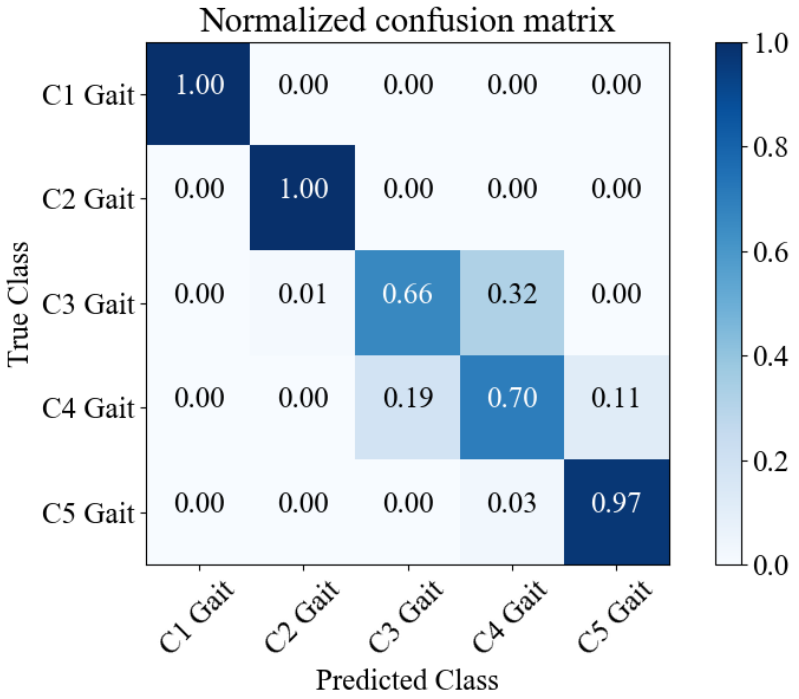
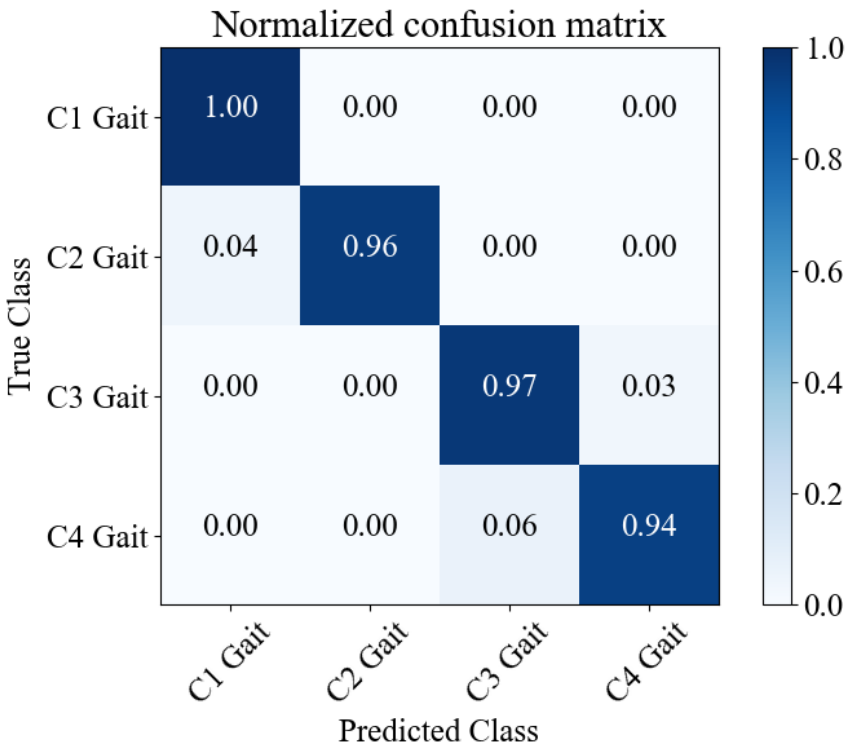


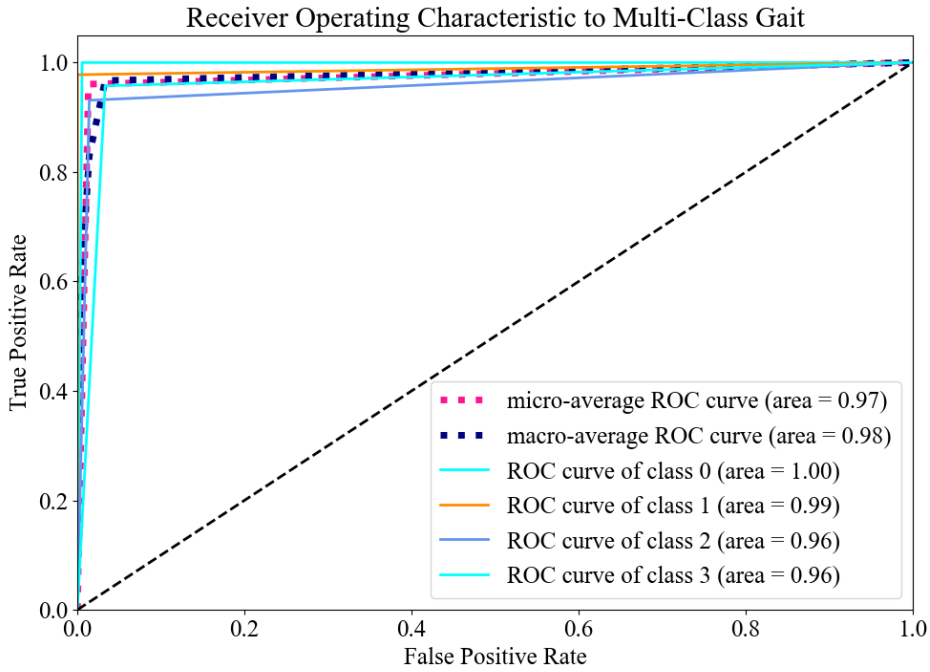

Figure 7.4. Gait speed classification in 5 classes: confusion matrix using hybrid CNN+LSTM.

The Vision Transformer performed poorly on this classification task, which can be attributed to the transformers model being designed with larger data capacity in mind. It is known to perform competitively when pre-trained on large-scale, weakly labelled and fine-tuned on a smaller data [211]. Compared with CNN (on image classification task [209]), the transformers tend to perform well on a large data, whereas the CNN perform well on smaller data. This can be advantageous, such that in their generalization the transformers can be picked up for problems on larger data, as per its success in natural language processing [210],[212]. In such a case a larger spatiotemporal gait data will be uses to train the transformer and a smaller task specific data will be used to fine-tuned the transformer for an application, such as gait speed recognition.

In our implementation, 10 subjects never seen by the model at the training stage are used to evaluate the model's accuracy. As noted in chapter 3, most of the gait classification literature is based on using subjects' data in both training and testing stage. This might yield misleading results, so that when tested in real world scenario the predictions become inaccurate and possibly problematic, especially in pathological gait classification and mental illness.




(a)




(b)

Figure 7.5. Gait speed classification in 4 classes. (a) confusion matrix and (b) ROC using Multi-Stream CNN.

## 7.5 Summary and Conclusions

In this paper, we study the fusion of reflective markers' trajectories, force-plates, and electromyographic sensors. Our data processing experiments focus on predicting the gait speed of 10 unseen subjects, consistent with a real-world scenario. We compare the performance of a number of supervised algorithms, namely: hybrid CNN+LSTM and Multi-stream CNN, single-stream CNN, Vision Transformer, and statistical classifiers.

Multimodality fusion achieved higher prediction accuracy, compared with the single modality methods, such as single-stream CNN, Vision Transformer, and statistical classifiers. It can be concluded that within our choice of alternatives, the deep learning models outperformed the classical classifies due to their ability to learn gait patterns throughout the training.

This work is limited to multimodal data fusion and classification for gait speed prediction. However, the methods investigated here are amenable to broadening their remit and provide a starting point for a broad set of applications, such as car navigation in scenes involving pedestrians, human-robot interaction, clinical practice to diagnose pathology or monitor cognitive impairment and stroke in elderly patients.

# Chapter. 8 Summary and Contributions, Future Directions, and Conclusions

# 8.1 Summary and Contributions

The literature search suggested that the research should focus on applying deep learning to analyse gait signals harvested by VS, WS and FS as presented in Chapter 3. The aim of this research was to extract gait features automatically and classify gait for security and healthcare applications using the state-of-the-art artificial neural network's methods for improved accuracy. Furthermore, the methodologies proposed in the reviewed literature focused on the challenges imposed by the environmental disturbance to capture gait signals, e.g., clothing's conditions, angles in video sequences, the limitation in the number of sensors and the format of the captured signals in wearable sensors, or the presence of other environmental noise.

To overcome these challenges, only recently - as highlighted in Chapter 3 (section 3.5) - the gap in literature indicates that more research is needed for gait captured by sensors under the foot and sensors fusion using deep learning. Therefore, this thesis focusses on the analysis of gait deterioration due to cognitive load and cognitive decline. A comprehensive analysis of gait spatiotemporal signal using deep CNN methodologies are presented in two cases: PD patients (chapter 4) and healthy humans (chapter 5 and 6). The proposed methodologies are implemented to accomplish the following.

1. Fusion of the sensors' raw spatiotemporal signals while down sampling to map the sensor signal to a class represented by a single digit number or a case of cognitive load or cognitive decline.
2. Gait pattern recognition based on classifying the down sampled gait data by achieving state of the art model performance.
3. Identifying subjects, using their individual cost of cognitive processing of information while walking with 100% F1 scores prediction from floor sensors spatiotemporal data.
4. Staging PD severity using gait data, with ground truth marked by clinicians, achieving 98% F1 scores prediction for sensors under the foot.
5. Models which outperform previous manual feature extraction methods and exhibit robustness to perturbation noise, overcoming the variability of human gait from person to person in PD analysis.

6. Multimodality sensor fusion with CNN and CNN+LSTM (long short term memory) networks for classifying gait speed, as presented in Chapter 7. The results are compared to statistical analysis and transformers, the proposed models manifesting superior results, with 97% F1 scores prediction of 10 subjects' gait never seen by the model.

The other main contribution of this thesis is that there is no work in literature on why these classification decisions based on gait are made by the deep learning models. It has been noted that gait as a characteristic exists in all humans, and it has identical repeating pattern defined by the gait cycle events, which is established based on visual gait analysis. Therefore, in this thesis we link the deep learning models outputs to key known events in the gait cycle using LRP and perturbation, as an attempt to understand and possibly build bridges between the output of artificial intelligence systems for processing of high quality gait data and decisions based on visual observations by humans, or quantitative parameters derived from such observations, typical in current practice. Few gait events have been identified to deteriorate due to cognitive load or cognitive decline, e.g., the heel strike or toe-off. In other words, single support is mostly affected in both cognitive load cases presented in this thesis. The LRP heat maps have shown that gait deterioration increases together with the task complexity as the body balance and posture deteriorate due to cognitive load or PD effect on heel strike or toe-off.

## 8.2 Future direction

Gait analysis based on cognition is a very young field and the growing advancement of sensors and computing technologies will significantly impact sensors fusion and gait deterioration research. The presented research in this thesis has set a solid base for further research from different points of view and motivation. A natural follow-ups to the work presented in this thesis are listed in the following points.

1. The recording of gait data from a large healthy cohort using floor sensors and sensor fusion under routine conditions over intervals spanning periods of physical and mental changes due to ageing. Analysis of such longitudinal gait data, possibly corrected for mood, emotional state, and cognitive load personalized influences on gait, could contribute to earlier detection of disease onset. Ideally, floor sensors

are suitable for this task, thus deploying other wearable sensors, and motion capturing systems would provide more insight on the movements of the lower limbs, pelvis, chest, and arms swing.

2. Sensor fusion is required in many applications including gait analysis. Therefore, more advanced methods will be recommended for multimodality sensor fusion. Among such methods are our proposed models, e.g. parallel DCNN, quadruplet DCNN, Multi-stream CNN, LSTMs, Hybrid CNN+LSTM. A future implementation of transformers networks for large data analysis, as well can be generalized on similar sensors with similar data characteristics of gait.
3. A possible future direction is to expand the knowledge on the running gait with new hypotheses of the athlete's performance associated with musculoskeletal evaluation and the risk factors of injury in sport. The aim is to use the machine learning outputs using LRP for the link between the running gait and gait cycle events. Floor sensors, motion capturing system, IMUs, EMG, etc. can be fused to gain more insight on the performance of athletes and their body balance and muscle while running.
4. The deployment of gait analysis in robotic applications for human-robot interaction, e.g. on a shared and busy factory floor or navigating in a busy public space, to secure safety based on the prediction of human speed and movement direction.

The presented gait analysis methods for spatiotemporal signals can be generalized and optimized on shared-control-based bimanual robot manipulation [213]. With the capacity to fuse sensor spatiotemporal signals, e.g. wearable sensors and motion capturing systems attached to the arms or force sensing sensors with spatiotemporal characteristics. To control robot's arms movement to mimic the operator's arms movements, to solve the challenges of latency in signals delivered from clouds based on remote teleoperation robotic arms and provide assistance to help the user complete tasks more easily. Similar to gait, a number of subjects arms movements will be recorded based on how people perform two-hand manipulations for a certain task. The data will be processed using CNNs and the arm movement prediction will fill in the latency gap or can be used to enforce more accuracy in the robotic arm movement.

## 8.3 Conclusion

The literature revealed that sensors under the foot, sensors fusion, and deep learning models have shown promising results, which make the contributions of this thesis particularly timely. This thesis emphasises the impact of sensors fusion and cognition for better classification performance, by not only assessing accuracies in the best ways, but also including other factors, such as: (i) the impact of cognition on the event of gait cycle that need more attention for sensor design and gait recording protocols. (ii) aiding clinical practitioners with the identification of the most relevant gait cycle events in diagnosing and monitoring the progression of gait abnormality, e.g. cognitive impairment, prosthetic gait, advanced age and aging, stroke, Alzheimer's disease, dementia. (iii) floor sensors and gait under cognitive load can be used for subject's identifications by capturing the changes in the individual's unique gait signature due to the need to process additional cognitive information in performing additional ("dual") tasks. The contributions of other's work are acknowledged and presented in this thesis as a moving step forward in the innovation of gait analysis in such a significant way.

# References


[1] Bhagwati Charan Patel, G R Sinha and Naveen Goel, “Advances in Modern Sensors” Physics, design, simulation and applications, (IOP Publishing Ltd), Online ISBN: 978-0-7503-2707-7 • Print ISBN: 978-0-7503-2705-3, 2020. https://iopscience.iop.org/book/978-0-7503-2707-7.

[2] J. Brignell and N. White, “Intelligent Sensor Systems 2nd” (Sensors), Institute of Physics Publishing (Bristol: IOP Publishing), 1996. http://eprints.soton.ac.uk/id/eprint/251222

[3] R. Caldas, T. Fadel, F. Buarque and B. Markert, “Adaptive predictive systems applied to gait analysis: A systematic review,” *Gait & Posture*, vol. 77, pp. 75-82, ISSN 0966-6362, 2020, doi:10.1016/j.gaitpost.2020.01.021.

[4] M. J. Major, P. Raghavan and S. Gard, "Assessing a low-cost accelerometer-based technique to estimate spatial gait parameters of lower-limb prosthesis users," *Pros. Ortho. Intern.*, vol. 40, no. 5, pp. 643–648, 2015, doi: 10.1177/0309364614568411.

[5] R. Camicioli, D. Howieson, S. Lehman and J. Kaye, ”Talking while walking: The effect of a dual task in aging and Alzheimer's disease,” *Neurology*, vol. 48, no. 4, pp. 955-958, 1997, doi: 10.1212/WNL.48.4.955.

[6] D. Gouwanda and S. M. N. A. Senanayake, “Emerging Trends of Body-Mounted Sensors in Sports and Human Gait Analysis,” *4th Kuala Lumpur International Conference on Biomedical Engineering 2008, IFMBE Proc.,* vol. 2, pp. 715–718, Springer 2008, doi: 10.1007/978-3-540-69139-6_178

[7] H. Lee, S. J. Sullivan and A. G. Schneiders, “The use of the dual-task paradigm in detecting gait performance deficits following a sports-related concussion: A systematic review and meta-analysis,” *J. Sci. Med. Sport,* vol. 16, no. 1, pp. 2-7, 2013, doi: 10.1016/j.jsams.2012.03.013.

[8] M. Alotaibi and A. Mahmood, “Improved Gait Recognition Based on Specialized Deep Convolutional Neural Network,” in *Comp. Vis. Imag. Understanding,* vol. 164, pp. 103–110, 2017, doi: 10.1016/ j.cviu.2017.10.004.

[9] G. Qian, J. Zahang and A. Kidane, “People Identification Using Floor Pressure Sensing and Analysis,” *IEEE Sensors J.,* vol. 10, pp. 1447–60, 2010, doi: 10.1109/JSEN.2010.2045158.

[10] S. Li, B. Schouten, and M. Tistarelli, “Biometrics at a distance: Issues, challenges, and prospects,” in Handbook of Remote Biometrics (Advances in Pattern Recognition), M. Tistarelli, S. Li, and R. Chellappa, Eds. London, U.K.: Springer, 2009, pp. 3–21. doi: 10.1007/978-1-84882-385-3_1

[11] N. Céspedes, D. Raigoso, M. Múnera and C. A. Cifuentes , “Long-Term Social Human-Robot Interaction for Neurorehabilitation: Robots as a Tool to Support Gait Therapy in the Pandemic,” *Frontiers in Neurorobotics*, 23 February 2021, doi: 10.3389/fnbot.2021.612034.

[12] N. Céspedes, M. Múnera, C. Gómez, and C. A. Cifuentes, “Social Human-Robot Interaction for Gait Rehabilitation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering.* vol. 28, np. 6, pp. 1299 – 1307, June 2020. doi: 10.1109/TNSRE.2020.2987428.

[13] K. H. Yu, A. L. Beam and I. S. Kohane, “Artificial intelligence in healthcare,” Nature Biomedical Engineering. vol. 2, pp. 719–731, 2018. doi:10.1038/s41551-018-0305-z.

[14] O. Simeone, "A Very Brief Introduction to Machine Learning With Applications to Communication Systems," *IEEE Trans. Cogn. Commun. Netw*. (IEEE Xplore early access), 2018, doi: 10.1109/TCCN.2018.2881442.

[15] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning,”. *Nature*, vol. 521, pp. 436–444, 2015, doi: 10.1038/nature14539.

[16] I. Goodfellow, Y. Bengio and A. Courville, “Deep learning,” in *Deep Learning*, *MIT Press*, ch. 2, pp. 326-415, 2016, ISBN: 978-0262035613.

[17] N. Yi, C. Li, X. Feng and M. Shi, “Research and Improvement of Convolutional Neural Network,” *17th IEEE/ACIS Int. Conf. Comput. Inf. Sci. (ICIS 2018)*, 2018, pp. 637-640. doi: 10.1109/ICIS.2018.8466474.

[18] S. Ruskaa, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, J. Mononen, “Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle,” *Behav. Processes*, vol. 148, pp. 56-62, March 2018, doi: 10.1016/j.beproc.2018.01.004.

[19] Y. Ho and S. Wookey, “The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling” *IEEE Access*, vol.8, pp. 4806 - 4813, December 2019. doi: 10.1109/ACCESS.2019.2962617.

[20] G. Montavon, W. Samek, K. R. Müller, “Methods for interpreting and understanding deep neural networks,” *Digital Signal Processing*, vo. l73, pp. 1-15, ISSN. 1051-2004, 2018. doi: 10.1016/j.dsp.2017.10.011.

[21] G. Montavon, S. Lapuschkin, A. Binder, W. Samek and K. R. Müller, “Explaining nonlinear classification decisions with deep Taylor decomposition,” *Pattern Recognition*. vol. 65, pp. 211-222, ISSN 0031-3203, 2017. doi: 10.1016/j.patcog.2016.11.008.

[22] W. Samek, A. Binder, G. Montavon, S. Lapuschkin and K. Müller, "Evaluating the Visualization of What a Deep Neural Network Has Learned," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 11, pp. 2660-2673, 2017. doi: 10.1109/TNNLS.2016.2599820.

[23] Z. Zhang and M. R. Sabuncu, "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels," *2nd Conference on Neural Information Processing Systems* (NeurIPS 2018). arXiv:1805.07836.

[24] W Samek, G Montavon, S Lapuschkin, C Anders, KR Müller, “Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications,” Proceedings of the IEEE, vol. 109, no. 3, March, pp. 247-278, 2021. doi: 10.1109/JPROC.2021.3060483.

[25] Aristotle, "On the Gait of Animals," 350BC. [Online]. Available: http://classics.mit.edu/Aristotle/motion_animals.html [Accessed 08 Oct. 2018].

[26] M. H. Pope, "Giovanni Alfonso Borelli—the father of biomechanics," *Spine*, vol. 30, no. 20, pp. 2350-2355, 2005, doi:10.1097/01.brs.0000182314.49515.d8.

[27] M. W. Whittle, "Whittle's Gait Analysis," p. 30. *5th ed.,* Churchill- Livingstone, Elsevier, 2012, ISBN 978-0-7020-4265-2.

[28] W. Weber and E Weber, "Über die Mechanik der menschlichen Gehwerkzeuge, nebst der Beschreibung eines Versuchs über das Herausfallen des Schenkelkopfs aus der Pfanne im luftverdünnten Raume," *Annal. Phys Chem.* vol. 116, no. 1, pp. 1-13, 1837, doi: 10.1002/andp.18371160102; Also see: *Mechanics of the Human Walking Apparatus (Engl. Transl: P. Maquet and R Furlong).* Springer, 1992, ISBN 3-540-53541-1.

[29] P. Connor and A. Ross, "Biometric recognition by gait: A survey of modalities and features," *Comp. Vision Imag. Unders.*, vol. 167, pp. 1-27, 2018, doi: 10.1016/j.cviu.2018.01.007.

[30] V. L. S. Profeta and M. T. Turvey, "Bernstein's Levels of Movement Construction: A Contemporary Perspective," *Hum. Mov. Sci.*, vol. 57, pp. 111–133, 2017, doi:10.1016/j.humov.2017.11.013.

[31] P. R. Cavanagh and M. A. Lafortune, "Ground reaction forces in distance running," *J. Biomech.*, vol. 13, no.5, pp. 397-406, 1980, doi: 10.1016/0021-9290(80)90033-0.

[32] H. Elftman and J. T. Manter, " The axis of the human foot," *Science,* vol. 80, no. 2082. pp. 484, 1934, doi: 10.1126/science.80.2082.484.

[33] D. H. Sutherland, "The evolution of clinical gait analysis part III—kinetics and energy assessment," *Gait & Post.*, no. 4 , pp. 447–461, 2005, doi: 10.1016/j.gaitpost.2004.07.008.

[34] A. Muro-de-la-Herran, B. G. Zapirain and A. M. Zorrilla, "Gait Analysis Methods: An Overview of Wearable and Non-Wearable Systems, Highlighting Clinical

Applications," *Sensors, Basel, Switzerland*, vol. 14, np. 2, *pp*. 3362–3394, 2014, doi:10.3390/s140203362.

[35] E. Abdulhay, N. Arunkumar, K. Narasimhan and E. Vellaiappan, "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease," *Future Gener. Comput. Syst.*, vol. 83, pp. 366–373, 2018. doi: 10.1016/j.future.2018.02.009.

[36] G. Bastas, J. J. Fleck, R. A. Peters and K.E. Zelik," IMU-based gait analysis in lower limb prosthesis users: Comparison of step demarcation algorithms," *Gait & Post.*, vol. 64, pp. 30–37, 2018, doi: 10.1016/j.gaitpost.2018.05.025.

[37] C. Yan, B. Zhang and F. Coenen, "Multi-attributes gait identification by convolutional neural networks," in *Proc. Int. Cong. on Img. Sig. Proc. (CISP)*, Shenyang, pp. 642-647, 2016, doi: 10.1109/CISP.2015.7407957.

[38] F. Battistone and A. Petrosino, "TGLSTM : A time based graph deep learning approach to gait recognition," *Pattern. Recog. Lett.*, 2018. doi: 10.1016/j.patrec.2018.05.004.

[39] J. Suutala, S. Pirttikangas, J. Riekki and J. Ronin, "Reject-Optional LVQ-Based Two-Level Classifier to Improve Reliability in Footstep Identification," in *Proc. 2nd Int. Conf. Pervasive. Comp.*, *(Pervasive 2004)*, pp. 818-821, 2004, doi: 10.1007/978-3-540-24646-6_12.

[40] A. Mostayed, S. Kim, M. M. G. Mazumder and S. J. Park," Foot step based person identification using histogram similarity and wavelet decomposition," *Proc. 2nd Int. Conf. Inf. Sec. and Assur. (ISA 2008)*, pp. 307–311, 2008, doi: 10.1109/ISA.2008.46.

[41] S. C. White and D. A. Winter, "Predicting muscle forces in gait from EMG signals and musculotendon kinematics," *J. Electromyogr. Kinesiol*., vol. 2, no. 4, pp. 217–231, 1992, doi: 10.1016/1050-6411(92)90025-E.

[42] S. L. Di. Stasi, D. Logerstedt, E. S. Gardinier and L. S. Mackler, "Gait patterns differ between ACL-reconstructed athletes who pass return-to-sport criteria and those

who fail," *Am. J. Sports Med.,* vol. 41, no. 6, pp. 1310-1318, 2013, doi: 10.1177/0363546513482718.

[43] E. R. H. P. Isaac, S. Elias, S. Rajagopalan and K.S. Easwarakumar, "Multiview Gait-Based Gender Classification through Pose-Based Voting," *Pattern Recognit. Lett.*, 2018, doi: 10.1016/j.patrec.2018.04.020.

[44] I. Birch, T. Birch and D. Bray, "The identification of emotions from gait," *Sci. & Jus.*, vol. 56, no. 5, pp.351-356, 2016, doi: 10.1016/j.scijus.2016.05.006.

[45] J. Preis, M. Kessel, M. Werner and C. Linnhoff-Popien, "Gait recognition with Kinect," *Proc. 1st Workshop Kinect Pervasive Comput., Pervasive 2012*, https://www.martinwerner.de/pdf/GAITKINECT.pdf .

[46] B. Galai and C. Benedek, "Feature selection for lidar-based gait recognition," *Int. Workshop Comput. Intell. Multimedia Underst. (IEEE IWCIM, 2015*), doi: 10.1109/IWCIM.2015.7347076.

[47] E. Hossain and G. Chetty, "Multimodal feature learning for gait biometric based human identity recognition," in: M. Lee, A. Hirose, Z. G. Hou, R. M. Kil (eds), Neural Information Processing (ICONIP 2013), *Lect. Notes Comput. Sci.,* vol 8227, pp.721-728, Springer, 2013, ISBN 978-3-642-42042-9.

[48] W. Kusakunniran, "Recognizing gaits on spatio-temporal feature domain," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 9, pp. 1416 – 1423, 2014, doi: 10.1109/TIFS.2014.2336379.

[49] M. Jeevan, N. Jain, M. Hanmandlu and G. Chetty, "Gait recognition based on gait pal and pal entropy image," in *Proc. IEEE Imag. Process. (ICIP)*, pp. 4195-4199, 2013, doi: 10.1109/ICIP.2013.6738864.

[50] C. Wang, J. Zhan, J. Pu, X. Yuan and L. Wang, "Chrono-gait image: A novel temporal template for gait recognition," in Computer Vision – ECCV 2010. *Lect. Notes Comp. Sci*., vol. 6311, pp. 257–270, *Springer 2010*, ISBN: 978-3-642-15549-9.

[51] C. Wang, J. Zhang, L.Wang, J. Pu and X. Yuan, "Human identification using temporal information preserving gait template," *IEEE Trans. Pattern Anal. Mach. Intell..,* vol. 34, no. 11, pp. 2164–2176, 2012, doi: 10.1109/TPAMI.2011.260.

[52] R. Gross and J. Shi," The CMU Motion of Body (MoBo) Database," *Tech. Rep. CMU-RI-TR-01-18*, 2001. http://ri.cmu.edu/pub_files/pub3/gross_ralph_2001_3/gross_ralph_2001_3.pdf [Accessed 8 December 2018].

[53] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother and K.W. Bowyer, "The humanID gait challenge problem: Data sets, performance and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, 2005, doi: 10.1109/TPAMI.2005.39.

[54] 'CASIA Gait Database'. Available at: http://www.cbsr.ia.ac.cn/english/GaitDatabases.asp [Accessed 8 Dec. 2018]

[55] Y. Makihara, H. Mannami, A. Tsuji, A. Hossain, K. Sugiura, A. Mori and Y. Yagi, "The OU-ISIR Gait Database Comprising the Treadmill Dataset," *IPSJ Trans. Comp. Vision. App.*, vol. 4, pp. 53-62, 2012, doi: 10.2197/ipsjtcva.4.53.

[56] H. Iwama, M. Okumura, Y. Makihara and Y. Yagi, "The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition," I*EEE Trans. Inf. Forensics Security.*, vol. 7, no. 5, pp. 1511-1521, 2012, doi: 10.1109/TIFS.2012.2204253.

[57] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller and G. Rigoll, "The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits," *J. Vis. Commun. Imag. Repr.*, vol. 25, no. 1, pp. 195–206, 2014, doi: 10.1016/j.jvcir.2013.02.006.

[58] T. Yeoh, H. E. Aguirre and K. Tanaka, "Clothing-invariant gait recognition using convolutional neural network," in *Proc. Int. Symp. ISPACS,* pp. 1-5, 2016, doi: 10.1109/ISPACS.2016.7824728.

[59] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo and Y. Yagi, "Geinet: View-invariant gait recognition using a convolutional neural network," in *Proc. 2016 Int. Conf. Biometrics (ICB)*, pp. 1–8, 2016, doi: 10.1109/ICB.2016.7550060.

[60] T. Wolf, M. Babaee and G. Rigoll, "Multi-view gait recognition using 3D convolutional neural networks," *Proc.2016 IEEE Int. Conf. Imag. Process. (ICIP)*, pp. 4165–4169, 2016, doi: 10.1109/ICIP.2016.7533144.

[61] F. M. Castro, M. J. M. Jiménez, N. Guil and N. P. D. L. Blanca "Automatic learning of gait signatures for people identification," In book: Advances in Computational Intelligence, I. Rojas et al. (Eds.): IWANN 2017, Part II, *Lect. Notes in Computer Sci.* vol. 10306, pp. 257–270, Springer, 2017, doi: 10.1007/978-3-319-59147-6_23.

[62] D. Muramatsu, Y. Makihara and Y. Yagi, "View transformation model incorporating quality measures for cross-view gait recognition," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1602–1615, 2016, doi: 10.1109/TCYB.2015.2452577.

[63] M. Hu, Y. Wang, Z. Zhang, D. Zhang and J. J. Little, "Incremental learning for video based gait recognition with LBP flow," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 77–89, 2013, doi: 10.1109/TSMCB.2012.2199310.

[64] D. Muramatsu, Y. Makihara and Y. Yagi , "Cross-view gait recognition by fusion of multiple transformation consistency measures," *IET Biometrics*, vol. 4, no. 2, pp. 62–73, 2015, doi: 10.1049/iet-bmt.2014.0042

[65] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo and Y. Yagi, "On Input/Output Architectures for Convolutional Neural Network-Based Cross-View Gait Recognition," *IEEE Trans. Circuits Syst. Video Technol. (early access),* doi: 10.1109/TCSVT.2017.2760835.

[66] C. Zhang, W. Liu, H. Ma and H. Fu, "Siamese neural network based gait recognition for human identification," in *Proc. IEEE Int. Conf. on Acoust., Speech Sig. Process. (ICASSP 2016)*, pp. 2832–2836, 2016, doi: 10.1109/ICASSP.2016.7472194.

[67] S. Chopra, R. Hadsell and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Comp. Soc. Conf. on Comput.*

*Vision Pattern Recogn. (CVPR'05).*, pp. 539–546, 2005, doi: 10.1109/CVPR.2005.202

[68] Z. Wu, Y. Huang and L. Wang, "Learning representative deep features for image set analysis," *IEEE Trans. Multimedia,* vol. 17, no. 11, pp. 1960–1968, 2015, doi: 10.1109/TMM.2015.2477681.

[69] Z. Wu, Y. Huang, L. Wang, X. Wang and T. Tan, "A comprehensive study on cross-view gait based human identification with deep cnns," *IEEE Trans*, *Pattern Anal. Mach. Intell.* vol. 39, no.2, pp. 209-226, 2017, doi: 10.1109/TPAMI.2016.2545669.

[70] K. Weiss, T. M. Khoshgoftaar and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1-40, 2016, doi: 10.1186/s40537-016-0043-6.

[71] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," in *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345-1359, 2010, doi: 10.1109/TKDE.2009.191.

[72] C. Li, X. Min, S. Sun, W. Lin and Z. Tang, "DeepGait: A Learning Deep Convolutional Resentation for Gait Recognition Using Joint Bayesian," *App. Sci.*, vol.7, no.13, 210, pp1-13, 2017, doi: 10.3390/app7030210.

[73] A. Schrag, M. Choudhury, D. Kaski and D. A. Gallagher, "Why Do Patients with Parkinson's Disease Fall? A Cross-Sectional Analysis of Possible Causes of Falls," *NPJ Parkinson's dis.* vol. 1, no. 15011, 2015, doi: 10.1038/npjparkd.2015.11.

[74] G. P. Panebianco, M. C. Bisi, R. Stagni and S. Fantozzi, "Analysis of the performance of 17 algorithms fr `om a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements," *Gait & Post.*, vol. 66, pp. 76-82, 2018, doi: 10.1016/j.gaitpost.2018.08.025.

[75] T. Zebin, P. J. Scully and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," *Proc. IEEE SENSORS*, pp.1-3, 2016, doi: 10.1109/ICSENS.2016.7808590.

[76] F. J. Ordóñez and D. Roggen, “Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition,” *Sensors*, vol. 16, no. 1, p. 115, 2016, doi: 10.3390/s16010115.

[77] M. Gadaleta, L. Merelli and M. Rossi, "Human authentication from ankle motion data using convolutional neural networks," *IEEE Statist. Sig. Process. Workshop (SSP)*, 2016, pp. 1-5, 2016, doi: 10.1109/SSP.2016.7551815.

[78] M. Gadaleta and M. Rossi. “IDNet: Smartphone-Based Gait Recognition with Convolutional Neural Networks,” *Pattern Recognition*, vol. 74, pp. 25–37, 2018, doi: 10.1016/j.patcog.2017.09.005.

[79] Y. Zhao and S. Zhou, “Wearable Device-Based Gait Recognition Using Angle Embedded Gait Dynamic Images and a Convolutional Neural Network,” *Sensors*, vol. 17, no. 3, p. 478, 2017, doi: 10.3390/s17030478.

[80] J. Frank, S. Mannor, J. Pineau and D. Precup, “Time Series Analysis Using Geometric Template Matching,” *IEEE Trans. Patt. Anal. Mach. Intell*., vol. 35, no. 3, pp. 740-754, 2013, doi: 10.1109/TPAMI.2012.121.

[81] O. Dehzangi, M. Taherisadr and R. ChangalVala, “IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion,” *Sensors,* vol. 17, no. 12, p. 2735, 2017, doi: 10.3390/s17122735.

[82] P. Lorenzi et al., “Smart sensors for the recognition of specific human motion disorders in Parkinson’s disease,” *Intern. Workshop Adv. Sens. Interfaces (IWASI 2015)*, pp. 131-136, 2015, doi: 10.1109/IWASI.2015.7184973.

[83] J. Camps et al., “Deep learning for freezing of gait detection in Parkinson’s disease patients in their homes using a waist-worn inertial measurement unit,” *Knowledge Based Syst.*, vol. 139, pp. 119-131, 2018, doi: 10.1016/j.knosys.2017.10.017.

[84] H.B. Kim et al., “Validation of Freezing-of-Gait Monitoring Using Smartphone,” *Telemed. e-Health.*. vol. 24, no. 11, 2018, doi: 10.1089/tmj.2017.0215.

[85] Y. Xia, J. Zhang, Q. Ye, N. Cheng, Y. Lu and D. Zhang, "Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients," *Biomed. Sig. Process. Control*, vol. 46, pp. 221-230, 2018, doi: 10.1016/j.bspc.2018.07.015.

[86] M. Bachlin et al.,"Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 436-446, 2010, doi: 10.1109/TITB.2009.2036165.

[87] N. M. Rad, T. v. Laarhoven, C. Furlanello and E. Marchiori, "Novelty Detection using Deep Normative Modelling for IMU-Based Abnormal Movement Monitoring in Parkinson's Disease and Autism Spectrum Disorders," *Sensors,* vol. 18, no. 10, 3533, 2018, doi: 10.3390/s18103533.

[88] D. Ravì, C. Wong, B. Lo and G. Yang, "A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 56-64, 2017, doi: 10.1109/JBHI.2016.2633287.

[89] N. Y. Hammerla, S. Halloran and T. Plotz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in S. Kambhampati (ed.) *Proc 25th Int. Joint Conf. Art. Intell.*, (*IJCAI* 2016), AAAI Press , pp. 1533–1540, 2016, ISBN 978-1-57735-770-4.

[90] A. Murad and J. Pyun, "Deep Recurrent Neural Networks for Human Activity Recognition," *Sensors*, vol. 17, no. 11, 2556, 2017, doi: 10.3390/s17112556.

[91] J. Hannink et al., "Mobile Stride Length Estimation with Deep Convolutional Neural Networks," in *IEEE J. Biomed. Health. Informat.*, vol. 22, no. 2, pp. 354-362, 2018, doi: 10.1109/JBHI.2017.2679486.

[92] A. Rampp, J. Barth, S. Schülein, K.-G. Gaßmann, J Klucken and B. Eskofier, "Inertial Sensor-Based Stride Parameter Calculation from Gait Sequences in Geriatric Patients," *IEEE Trans. Biomed. Eng.,* vol. 62, no. 4, pp. 1089-1097, 2015, doi: 10.1109/TBME.2014.2368211.

[93] A. N. Aicha, G. Englebienne, K. S. van Schooten, M. Pijnappels and B. Kröse, "Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry," *Sensors,* vol. 18, no. 5, 1654, 2018, doi: 10.3390/s18051654.

[94] B. Hu, P.C. Dixon, J.V. Jacobs, J.T. Dennerlein and J.M. Schiffman. "Machine Learning Algorithms Based on Signals from a Single Wearable Inertial Sensor Can Detect Surface- and Age-Related Differences in Walking," *J. Biomech.*, vol. 71, pp. 37–42, 2018, doi: 10.1016/j.jbiomech.2018.01.005.

[95] R. Vera-Rodriguez, J.S.D. Mason, J. Fierrez and J. Ortega-Garcia, "Analysis of Time Domain Information for Footstep Recognition," *Proc. Int. Symp. Visual Comp.*, In: Bebis G. et al. (eds) Advances in Visual Computing, ISVC 2010. *Lect. Notes Comput. Sci.,* vol. 6453, pp. 489-498, 2010, doi: 10.1007/978-3-642-17289-2_47.

[96] R. Vera-Rodriguez, J. S. D. Mason, J. Fierrez and J. Ortega-Garcia, "Analysis of Spatial Domain Information for Footstep Recognition," *IET Comp. Vision*, vol. 5, no. 6, pp. 380-388, 2011, doi: 10.1049/iet-cvi.2010.0189 .

[97] R. Vera-Rodriguez, J. S. D. Mason, J. Fierrez and J. Ortega-Garcia, "Comparative analysis and fusion of spatiotemporal information for footstep recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 823–834, 2013, doi: 10.1109/TPAMI.2012.164.

[98] A. Bertani, A. Cappello, M.G. Benedetti, L. Simoncini and F. Catani, "Flat foot functional evaluation using pattern recognition of ground reaction data," *Clin. Biomech.*, vol. 14, no. 7, pp. 484–493, 1999, doi: 10.1016/S0268-0033(98)90099-7.

[99] P. Leusmann, C. Mollering, L. Klack, K. Kasugai, M. Ziefle and B. Rumpe, "Your floor knows where you are: sensing and acquisition of movement data," *Proc. 12th IEEE Int. Conf. Mobile. Data. Manag.*, vol. 2, pp. 61–66, 2011, doi: 10.1109/MDM.2011.29.

[100] S. O'Shea, M. E. Morris and R. Iansek, "Dual task interference during gait in people with Parkinson disease: effects of motor versus cognitive secondary tasks," *Phys. Ther.*, vol. 82, no. 9, pp. 888–897, 2002, doi: 10.1093/ptj/82.9.888.

[101] J. H. Hollman, F. M. Kovash, J. J. Kubik and R. A. Linbo, "Age-related differences in spatiotemporal markers of gait stability during dual task walking," *Gait & Post.*, vol. 26, no. 1, pp. 113–119, 2007, doi: 10.1016/j.gaitpost.2006.08.005.

[102] J. Suutala, K. Fujinami and J. Rning, "Gaussian Process Person Identifier Based on Simple Floor Sensors," in: D. Roggen, C. Lombriser, G. Tröster, G. Kortuem and P. Havinga (eds), Smart Sensing and Context. EuroSSC 2008. *Lect. Notes Comput. Sci.*, vol. 5279, pp. 55-68, Springer, 2008, doi: 10.1007/978-3-540-88793-5_5.

[103] J.S. Yun, S.H. Lee, W.T. Woo and J.H. Ryu, "User Identification Using Walking Pattern over the ubiFloorII," in: Hao Y. et al. (eds) Computational Intelligence and Security. CIS 2005. *Lect. Notes Comput. Sci.*, vol. 3801, pp. 949-956, Springer, 2005, doi:/10.1007/11596448_141.

[104] L. Middleton, A. A. Buss, A. Bazin and M. S. Nixon, "A floor sensor system for gait recognition," *4th IEEE Workshop Autom. Ident. Adv. Technol. (AutoID'05)*, pp. 171-176, 2005, doi: 10.1109/AUTOID.2005.2.

[105] M. S. Singh, V. Pondenkandath, B. Zhou, P. Lukowic and Marcus Liwick, "Transforming sensor data to the image domain for deep learning — An application to footstep detection," *Proc. Int. Joint Conf. Neural Netw. (IJCNN 2017)*, pp. 2665-2672, 2017, doi: 10.1109/IJCNN.2017.7966182.

[106] J. A. Cantoral-Ceballos, N. Nurgiyatna, P. Wright, J. Vaughan, C. Brown-Wilson, P. J. Scully and K. B. Ozanyan, "Intelligent Carpet System, Based on Photonic Guided-Path Tomography, for Gait and Balance Monitoring in Home Environments," *IEEE Sensors J.*, vol. 15, no. 1, pp. 279-289, 2015, doi: 10.1109/JSEN.2014.2341455.

[107] J. A. Cantoral-Ceballos, N. Nurgiyatna, P. Scully and K. B. Ozanyan, "Smart carpet for imaging of objects' footprint by photonic guided-path tomography," Proc. *IEEE Africon '11*, pp. 1-6, 2011, doi:10.1109/AFRCON.2011.6071986.

[108] O. Costilla-Reyes, P. Scully and K. B. Ozanyan, "Temporal Pattern Recognition in Gait Activities Recorded With a Footprint Imaging Sensor System," *IEEE Sensors J.*, vol. 16, no. 24, pp. 8815-8822, 2016, doi: 10.1109/JSEN.2016.2583260.

[109] O. Costilla-Reyes, P. Scully and K. B. Ozanyan, “Deep Neural Networks for Learning Spatio-Temporal Features From Tomography Sensors,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 645–653, 2018, doi: 10.1109/TIE.2017.2716907.

[110] O. Costilla Reyes, R. Vera-Rodriguez, P. Scully and K. B. Ozanyan, “Analysis of Spatio-temporal Representations for Robust Footstep Recognition with Deep Residual Neural Networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol: 41 , Issue. 2 pp. 285 - 296, 2019, doi: 10.1109/TPAMI.2018.2799847.

[111] J. Llinas and D. L. Hall, "An introduction to multi-sensor data fusion," *Proc.1998 IEEE Int. Symp. Circuits Syst.* (*ISCAS '98)*, vol.6. pp. 537-540, 1998, doi: 10.1109/ISCAS.1998.705329.

[112] Z. Ding et al., “The Real Time Gait Phase Detection Based on Long Short-Term Memory,” *Proc. 2018 IEEE Third Int. Conf. Data Sci. in Cybern. (DSC)*, pp. 33–38, 2018, doi: 10.1109/DSC.2018.00014.

[113] H. T. T. Vu, F. Gomez, P. Cherelle, D. Lefeber, A. Nowé and B. Vanderborght, “ED-FNN: A New Deep Learning Algorithm to Detect Percentage of the Gait Cycle for Powered Prostheses,” *Sensors*, vol. 18, no. 7, 2389, 2018, doi: 10.3390/s18072389.

[114] O. Mazumder, A. Sankar, K. P. Kumar Lenka and S. Bhaumik, “Multi-Channel Fusion Based Adaptive Gait Trajectory Generation Using Wearable Sensors,” *J. Intell. & Robot. Syst.*, vol. 86, no. 3-4, pp. 335–351, 2016, doi: 10.1007/s10846-016-0436-y.

[115] K. R. Mun, G. Song, S. Chun and J. Kim, “Gait Estimation from Anatomical Foot Parameters Measured by a Foot Feature Measurement System using a Deep Neural Network Model,” Nature, vol. 8, 9879, 29 2018, doi: 10.1038/s41598-018-28222-2.

[116] P. Kumar, S. Mukherjee, R. Saini, P. Kaushik, P. P. Roy and D. P. Dogra, "Multimodal Gait Recognition with Inertial Sensor Data and Video using Evolutionary

Algorithm," in *IEEE Trans. Fuzzy Syst.* vol. 27, Issue: 5, pp. 956 – 965, 2019, doi: 10.1109/TFUZZ.2018.2870590.

[117] S. Mirjalili, S. M. Mirjalili and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[118] R. Vera-Rodriguez, J. Fierrez, J. S. D. Mason and J. Orteua-Garcia, "A novel approach of gait recognition through fusion with footstep information," *Proc. 2013 Inter. Conf. Biometrics (ICB)*, pp. 1-6, 2013, doi: 10.1109/ICB.2013.6613014.

[119] ABC news: "Chinese authorities use 'gait' surveillance to identify people by their body shape and walk". https://www.abc.net.au/news/2018-11-06/chinese-gait-recognition-tech-ids-people-by-how-they-walk/10469974. [Accessed 14 Dec. 2018].

[120] P. Zhang, Q. Wu and J. Xu, "VT-GAN: View Transformation GAN for Gait Recognition Across Views," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, pp. 1-8,. 2019. doi: 10.1109/IJCNN.2019.8852258.

[121] M. Babaee, Y. Zhu, O. Köpüklü, S. Hörmann and G. Rigoll, "Gait Energy Image Restoration Using Generative Adversarial Networks," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, pp. 2596-2600, 2019. doi: 10.1109/ICIP.2019.8803236.

[122] X. Chen et al, "Multi-View Gait Image Generation for Cross-View Gait Recognition" IEEE Transactions on Image Processing, Vol. 30, pp. 3041 – 3055, 2021. doi: 10.1109/TIP.2021.3055936.

[123] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition", Ipsj Trans. Comput. Vis. Appl., vol. 10, no. 1, pp. 4, 2018. doi: 10.1186/s41074-018-0039-6.

[124] A. Balakrishnan, J. Medikonda and P. K. Namboothiri, "Analysis of the effect of muscle fatigue on gait characteristics using data acquired by wearable sensors,"

2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), Udupi, India, 2020, pp. 137-140, doi: 10.1109/DISCOVER50404.2020.9278096.

[125] F. Muheidat and L. A. Tawalbeh, "In-Home Floor Based Sensor System-Smart Carpet- to Facilitate Healthy Aging in Place (AIP)," in *IEEE Access*, vol. 8, pp. 178627-178638, 2020. doi: 10.1109/ACCESS.2020.3027535.

[126] O. Costilla-Reyes, P. Scully, I. Leroi and K. B. Ozanyan, "Age-Related Differences in Healthy Adults Walking Patterns Under a Cognitive Task With Deep Neural Networks," in *IEEE Sensors Journal*, vol. 21, no. 2, pp. 2353-2363, 15 Jan.15, 2021. doi: 10.1109/JSEN.2020.3021349.

[127] A. Turner and S. Hayes, "The Classification of Minor Gait Alterations Using Wearable Sensors and Deep Learning" *IEEE Transactions on Biomedical Engineering*, vol. 66, no.11, pp. 3136 – 3145. Nov. 2019. doi: 10.1109/TBME.2019.2900863.

[128] L. Tran, T. Hoang, T. Nguyen, H. Kim and D. Choi, "Multi-Model Long Short-Term Memory Network for Gait Recognition Using Window-Based Data Segment" IEEE Access , vol. 9, pp. 23826 – 23839 2021. doi: 10.1109/ACCESS.2021.3056880.

[129] Q. Zou, Y. Wang, Q. Wang, Y. Zhao and Q. Li, "Deep learning-based gait recognition using smartphones in the wild", IEEE Trans. Inf. Forensics Security, vol. 15, pp. 3197-3212, 2020. doi: 10.1109/TIFS.2020.2985628.

[130] T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa and Y. Yagi, "The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication", Pattern Recognit., vol. 47, no. 1, pp. 228-237, 2014. doi: 10.1016/j.patcog.2013.06.028.

[131] A. S. Alharthi, S. U. Yunas and K. B. Ozanyan, "Deep Learning for Monitoring of Human Gait: A Review," IEEE Sensors Journal, vol. 19, no. 21, pp. 9575-9591, 2019. doi: 10.1109/JSEN.2019.2928777.

[132] S. M. Waechter *et al.* "The impact of dual tasking on cognitive performance in a Parkinson's disease cohort with and without freezing of gait: An EEG and behavioral based approach," *7th International IEEE/EMBS Conference on Neural Engineering (NER)*. pp. 1072-1075, April 2015. doi: 10.1109/NER.2015.7146813.

[133] M. J. Devine, H. Plun-Favreau, and N. W. Wood, "Parkinson's disease and cancer: two wars, one front," *Nature Reviews Cancer.* vol. 11, no. 11, pp. 812-823, 2011. doi: 10.1038/nrc3150.

[134] Lonneke ML de Lau and M. M. B. Breteler, "Epidemiology of Parkinson's disease," *The Lancet Neurology*, vol. 5, Issue 6, pp. 525-535, ISSN 1474-4422, 2006. doi: 10.1016/S1474-4422(06)70471-9.

[135] J. S. Perlmutter, "Assessment of Parkinson disease manifestations," *Current protocols in neuroscience*. Chapter. 10, Unit10.1, October 2009. doi:10.1002/0471142301.ns1001s49.

[136] G. Ebersbach, C. Moreau, F. Gandor, L. Defebvre, and D. Devos, "Clinical syndromes: parkinsonian gait," *Movement Disorder*. vol. 28, Issue. 11, pp. 1552–1559, September 2013. DOI: 10.1002/mds.25675.

[137] L. Medeiros, H. Almeida, L. Dias, M. Perkusich, and R. Fischer, "A Gait Analysis Approach to Track Parkinson's Disease Evolution Using Principal Component Analysis," *IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)*. pp. 48-53, June 2016. doi: 10.1109/CBMS.2016.14.

[138] Y. Wu, P. Chen, X. Luo, M. Wu, L. Liao, S. Yang, and R.M. Rangayyan, "Measuring signal fluctuations in gait rhythm time series of patients with parkinson's disease using entropy parameters," *Biomedical Signal Processing and Control*. vol. 31, pp. 265-271, January 2017. doi: org/10.1016/j.bspc.2016.08.02.

[139] A. Zhao, L. Qi, J. Li, J. Dong and H. Yu, "A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data," *Neurocomputing*. vol. 315, pp. 1-8, ISSN 0925-2312, 2018. doi: 10.1016/j.neucom.2018.03.032.

[140] S. Bach, A. Binder, G. Montavon, F. Klauschen, K. R Müller, W. Samek, "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation," *PLoS ONE.* vol. 10 np. 7, 2015. doi: 10.1371/journal.pone.0130140.

[141] M. Kohlbrenner, A. Bauer, S. Nakajima, A. Binder, W. Samek and S. Lapuschkin, "Towards best practice in explaining neural network decisions with LRP," Computer Vision and Pattern Recognition (cs.CV); Machine Learning (stat.ML), 2019. arXiv:1910.09840.

[142] S. Jolly, B. K. Iwana, R. Kuroki, and S. Uchida, "How do Convolutional Neural Networks Learn Design?" 24th International Conference on Pattern Recognition (ICPR). pp. 1085-1090, August 2018. doi: 10.1109/ICPR.2018.8545624.

[143] W. Samek, A. Binder, S. Lapuschkin and K. Müller, "Understanding and Comparing Deep Neural Networks for Age and Gender Classification," 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). pp. 1629-1638, January 2018 doi: 10.1109/ICCVW.2017.191.

[144] F. Horst, S. Lapuschkin, W. Samek, K. R. Müller and W. I. Schöllhorn, "Explaining the unique nature of individual gait patterns with deep learning," Nature. Scientific Reports 9, Article number: 2391, February 2019. doi: 10.1038/s41598-019-38748-8.

[145] A. S. Alharthi, Krikor B. Ozanyan, "Deep Learning for Ground Reaction Force Data Analysis: Application to Wide-Area Floor Sensing," 28th International Symposium on Industrial Electronics (ISIE), 2019 doi: 10.1109/ISIE.2019.8781511.

[146] G, Yogev, N. Giladi, N. Peretz, S. Springer, E. S. Simon, and M. J. Hausdorff, "Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?" *European Journal of Neuroscience.* vol. 22, pp. 1248-1256, September 2005. doi: 10.1111/j.1460-9568.2005.04298.

[147] M. J. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz, and N. Giladi, "Rhythmic auditory stimulation modulates gait variability in Parkinson's

disease," *European Journal of Neuroscience.* vol. 26, pp. 2369-2375, October 2007. doi: 10.1111/j.1460-9568.2007.05810.

[148] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. M. Hausdorff, "Treadmill walking as a pacemaker to improve gait rhythm and stability in Parkinson's disease," *Movement Disorders*. vol. 20, pp. 1109-1114, September 2005. doi: 10.1002/mds.20507.

[149] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). Circulation. 101(23):e215-e220. doi: 10.13026/C24H3N.

[150] A. S. Alharthi, Krikor B. Ozanyan, "Deep learning and Sensor Fusion Methods for Cognitive Load Gait Difference in Males and Females," 20th International Conference on Intelligent Data Engineering and Automated Learning, 2019. doi: 10.1007/978-3-030-33607-3_25.

[151] Abdullah S. Alharthi, Krikor B. Ozanyan, "Sensor Fusion for Analysis of Gait under Cognitive Load Deep Learning Approach," *IEEE Sensors Applications Symposium proceedings,* 2020. doi: 10.1109/SAS48726.2020.9220046.

[152] F. Chollet, Keras "The python deep learning library", [Online]. Available: https://keras.io/

[153] Maximilian, A., et al. "iNNvestigate neural networks". https://github.com/albermax/innvestigate. (2018).

[154] C. Szegedy et al., "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, pp. 1-9, 2015. doi: 10.1109/CVPR.2015.7298594.

[155] J.Bromley, I. Guyon, Y. LeCun, E. Sickinger and R. Shah, "Signature verification using a "siamese" time delay neural network," *Advances in Neural Information Processing Systems*. vol. 6, pp. 737–744, 1994. doi: 10.1142/S0218001493000339.

[156] E. Hoffer and N. Ailon, "Deep Metric Learning Using Triplet Network," *In: Feragen A., Pelillo M., Loog M. (eds) Similarity-Based Pattern Recognition. SIMBAD 2015. Lecture Notes in Computer Science*. vol 9370. 2015, Springer, Cham. doi: 10.1007/978-3-319-24261-3_7.

[157] D. P. Kingma and J. L. Ba, "Adam: a Method for Stochastic Optimization," 3rd International Conference on Learning Representations (ICLR). 2015. https://arxiv.org/abs/1412.6980.

[158] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in ECCV, pp. 818–833, 2014. arXiv:1311.2901.

[159] J. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller, "Striving for Simplicity: The All Convolutional Net," 2015. arXiv:1412.6806.

[160] Y. N. Jane, H. K. Nehemiah, and K. Arputharaj, "A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease," *Journal of Biomedical Informatics*. vol. 60, pp. 169-176, ISSN 1532-0464, 2016. doi:10.1016/j.jbi.2016.01.014.

[161] Ö. F. Ertuğrul, Y. Kaya, R. Tekin, and M. N. Almalı, "Detection of Parkinson's disease by Shifted One Dimensional Local Binary Patterns from gait," *Expert Systems with Applications*. vol. 56, pp. 156-163, September 2016. doi: 10.1016/j.eswa.2016.03.018.

[162] 'European Parkinson's Disease Association, Rating scales'. Available at: https://www.epda.eu.com/about-parkinsons/symptoms/rating-scales/[Accessed 25 Jun. 2020].

[163] J. M. Hausdorff, A. Schweiger, T. Herman, G. Yogev-Seligmann and N. Giladi, "Dual-task decrements in gait: contributing factors among healthy older adults," Journal of gerontology, Series A, Biological sciences and medical sciences. vol. 63, Issue. 12, pp. 1335-43, December 2008. doi: 10.1093/gerona/63.12.1335.

[164] G. Yogev, J. M. Hausdorff, and , N. Giladi "The role of executive function and attention in gait" *Official Journal of the Movement Disorder Society*. Vol. 23, no.3, pp. 329–342, 2008. doi: 10.1002/mds.21720.

[165] M. Woollacott, A. Shumway-Cook, "Attention and the control of posture and gait: a review of an emerging area of research" *Gait & Posture*. Vol. 16, no. 1, pp. 1–14, 2002. doi: 10.1016/S0966-6362(01)00156-4.

[166] S. Ho, A. Mohtadi, K. Daud, U. Leonards and T. C. Handy, "Using smartphone accelerometry to assess the relationship between cognitive load and gait dynamics during outdoor walking," Nature. Scientific Reports 9, Article number: 3119, February 2019. doi: 10.1038/s41598-019-39718-w.

[167] A. M. F. Barela, G. L. Gama, D. V. Russo-Junior, M. L. Celestino and J. A. Barela,"Gait alterations during walking with partial body weight supported on a treadmill and over the ground," Nature. Scientific Reports 9, Article number: 8139, May 2019. doi: 10.1038/s41598-019-44652-y.

[168] C. Meyer, T. Killeen, C. S. Easthope, A. Curt, M. Bolliger, M. Linnebank, B. Zörner and Linard Filli, "Familiarization with treadmill walking: How much is enough?" Nature. Scientific Reports 9, Article number: 5232, March 2019. doi: 10.1038/s41598-019-41721-0.

[169] J. M. Hausdorff, D. Rios, and H. K. Edelberg, "Gait variability and fall risk in community-living older adults: a 1-year prospective study," Archives of Physical Medicine and Rehabilitation. vol. 82, Issue. 8, pp. 1050-1056, August 2001. doi: 10.1053/apmr.2001.24893.

[170] Y. Barak, R. C. Wagenaar, and K. G. Holt, "Gait characteristics of elderly people with a history of falls: a dynamic approach," Physical Therapy. vol. 86, Issue. 11, pp. 1501–1510, November 2006. doi: 10.2522/ptj.20050387.

[171] R. McArdlea, R. Morrisa, J. Wilsona, B. Galnaa and A. J. Thomasa, "What Can Quantitative Gait Analysis Tell Us about Dementia and Its Subtypes? A Structured Review," Journal of Alzheimer's Disease. vol. 60, pp. 1295–1312, 2017. doi:10.3233/JAD-170541.

[172] L. Rochester, B. Galna, S. Lord, and D. Burn, "The nature of dual-task interference during gait in incident Parkinson's disease," Neuroscience. vol. 265, pp. 83-94, April 2014. doi: 10.1016/j.neuroscience.2014.01.041.

[173] S. U. Yunas, A. Alharthi and K. B. Ozanyan, "Multi-modality fusion of floor and ambulatory sensors for gait classification," 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), Vancouver, BC, Canada, 2019, pp. 1467-1472. doi: 10.1109/ISIE.2019.8781127.

[174] C. François, "Keras". https://keras.io/, 2015.

[175] P. Chopra, D. M. Castelli and J. B. Dingwell, "Cognitively Demanding Object Negotiation While Walking and Texting," Scientific Reports. vol. 8, no. 17880, 2018. doi: 10.1038/s41598-018-36230-5.

[176] S. E. Soh, M. E. Morris, and J. L. McGinley, "Determinants of health-related quality of life in Parkinson's disease: A systematic review," Parkinsonism & Related Disorders. vol. 17, Issue. 1, pp. 1–9, January 201. doi: 10.1016/j.parkreldis.2010.08.012.

[177] L. M. Allan, Ballard, C. G. Burn, D. J. Burn, and R.A. Kenny, "Prevalence and Severity of Gait Disorders in Alzheimer's and Non-Alzheimer's Dementias," Journal of the American Geriatrics Society. vol. 53 Issue. 10, pp. 1681-1687, October 2005. doi: 10.1111/j.1532-5415.2005.53552.x.

[178] E. Heremans, A. Nieuwboer, and S. Vercruysse, "Freezing of Gait in Parkinson's Disease: Where Are We Now?" Current Neurology and Neuroscience Reports. Online ISSN1534-6293, April 2013. doi: 10.1007/s11910-013-035.

[179] Y. Shoji ; T. Takasuka ; H. Yasukawa, "Personal identification using footstep detection," Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. 18-19 Nov. 2004 doi: 10.1109/ISPACS.2004.1439012.

[180] A. Elkholy, M. E. Hussein, W. Gomaa, D. Damen and E. Saba, "Efficient and Robust Skeleton-Based Quality Assessment and Abnormality Detection in Human Action Performance," IEEE Journal of Biomedical and Health Informatics. Vol. 24, no. 1, pp. 280-291, Jan 2020. doi: 10.1109/JBHI.2019.2904321.

[181] J.E.A. Bertram, A Ruina, “Multiple walking speed–frequency relations are predicted byconstrained optimization,” Journal of Theoretical Biology. Vol. 209, Issue 4, pp. 445–453, 2001. doi: 10.1006/jtbi.2001.2279.

[182] K. G. Holt, S. F. Jeng, R. Ratcliffe, J. Hamill, “Energetic Cost and Stability during Human Walking at the Preferred Stride Frequency” Journal of Motor Beha. Vol. 27, no. 2, pp. 164–178, 1995. doi: 10.1080/00222895.1995.9941708.

[183] J. C. Selinger, , S. M. O’Connor, J. D. Wong, J. M. Donelan, “Humans Can Continuously Optimize Energetic Cost during Walking” *Current Biology*. Vol. 25 no. 18, pp. 2452–2456, 2015. doi: 10.1016/j.cub.2015.08.016.

[184] Marengoni, A. et al. “Aging with multimorbidity: A systematic review of the literature,” Ageing Research Reviews. vol. 10, no. 4 pp. 430–439, 2011. doi: 10.1016/j.arr.2011.03.003.

[185] K. B. Ozanyan, "Tomography defined as sensor fusion," 2015 IEEE SENSORS, Busan, Korea (South), 2015, pp. 1-4, doi: 10.1109/ICSENS.2015.7370554.

[186] K. B. Ozanyan, S. G. Castillo and F. J. P. Ortiz, "Guided-path tomography sensors for nonplanar mapping," in IEEE Sensors Journal, vol. 5, no. 2, pp. 167-174, April 2005, doi: 10.1109/JSEN.2005.843895.

[187] J. Cantoral Ceballos, N. Nurgiyatna, P. Scully and K. B. Ozanyan, "Smart carpet for imaging of objects' footprint by photonic guided-path tomography," IEEE Africon '11, Victoria Falls, Zambia, 2011, pp. 1-6, doi: 10.1109/AFRCON.2011.6071986.

[188] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, K. R. Müller, “Explainable AI: Interpreting, Explaining and Visualizing Deep Learning,” Springer, Cham LNCS, 11700, 2019. doi: 10.1007/978-3-030-28954-6.

[189] C. Seibold, W. Samek, A. Hilsmann, P. Eisert, “Accurate and Robust Neural Networks for Face Morphing Attack Detection,” Journal of Information Security and Applications. Vol. 53, 2020. doi: 10.1016/j.jisa.2020.102526.

[190] L. Arras, F. Horn, G. Montavon, KR. Müller, W Samek, "What is Relevant in a Text Document?": An Interpretable Machine Learning Approach," PLOS ONE, vol. 12, Issue 8, 2017. doi: 10.1371/journal.pone.0181142.

[191] S. Becker, M. Ackermann, S. Lapuschkin, KR. Müller, W. Samek, "Interpreting and Explaining Deep Neural Networks for Classification of Audio Signals," 2018, arXiv:1807.03418.

[192] Lutz Roeder, "Netron," https://github.com/lutzroeder/netron, 2020.

[193] A. S. Alharthi, A. J. Casson K. B. Ozanyan, "Gait Spatiotemporal Signal Analysis for Parkinsons Disease Detection and Severity Rating," IEEE Sensors Journal, vol. 21, Issue. 2, pp. 1838 - 1848 2020. doi: 10.1109/JSEN.2020.3018262.

[194] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015. arXiv:1502.03167.

[195] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research. vol. 15 Issue 56, pp. 1929–1958, 2014. https://jmlr.org/papers/v15/srivastava14a.html

[196] Z. Zhang and M. R. Sabuncu, "Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels," 2nd Conference on Neural Information Processing Systems (NeurIPS 2018). arXiv:1805.07836.

[197] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, ser. Proceedings of Machine Learning Research, Y.W. Teh and M. Titterington, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, pp. 249– 256, 2010. https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf

[198] T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," Twenty-first International Conference on Machine Learning, ser. ICML '04, 2004. doi: 10.1145/1015330.1015332.

[199] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,'' Amer. Statist., vol. 46, no. 3, pp. 175_185, 1992. https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879

[200] G. Xiao, Q. Cheng and C. Zhang, “Detecting Travel Modes Using Rule-Based Classification System and Gaussian Process Classifier,” IEEE Access. Vol. 7, pp. 116741 – 116752, 2019. doi: 10.1109/ACCESS.2019.2936443.

[201] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: A comprehensive review,” Neural Computation. vol. 29, no. 9, pp. 2352–2449, 2017. doi: 10.1162/neco_a_00990.

[202] A. R. Anwary, H. Yu, A. Callaway and M. Vassallo, "Validity and Consistency of Concurrent Extraction of Gait Features Using Inertial Measurement Units and Motion Capture System," in IEEE Sensors Journal, vol. 21, no. 2, pp. 1625-1634, 15 Jan.15, 2021, doi: 10.1109/JSEN.2020.3021501.

[203] E. J. L. Pulgarin, G. Herrmann and U. Leonards, “Drivers' Manoeuvre Prediction for Safe HRI” 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1-5 Oct. 2018. doi: 10.1109/IROS.2018.8593957.

[204] S. Shinkai et al. “Walking speed as a good predictor for the onset of functional dependence in a Japanese rural community population” *Age and Ageing,* vol. 29, pp. 441–446, 2000, https://doi.org/10.1093/ageing/29.5.441.

[205] Gras, L. *et al.* “Balance and gait of adults with very mild Alzheimer’s disease” *Journal of Geriatric Physical Therapy.* 38, 1–7 (2015). doi: 10.1519/JPT.0000000000000020.

[206] C. Schreiber and F. Moissenet, ”A multimodal dataset of human gait at different walking speeds established on injury-free adult participants,” Scientific Data Nature, vol. 6, no. 111, 2019. doi: 10.1038/s41597-019-0124-4.

[207] C. Schreiber and F. Moissenet, ”A multimodal dataset of human gait at different walking speeds established on injury-free adult participants,”*figshare*. doi: 10.6084/m9.figshare.7734767, 2019.

[208] "Mokka - Motion Kinematic & Kinetic Analyzer" http://biomechanical-toolkit.github.io/mokka/index.html [access May 2021]

[209] A. Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scalear" arXiv preprint, 2020, Xiv:2010.11929.

[210] A. Vaswani et al. "Attention Is All You Needar" 2017, Xiv:1706.03762.

[211] Z. Dai, H. Liu, Q. V. Le and M. Tan, "CoAtNet: Marrying Convolution and Attention for All Data Sizes" arXiv preprint, 2021, arXiv:2106.04803.

[212] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding" arXiv preprint, 2018. arXiv:1810.04805.

[213] D. Rakita, B. Mutlu, M. Gleicher, L. M. Hiatt "Shared-Control-Based Bimanual Robot Manipulation" *Science Robotics.* Vol. 4, no. 30, eaaw0955. May 2019. doi: 10.1126/scirobotics.aaw0955.