

**BIOLOGICAL AND BIOMIMETIC  
MACHINE LEARNING FOR  
AUTOMATIC CLASSIFICATION OF  
HUMAN GAIT**

**Viswadeep Sarangi**

Doctor of Philosophy

University of York

Electronic Engineering

December 2020

# Abstract

Machine learning (ML) research has benefited from a deep understanding of biological mechanisms that have evolved to perform comparable tasks. Recent successes of ML models, superseding human performance in human perception based tasks has garnered interest in improving them further. However, the approach to improving ML models tends to be unstructured, particularly for the models that aim to mimic biology. This thesis proposes and applies a bidirectional learning paradigm to streamline the process of improving ML models' performance in classification of a task, which humans are already adept at. The approach is validated taking human gait classification as the exemplar task. This paradigm possesses the additional benefit of investigating underlying mechanisms in human perception (HP) using the ML models. Assessment of several biomimetic (BM) and non-biomimetic (NBM) machine learning models on an intrinsic feature of gait, namely the gender of the walker, establishes a functional overlap in the perception of gait between HP and BM, selecting the Long-Short-Term-Memory (LSTM) architecture as the BM of choice for this study, when compared with other models such as support vector machines, decision trees and multi-layer perceptron models. Psychophysics and computational experiments are conducted to understand the overlap between human and machine models. The BM and HP derived from psychophysics experiments, share qualitatively similar profiles of gender classification accuracy across varying stimulus exposure durations. They also share the preference for motion-based cues over structural cues (BM=H>NBM). Further evaluation reveals a human-like expression of the inversion effect, a well-studied cognitive bias in HP that reduces the gender classification accuracy to 37% ( $p<0.05$ , chance at 50%) when exposed to inverted stimulus. Its expression in the BM supports the argument for learned rather than hard-wired mechanisms in HP. Particularly given the emergence of the effect in every BM, after training multiple randomly initialised BM models without prior anthropomorphic expectations of gait.

The above aspects of HP, namely the preference for motion cues over structural cues and the lack of prior anthropomorphic expectations, were selected to improve BM performance. Representing gait explicitly as motion-based cues of a non-anthropomorphic, gender-neutral skeleton not only mitigates the inversion effect in BM, but also improves significantly the classification accuracy. In the case of gender classification of upright stimuli, mean accuracy improved by 6%, from 76% to 82% ( $F_{1,18} = 16$ ,  $p<0.05$ ). For inverted stimuli, mean accuracy improved by 45%, from 37% to 82% ( $F_{1,18} = 20$ ,  $p<0.05$ ). The model was further tested on a more challenging, extrinsic feature task; the classification of the emotional state of a walker. Emotions were visually induced in subjects through exposure to emotive or neutral images from the International Affective Picture System (IAPS) database. The classification accuracy of the BM was significantly above chance at 43% accuracy ( $p<0.05$ , chance at 33.3%). However, application of the proposed paradigm in further binary emotive state classification experiments, improved mean accuracy further by 23%, from 43% to 65% ( $F_{1,18} = 7.4$ ,  $p<0.05$ ) for the positive vs. neutral task. Results validate the proposed paradigm of concurrent bidirectional investigation of HP and BM for the classification of human gait, suggesting future applications for automating perceptual tasks for which the human brain and body has evolved.

# Acknowledgement

Throughout this thesis I have received an immeasurable amount of support and assistance from people I shall always cherish. I would first like to thank my supervisor Dr. Adar Pelah, who has been a mentor to me in more than an academic capacity. I'm extremely grateful for the encouragement, the help in formulation of the research topic, the methodology and execution of the experiments. Thank you so much for all those discussions that extended well into midnight, with talks of the future of man and machine-kind, and the need for both to work together and learn from each other. I really appreciate the free flow of knowledge between us, about artificial intelligence, human neuroscience and life in general. It has left a lasting impression on me.

I owe immense gratitude to Dr. Thomas Stone for the exposure to the practical applications of gait analysis in a clinical setting. Seeing the possible end result of a good automatic gait analyser has been my constant driving force for the development of the computational model, for adhering to a high standard of analysis and rigor in evaluation of any improvements made.

I would like to thank our collaborators Prof. Elan Barenholtz and Dr. William Edward Hahn, from the Center for Complex Systems and Brain Sciences, Florida Atlantic University, for their wonderful contributions in spearheading the psychophysical and design of the machine learning side of the research and helping formulate and refine the research questions.

Thank you also to Dr. Terrence Barnhardt and his students David Bickham and Belen Wertheimer, in the Psychology Department, Florida Atlantic University, for their immense help in the collection of data for the emotional gait experiments. Their contribution in my analysis of the connection between human emotions and gait is extremely appreciated.

I would also like to thank my thesis advisory panel, Dr. Dimitar Kazakov for his valuable comments and input. Without them, this work simply would not have been possible.

All of the above would not have been possible without the valuable support of the team at Translate-MedTech and GrowMedTech. They provided me with the necessary guidance and tools to understand the impact of my research and how it could potentially change the world. Their support enabled my travels to meet the eminent clinicians and researchers of the world, in Florida, Cambridge and India, to bring the knowledge and experience back into my lab here in York.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. Importantly, there are my friends, who were in great support in deliberating over our problems and findings, both academically and personally. Finally, Milan, Vakilna, Kashi, Chow and Malin Aeri Imo Kristina Olsson Sundstrom, this journey has been transformative for me in so many ways and a lot of it goes out to you and I hope done justice in communicating my appreciation to you!

# Table of Contents

Abstract	ii
Acknowledgement	iii
List of Figures	ix
List of Tables	xi
Author's Declaration	xii
Chapter 1 Introduction	1
Chapter 2 Background	6
2.1 Gait Analysis	6
2.1.1 Human Gait	6
2.1.2 History of Human Gait Analysis	7
2.1.3 Gait Motion Capture Technology	8
2.1.3.1 Marker based Motion Capture Systems	8
2.1.3.2 Marker-less Motion Capture Systems	10
2.1.3.3 RGB based video capture	12
2.1.3.4 Inertial Measurement Unit (IMU)	13
2.1.3.5 Electromyogram (EMG)	13
2.1.4 Gait Analysis through Human Perception	14
2.2 Machine Learning	15
2.2.1 Classification	16
2.2.2 Support Vector Machines	17
2.2.3 Decision Trees (ID3)	17
2.2.4 Artificial Neural Networks	18
2.2.4.1 Recurrent Neural Network (RNN)	19
2.2.4.2 The Vanishing and Exploding Gradients Problem in RNNs	21
2.2.4.3 Long Short Term Memory (LSTM)	22
2.2.5 Model Comparison	23
2.2.6 Backpropagation and Genetic Algorithms	24
2.3 Related Work	25
2.3.1 Automation of Gait Classification	26
Chapter 3 Methods	30
3.1 Data Collection	30
3.1.1 Intrinsic Feature: Gender	33

3.1.2 Extrinsic Feature: Emotion	34
3.2 Stimuli Development	35
3.2.1 Psychophysical Experiments	35
3.2.2 Computational Experiments	36
3.3 Model Development	36
3.4 Data Analysis	37
Chapter 4 Biological, biomimetic and non-biomimetic learning	38
4.1 Experiment 1: Biological Models	39
4.1.1 Experiment Setup	40
4.1.1.1 Stimuli	40
4.1.1.2 Biological Model (HP)	41
4.1.2 Results	41
4.2 Experiment 2: Biomimetic Models	42
4.2.1 Choice of Biomimetic Model (BM)	43
4.2.2 Experiment Setup	44
4.2.2.1 Biomimetic model (BM)	44
4.2.2.2 Data Input.	44
4.2.2.3 Model Training and Testing	45
4.2.3 Results	45
4.3 Experiment 3: Non-Biomimetic Model (NBM)	47
4.3.1 Experiment Setup	47
4.3.1.1 Non-biomimetic Model (NBM)	47
4.3.1.2 Data Input	48
4.3.1.3 Model Training and Testing	48
4.3.2 Results	49
4.4 Experiment 4: Biomimetic (BM) and Non-Biomimetic Models (NBM) with Velocity Cues	50
4.4.1 Experiment Setup: Biomimetic Model (BM)	51
4.4.1.1 Biomimetic model (BM)	51
4.4.1.2 Biomimetic Data Input	51
4.4.2 Results	51
4.4.2.1 Biomimetic model (BM)	51
4.4.3 Method: Non-Biomimetic Model (NBM)	52
4.4.3.1 Non-Biomimetic Model (NBM)	52

4.4.3.2 Non-Biomimetic Data Input	52
4.4.4 Results	53
4.4.4.1 Non-biomimetic models (NBM)	53
4.5 Experiment 5: Biomimetic Models with Two-dimensional Input	54
4.5.1 Method	54
4.5.1.1 Biomimetic model (BM)	54
4.5.1.2 Data Input	54
4.5.2 Results	55
4.6 Discussion	56
4.7 Conclusions	58
Chapter 5 Biological and Biomimetic Perception of Gait	60
5.1 Experiment 1: Inversion Effect	61
5.1.1 Experiment Setup	62
5.1.1.1 Data Input	62
5.1.1.2 Biomimetic Model (BM)	62
5.1.2 Results	62
5.2 Experiment 2: Contribution of Structural Cues for Biomimetic Models	64
5.2.1 Method	65
5.2.1.1 Biomimetic Model	65
5.2.1.2 Data Input	65
5.2.2 Results	67
5.3 Experiment 3: Spatiotemporal Pre-processing Strategies for Biomimetic Models	67
5.3.1 Method	68
5.3.1.1 Biomimetic Model	68
5.3.1.2 Data Input	68
5.3.2 Results	69
5.3.2.1 Upright (Right-Side-Up) Skeleton	69
5.3.2.2 Vertically Inverted Skeleton	70
5.4 Experiment 4: Evaluating the Spatiotemporal Pre-processing Strategies for Biomimetic Models on 2D Gait Data	73
5.4.1 Results	74
5.4.1.1 Upright (right-side-up) Skeleton	74
5.4.1.2 Vertically Inverted Skeleton	76
5.5 Discussion	77

5.6 Conclusions	78
Chapter 6 Extrinsic Feature Classification Using Biomimetic Models: Emotion Classification from Human Gait	80
6.1 Experiment 1: Classification of Emotion from Human Gait Using Biomimetic Models	80
6.1.1 Data Analysis	81
6.1.1.1 Data Pre-processing	81
6.1.1.2 Biomimetic Model	81
6.1.1.3 Model Training and Testing	82
6.1.2 Results	83
6.2 Experiment 2: Pairwise Comparison of Positive and Negative Emotion with Neutral, Using Biomimetic Models	84
6.2.1 Experiment Setup	84
6.2.1.1 Data Input	84
6.2.1.2 Biomimetic Model	84
6.2.2 Results	85
6.2.2.1 Neutral versus Positive	85
6.2.2.2 Neutral versus Negative	86
6.2.3 Discussion	87
Chapter 7 Discussion	89
Chapter 8 Conclusions	94
List of Abbreviations	95
References	96
Appendix	109
A Support Vector Machines	109
A.I Hyperplane	109
A.II Classification using a Separating Hyperplane	109
A.III Maximal Margin Classifier	110
A.IV Support Vector Classifier	111
A.V Solving for the Support Vector Classifier	112
A.VI Kernel functions: Classification with Nonlinear Decision Boundaries	114
B ID3 Decision Trees	116
B.I Information Entropy	116
B.II Gini Index/Gini Impurity	116
B.III Decision Tree Algorithm	117

B.IV CART - Classification and Regression Tree	117
B.V ID3 - Iterative Dichotomiser 3	122
C Random Forests	124
D Artificial Neural Networks	124
D.I From Biological Neurons to Mathematical Model Neurons	124
D.II Learning in Artificial Neural Networks	127
D.III Error Function	128
D.IV Backpropagation for updating the weights	129
E Artificial Neural Network Architectures	131
E.I Convolutional Neural Networks (CNN)	131
F Canonical Problems in Machine Learning	133
G Classification Machine Learning Algorithms	135
G.I (Regularised) Logistic Regression	135
G.II Classification Trees (Ensembles)	135
G.III Deep Learning (Artificial Neural Networks)	135
G.IV Support Vector Machines (SVMs)	135
H Use of Machine Learning in Clinical Gait Analysis	136
I Ethics Application, Protocol and Consent Forms	139



# List of Figures

Figure 1: Schematic representation of the objectives and the proposed approach. The horizontal arrow represents the proposed bidirectional approach of improving machine perception of gait using human perception knowledge. The vertical arrow represents the corollary of understanding both human and machine perception better as a result of this approach .....	4
Figure 2: The different stages within the stance, swing phases of one gait cycle of a healthy adult, which includes one stride and two steps . The proportions of gait phases may vary in the case of motor dysfunctions .....	7
Figure 3: Chronophotography of human motions. Superimposition of several photographs of a man walking and running, late 19th century. ....	8
Figure 4: Vicon™ marker placement. Some Vicon™ configurations require very precise placement of the markers on the body .....	10
Figure 5: An illustration of the joint extraction process from the depth image obtained from the Kinect™ sensor. The picture shows the body, then their individual body areas and their joint representations highlighted as multi-coloured segments. This is the internal representation of the body used in extraction of the skeleton by the Kinect™ sensor. The output i.e. 3D joint proposals are used as data source for classification of gait.....	11
Figure 6: The names, positions and hierarchy of the joints represented in the skeleton extracted from the depth image of the Kinect™ sensor. There are a total of 25 joints that are extracted when including fingers.....	12
Figure 7: A representation of the ‘unfolding’ of the RNN network with each subsequent input. Given the characteristics of RNN networks, each unfold increases the prominence of the vanishing and exploding gradient problem.....	20
Figure 8: An illustration of the experiment setup for data collection, denoting the treadmill, walking subject, the 72 inch stimulus screen and the computer-controlled motion capture sensor. Sensor was kept at a vertical distance of 1 m and horizontal distance of 1.8 m from the center of the treadmill .....	31
Figure 9: Conversion of one skeletal pose into a frame in the point light animation to be used as visual stimulus for the psychophysics study.....	35
Figure 10: Point Light Displays (PLD) of a walking stimulus, labelled A to H, at eight different stages of a gait cycle. In psychophysics experiments, the eight different body poses shown here are an example of the sequence of frames shown in the animation. Left to right, top to bottom show each of the frames that are shown one after the other at a rate of 24 frames per second to create the illusion of motion. Here, the body poses labelled A to H are played in quick succession to give the illusion of walking motion .....	41
Figure 11: Implementation of the LSTM network architecture for processing gait sequences. Gait is provided as a temporal sequence of 3D body poses, to the LSTM and the final inner cell state of the LSTM is fully connected with the binary output layer.....	44
Figure 12: Correct gender classification accuracy (in %, mean and standard error) by the BM as a function of exposure duration in seconds. ....	46
Figure 13: Correct gender classification accuracy by NBM (in %, mean and standard error) as a function of duration of gait data (in seconds) used to generate the static representation. ....	50

Figure 14: Correct gender classification (% , mean and standard error) by LSTM models trained with position and velocity as a function of exposure duration in seconds.....	52
Figure 15: Correct gender classification accuracy (in % , mean and standard error) of NBM trained with position and velocity information, as a function of duration of gait considered for extracting the static representation.....	53
Figure 16: Comparison of correct gender classification accuracy (in % , mean and standard error) using biomimetic LSTM models trained in three- and two- dimensional position and velocity representations of the joint trajectories .....	56
Figure 17: Gender classification accuracy (in % , mean and standard error) of the BM when tested on veridical (upright skeleton) stimuli compared with (vertically) inverted stimuli. In both cases, the model is trained on veridical stimuli.....	64
Figure 18: Joint dependency tree of the human body representing the parent and child joints originating from the hip base .....	66
Figure 19: Point light display (PLD) of a walker with unit limb lengths at different stages of the gait cycle. This stimulus dataset is used for training and testing to evaluate for its dependence on structural cues of the walker. ....	66
Figure 20: Correct gender classification accuracy (in % , mean and standard error) of all the BM variations as a function of exposure duration (in seconds). The values are filtered through a one-dimensional Gaussian filter.....	70
Figure 21: Gender classification accuracy (in % , mean and standard error) in case of vertically inverted stimuli using BM trained with a variety of spatiotemporally modified gait. The x-axis represents the exposure duration of the stimuli for the model, in seconds .....	72
Figure 22: Gender classification accuracy (in % , mean and standard error) as a function of exposure duration (seconds) using BM trained using six varieties of spatiotemporal modifications on 2D gait.....	75
Figure 23: Gender classification accuracy of vertically inverted skeleton (in % , mean and standard error) using BM trained and tested on two-dimensional gait data (without depth information). Classification accuracy (%) as a function of exposure duration (seconds) of gait stimuli .....	76
Figure 24: Model architecture of the BM for emotion classification of human gait into ‘neutral’, ‘positive’ and ‘negative’ emotions.....	82
Figure 25: Concise representation of the main findings with reference to the stated research objectives (see Chapter 1). The horizontal arrow points to the improvements to machine perception of gait by leveraging human perception knowledge. The vertical arrow refers to the deeper understanding of human perception using insights gained from machine perception .....	93

# List of Tables

Table 1: Brief of the related literature, classification of gender from human gait.....	29
Table 2: Variations of extracted features used in different experiments, adhering to different model constraints .....	33
Table 3: Details of Data Collected, including their purpose and utilisation .....	33
Table 4: Description of the walking subjects taking part in the stimulus set. Both the HP and BM models are evaluated using this dataset. Age is displayed as median (mean $\pm$ standard deviation). .....	34
Table 5: Gender classification accuracy in % as a function of exposure duration of the stimulus .....	42
Table 6: Correct gender classification (in %) as a function of exposure duration of the stimulus (in terms of frames as well as seconds) with $p < 0.001$ (using two-tailed t-test compared with chance) for all the durations.....	47
Table 7: Correct gender classification accuracy (in %) by the NBM as a function of the duration of stimulus exposure (in number of frames). Statistical significance (p-value) is obtained using two-tailed t-test compared with chance (50%). .....	49
Table 8: Correct gender classification accuracy (in %) by NBM trained with position and velocity gait information.....	54
Table 9: Correct gender classification accuracy (in %) by the BM across all the stimulus exposure durations (in seconds).....	69
Table 10: Average Classification Inversion Probability (fraction of test stimulus where the gender classification is inverted on inversion of stimulus) of the different variations of the BM	71
Table 11: One-way ANOVA results on BM pairs (p-values). Null hypothesis assumes no significant difference in the means of the classification accuracies of the BM model pairs at 3.75 sec duration of exposure to vertically inverted stimuli.....	73
Table 12: Nomenclature of the various BM based on the spatiotemporal preprocessing steps used to for training and testing of the models .....	82
Table 13: Correct emotion classification accuracy (in %) by various BM trained with a specific spatiotemporal modifications made to the original gait dataset.....	83
Table 14: Nomenclature of the BM based on the temporal pre-processing of the dataset used for training and testing of the model .....	85
Table 15: Correct emotion classification accuracy of classification of ‘neutral’ and ‘positive’ emotion (in %) by various BM operated on temporally pre-processed human gait data .....	85
Table 16: Classification accuracy of the BM for positive-neutral emotions (in %) for stimuli exposure duration of 3.4 seconds. Chance accuracy at 50% .....	86
Table 17: Difference in the mean accuracies (as p-values of one-way ANOVA) of the variations of BM. Null hypothesis assumes no significant difference in their means. Chance performance at 50% .....	86
Table 18: Correct emotion classification accuracy of classification of ‘neutral’ and ‘negative’ emotion by various BM operated on temporally preprocessed human gait data.....	86

# Author's Declaration

This thesis was written by myself and represents my original work, with supervision from Dr. Adar Pelah. This thesis has not been submitted for any other award at this or any other institution. The work presented in this thesis has been presented, published or accepted for publication, as follows:

## Journal Articles:

1. Sarangi V, Pelah A, Hahn W, Barenholtz E., "Gender Perception from Gait: A comparison between biological, biomimetic and non-biomimetic learning paradigms" *Frontiers in Human Neuroscience - Motor Neuroscience* (Published in July 2020)
2. Sarangi V, Pelah A, Hahn W, Barenholtz E., "Biological and biomimetic perception: A comparative study of gender recognition from human gait." *Journal of Perceptual Imaging*, Vol. 1, Issue 3 (Published in February 2020)

## Conference Publications:

1. Sarangi V, Pelah A, Hahn W, Barenholtz E., "Biological and biomimetic perception: A comparative study through gender recognition from human gait." *Human Vision and Electronic Imaging 2020*. Society for Imaging Science and Technology, Burlingame, California USA, Jan. 2020
2. Sarangi V, Algahtani E, Kazakov D, Pelah A, "Explainable AI for Clinical Gait Analysis: Developing Diagnostics with Inductive Logic Programming." 29<sup>th</sup> ILP Conference, Bulgaria, Sept. 2019

## Poster Presentations:

1. Stone T, Sarangi V, Pelah A, Pelah A, "Application of Skeletal pose estimation from video clinical gait analysis in lower limb prosthetics in comparison to gold standard Vicon motion capture", *IET Conference on Human Motion Analysis for Healthcare Applications*, June 2019
2. Sarangi V, Pelah A, Hahn W, Barenholtz E, Stone T, Kazakov D, "Do machines "see" like us? A comparative study on classification of gender from gait between human, bio-inspired and non bio-inspired learning systems", *BMVA Conference on Visual Image Interpretation in Humans and Machines: Machines that see like us?*, April 2019
3. Sarangi V, Pelah A, Hahn W, Barenholtz E, Stone T, Kazakov D, "Clinical evaluation of machine learning approaches for the classification of gait using static & dynamic algorithms in comparison to human performance", *BMVA Conference on Deep Learning in 3 Dimensions*, Feb. 2019
4. Villiers E, Sarangi V, Wang N, Stone T, Pelah A, Shenker N, "Virtual Environment Rehabilitation for Patients with Motor Neglect Trial (VERMONT)" (To be presented)

# Chapter 1 Introduction

Human gait encodes a plethora of information such as one's gender [1–11], state of emotion [12–17], personal identity [18–20] and the state of health [21–24]. The extraction of relevant features from gait has found its applications in biometric security, crowd surveillance, medical diagnostics and physiotherapy rehabilitation progress monitoring. By far, visual observation by humans has been the most versatile, robust and widely studied medium for gait analysis.

Humans rely on visual perception of gait and utilise the neural mechanisms to make sense of it. The ability of the human brain to understand spatiotemporal patterns enables it to classify gait into a variety of categories. The categories vary from person identification, to a cohort of people or diagnosis of a particular clinical condition. Despite this ability, there are a few caveats to employing human observation for all gait analysis. Transferring acquired gait perception knowledge demands investment of resources for training new people. Examples of the need for such transfer, include clinical diagnosis or physiotherapy progress monitoring. Shortage of experienced professionals, perception errors due to fatigue, stochastic human errors and mortality further increase the challenge of depending solely on human observers as gait analysts. Thus encouraging the use of machines for automatic gait classification.

Automating some or all aspects of gait analysis using a computer based model would resolve the majority of these problems. In theory, computers can operate indefinitely with high fidelity, eradicating any problems that may arise due to fatigue and minimising stochastic errors. The transference of information between computers is inexpensive, compared to humans, leveraging highly reliable digital information copying techniques. Additionally, the ability of these computational machines to communicate with each other in real-time parallel, enables fast and practically seamless updates of their knowledge base at a rate much higher than human communication and learning. If designed appropriately, the computational model can also help increase our understanding of the model of human perception (HP) of gait. There are numerous ways of developing computational models to classify human gait. One approach is through precise modelling of human gait as a set of stochastic equations for normative and abnormal gait patterns and finding the best fit for an individual's gait. This approach involves explicit human expertise and intervention, deriving the biomechanical equations of gait. However, it's a challenging approach as the equations have to be general enough to model the varieties of gait, while being specific enough to model the subtle deviations from the norm to model specific abnormalities. For example, in the case of the medical condition of knee Osteoarthritis, the biomechanical equations that model the specific deviation from the norm because of knee pain is to be modelled, ensuring a significant separation from another type of medical condition affecting gait, such as Complex Regional Pain Syndrome (CRPS) of the lower limbs. Not only does this require a translation of knowledge from an expert clinician into computer understandable language, which is extremely challenging, but also limited by mathematical expertise of the language formulator, possibly rendering itself non-generalisable to other conditions due to formulator biases. The challenge and investment required in modelling all the characteristics that can be derived from gait deters the exploration away from expert system-

based solutions, where a human expert is required to formulate the rules for the computational model. In contradiction to the expert system type of models, is the data driven computational modelling of the biological approach of humans, where humans develop an internal representation of human gait by observing people walk without being told explicitly what biomechanical equations govern human gait. A hybrid approach that combines machine-learning abilities with expert domain knowledge exists, termed as Inductive Logic Programming (ILP) and can prove useful to explore in terms of gait analysis [25], however the exploration of that approach is beyond the scope of this work. The selection criteria for the evaluation of computational models are the models that can learn to classify different features from a generic and holistic representation of gait, emphasising on minimal human intervention for handcrafting objective specific features. Towards this objective, the focus is on utilising machine learning models.

Machine learning (ML) is a branch of artificial intelligence that provides computational algorithms, the ability to learn and improve from experience without being explicitly programmed by a human domain expert. ML develops learning mechanisms for a model and exposes it to representational examples of information to learn from. Essentially, it attempts to mimic the ‘learning from experience’ mechanism in humans. The various types of ML models are differentiated by their structure and the approach to learning. The nature of the model also determines the nature of information that can be processed and learned from. The underlying learning mechanisms of the ML models may either attempt to fit a hyperplane to best separate the data or model a possible nonlinear dependency between the input and output by closely resembling the biological neural network. For example, models such as support vector machines (SVMs) rely on static data for classification and regression, and depend on linear hyperplane separation for classification[26], whereas more complex models such as the recurrent artificial neural networks have the ability to process multi-dimensional dynamic temporal data and can fit a highly non-linear function to the training data for classification and regression[27]. On the other hand, decision trees (DTs) work akin to the process of deduction in logic, by using a measure of ‘purity’ to maximise the likelihood of belonging to a certain class [28]. The difference in learning principles make the models more or less suitable for operation on certain types of data.

The ML algorithms have benefitted from our deep understanding of biological mechanisms that have evolved to perform a certain task, such as the analysis of human gait. In recent literature, ML techniques such as recurrent neural networks (RNNs) [27] and convolutional neural networks (CNNs) [27,29] have been developed based on human neuronal models. RNNs attempt to emulate the memory capability of the human brain by maintaining a cell state throughout the course of operating on a dynamic signal. The CNNs attempt to emulate the human visual perceptual mechanism by mimicking the neural layers in the occipital lobe i.e. V1 through V5 [30]. Thus, making such models structurally biomimetic, i.e. mimicking or emulating biological systems in neuroscience. This is defined in detail later in the methods chapter (chapter 3).

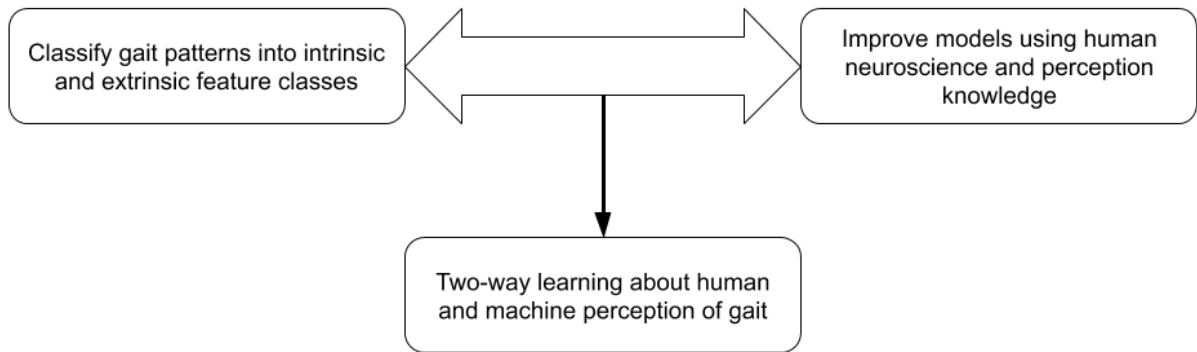
Given the ability of HP to learn to distinguish between gait patterns [10,12,13,18,21,32–34], one is encouraged to mimic HP and learning ability in machines when designing an artificial gait

classifier. However, the underlying biological learning mechanisms are not completely understood. For example, in case of gender classification from gait, studies have made contradictory claims of the perception being inherent in the brain, and being acquired through experiential learning [35–37]. The lack of understanding of the biological learning mechanism makes it difficult to establish the extent of reliability and common biases of the human observer. Thus, lacking concrete perceptual framework on which the computational model design can be based on. Consequently, one is also encouraged to gain an understanding of HP utilising the biomimetic machine learning models, in the course of their development.

While there has been considerable development of biomimetic models (BM), such as RNNs and CNNs, the research related to enhancement of the models for a specific purpose is still limited. In theory, one can attempt an infinite amount of computational enhancements before narrowing down to the ideal technique to improve the performance of the BM. The process relies heavily on intuition, thus making the process of tuning the hyper-parameters of an ML model, more art than science. On the other extreme, one can develop a model that tunes the parameters automatically to achieve the best accuracy. While hyper-parameter optimisation is a viable option of tuning the ML model itself [38], it does not provide insights about the pre-processing or feature extraction step in training the model. There needs to be a middle ground, where one does not have to explore potentially infinite options for improvement and can take solace in a systematic approach to improving one's models, while being able to understand the improvements made at a more human-understandable intuitive level. As mentioned previously, adhering to closely mimicking HP could provide a systematic approach to improving upon the model and thus be that middle ground. The hypothesis is proposed as follows, the biomimetic nature of the algorithms should be conducive to being improved upon by utilising knowledge of human neuroscience and perception. Understanding biological mechanisms and translating the insights into computational models should be able to enhance the models. Established human perceptual preferences, weaknesses and strengths could prove beneficial for application (computationally) to enhance its capabilities. Moreover, experimenting with the BM may provide deeper insights into the underlying biological phenomenon, increasing our understanding of human neuroscience and perception. This thesis aims to verify the validity of this paradigm, of a bidirectional learning approach, for the development of a computational model for perceiving a phenomenon that biology has evolved to be good at. In addition, the insights gained from HP inspired computational enhancements could help make a case for the underlying learning mechanism in humans. Fig. 1 visually represents the proposed paradigm.

In order to establish the validity of the proposed paradigm, human gait is used as the exemplar task. The two types of features of human gait that the models are evaluated on, (1) intrinsic, and (2) extrinsic. An intrinsic feature is defined as a feature of an individual that does not tend to change, such as one's gender. Conversely, an extrinsic feature has a tendency to change frequently, sometimes in a matter of minutes, such as one's emotional state. The reason for choosing the two types of classification features is to evaluate the generalisability of the classification ability of the artificial model, similar to humans. Successful classification of gait for a feature that tends not to change (gender) and that tends to change frequently (emotion) by

the same model would support its generalisability as an artificial classifier. The two types of features mentioned are detailed further in the Methods chapter (Chapter 3).



*Figure 1: Schematic representation of the objectives and the proposed approach. The horizontal arrow represents the proposed bidirectional approach of improving machine perception of gait using human perception knowledge. The vertical arrow represents the corollary of understanding both human and machine perception better as a result of this approach*

The thesis follows a chapter-based structure, where the chapters are presented sequentially to demonstrate the evolving applications from the insights gained from the previous chapters, as follows.

Chapter 1 provides the introduction to the topic of automated gait analysis and the relevance and need of incorporating human perception into the automation procedure. It mentions the motivation for developing an automatic gait classifier and the need for a systematic approach for selection and improvement of the classification model.

Chapter 2 is divided into three sections providing: (1) the theoretical framework of human gait analysis; (2) overview of machine learning (ML), and; (3) relevant literature pertaining to application of machine learning in gait analysis. The first section provides an overall brief of gait analysis, by introducing the concept of human locomotion, delving into the history of its analysis, building up to the current practices and instruments used to measure gait. The second section, provides a brief overview and definitions of the widely used biomimetic (BM) and non-biomimetic machine (NBM) learning models currently available. The descriptions are functional and relevant in the context of gait analysis. A detailed mathematical operation of the various models can be found in the Appendix. Finally, the third section, provides a summary of the existing literature and work related to the automation of gait analysis using machine learning algorithms, summarising the strengths and weaknesses of each approach.

Chapter 3 lays the methodology for the experiments. It details the capture sensors used, protocol for data collection, development of the ML models and statistical methods used for the analysis of the results.

Chapter 4 compares various ML models to human perception (HP) on the basis of gender classification from gait. The comparison is performed with the dual objective of determining the



biomimetic nature of the ML models on a functional level, as well as establishing baseline performance accuracies for the ML models, both for BM as well as NBM models.

Chapter 5 tests the BM model (resulting from the previous chapter) for presence of human-like bias in gender-classification, and determines the validity of the proposed paradigm in potentially alleviating the bias while improving the model's classification accuracy. The well-studied *inversion effect* in humans is taken as the exemplar bias to evaluate the model on. The presence of the effect in the model would not only further establish the biomimetic nature of the model, but also provide a plausible explanation for the presence of the effect in humans. The proposed bidirectional paradigm is utilised to modify the model and compare its classification accuracy with its initial unmodified version. A statistically significant reduction in the bias, along with an improvement in performance accuracy would establish the validity of the paradigm in successfully improving the model, thus providing an approach to improving the BM models. A statistically insignificant change or reduction in performance of the modified model would argue against the efficacy of such a systematic approach and warrant for further exploration. If successful, this could also provide evidence for the *inversion effect* being a learned phenomenon in humans, as opposed to being innate in the neuronal structure.

Chapter 6 evaluates the model on an extrinsic feature of human gait; the emotional state of the subject. The BM is evaluated for its ability to classify the feature in a matter of minutes, by inducing human emotions in subjects through visual stimulus, and attempting to classify the emotion from the gait pattern. The validity of the proposed bidirectional paradigm is evaluated, similar to the previous chapter, by modifying the BM. A statistically significant improvement in the performance accuracy between the unmodified BM and modified BM would further establish the validity of the proposed paradigm.

Chapter 7 collates the results of the experiments in the previous three chapters and discusses them with respect to the objectives and hypotheses set in the Introduction (Chapter 1) to evaluate the validity of the hypotheses and the limitations of the claims made. Additionally, further avenues of exploration are discussed as future work.

Chapter 8 provides a conclusion, with a holistic overview of the motivation for developing an automatic gait classifier, proposals for a solution, experiments to determine the validity of the proposed solutions and insights discovered along the way.

# Chapter 2 Background

Understanding the current practices in analysing human gait could provide a better context for the experiments presented in future chapters. Especially as it's been mostly done by human observers so far. This chapter briefs on three related aspects of human gait analysis. Firstly, the practices used in gait analysis by humans, from history to modern technological innovations in motion capture in section 2.1. Secondly, a brief functional introduction to the machine learning techniques used in this work and their justifications in section 2.2. Finally, the relevant literature in use of various machine learning techniques in automation of gait analysis, their advantages and limitations in section 2.3. The history of gait analysis, from utilising film-based snapshots to cutting edge marker-less motion capture devices, provides an appreciation for the role and advancement in technology in the analysis. Notably, the data used in the experiments in this work is captured using marker-less sensors. In addition to the above mentioned sections, this chapter also very briefly introduces the studies on understanding human perception (HP) of gait, its preferences and weaknesses. It may also be noted that in this work, gait data is used mainly as a means of evaluating the relationship between human perception and biomimetic and non-biomimetic machine learning. The work attempts to focus on the learning, rather than gait analysis itself. This chapter provides a background to gait analysis for better context into understanding the complexities involved in gait analysis as a practice.

## 2.1 Gait Analysis

### 2.1.1 Human Gait

Human gait is a complex acquired behaviour that requires little conscious thought by the walker in terms of limb movements, during routine activities [39]. The action of walking comprises two key stages: the activation of the nervous system and the stimulation of the musculoskeletal system [39]. Gait depends on the effective communication between the different aspects of the nervous system involved in the musculoskeletal signal generation, and any congenital or acquired problems in this complex process can result in gait disorders and gait abnormalities. The gait cycle can be broken down into two phases: the stance phase and the swing phase. As shown in Fig. 2, the stance is from the foot contacting the floor till the start of the swing phase when the toe leaves the ground. The swing phase, on the other hand, starts when the toe of the foot leaves contact with the ground and continues as long as the foot is in the air, above the ground. A stride comprises the portion of gait from the contact of one foot with the ground until the same foot is in contact with the ground again, while a step is defined as contact on the floor by one foot until there is contact by the other. Fig. 2 shows two steps and one full stride.

Gait analysis can be described as the study of human locomotion, especially from a biomechanical perspective, and has historically relied on the human observer's or clinician's ability to assess an individual's walk. Recent technological advancements, however, have allowed the development of tools which can be used to quantify and process the measurable parameters of human movement. The technologies also include the system used for the collection

of data from the walking subjects and in the analysis of the data. Gait analysis can be used in sports to improve a professional's performance and also, as is the subject of this research, during the rehabilitation of people with gait disorders.

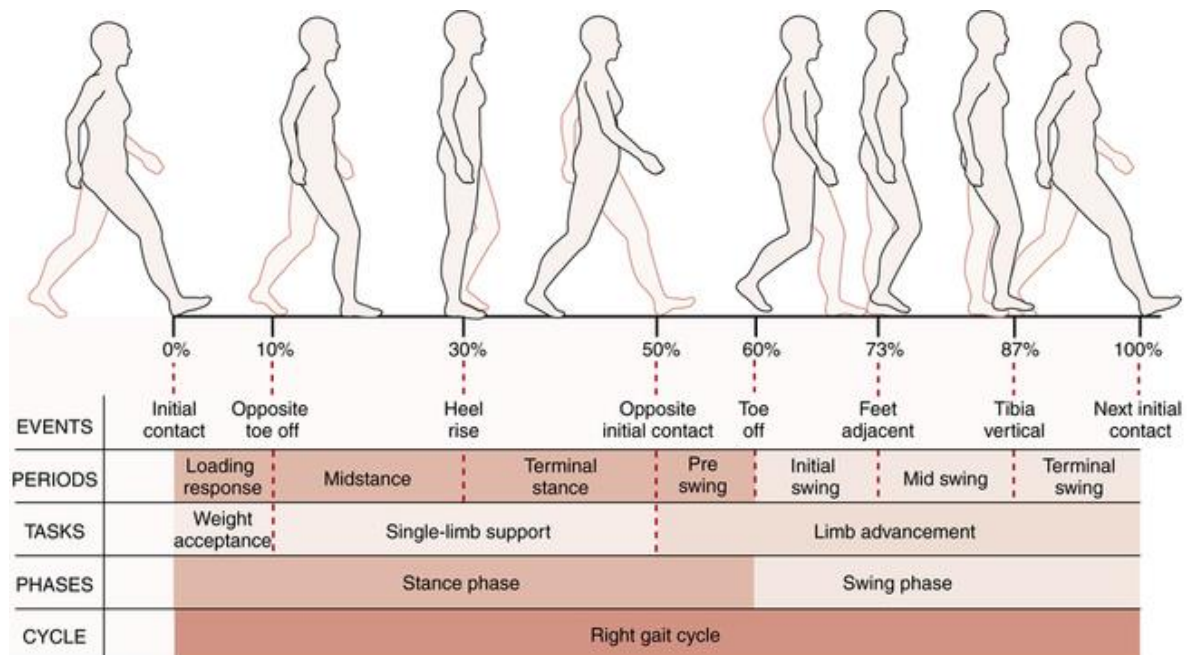
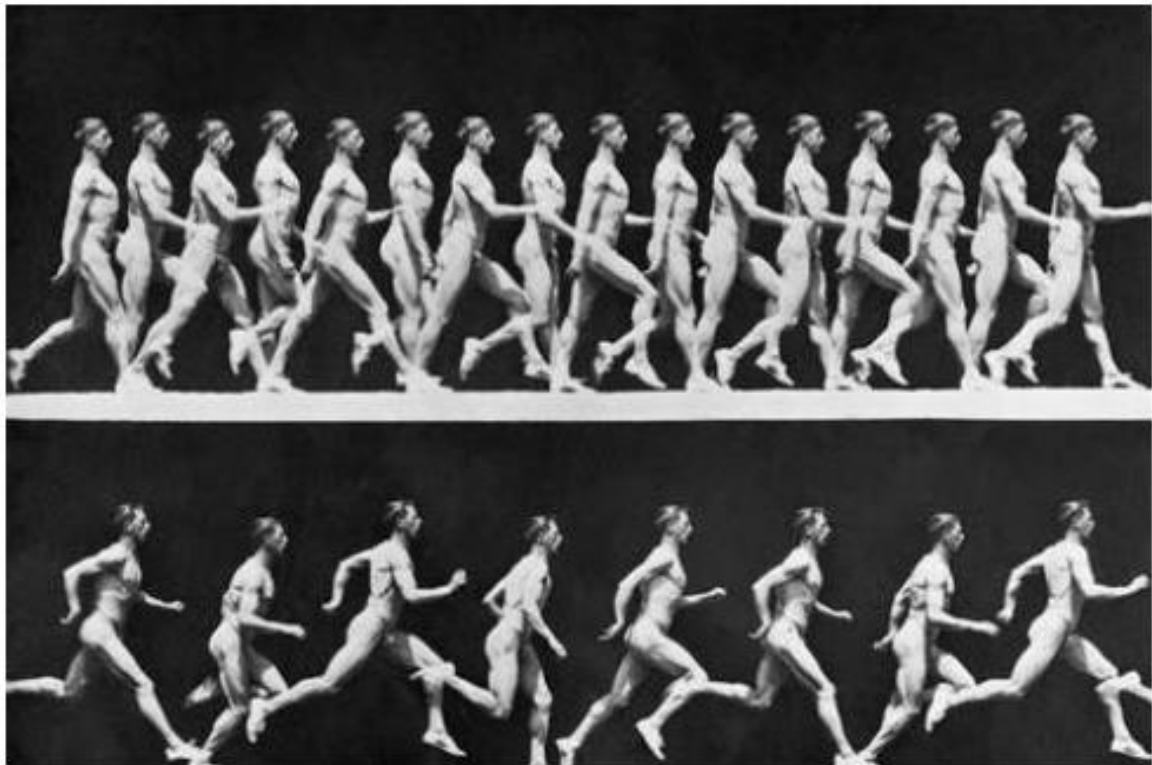


Figure 2: The different stages within the stance, swing phases of one gait cycle of a healthy adult, which includes one stride and two steps [40]. The proportions of gait phases may vary in the case of motor dysfunctions

### 2.1.2 History of Human Gait Analysis

Gait analysis has a long history [40], but previous investigations were in large part limited by human vision and judgement. Measurements were therefore not entirely objective. In the 17th century, Borelli calculated the centre of gravity of the human body and described the distinct gait cycles, and by the 19th century the first quantitative measurements were made using Borelli's description. The "stance" and "swing" phase of the gait were first used in these measurements by Wilhelm and Eduard Weber and continued as the foundation for later gait analysis approaches. Image analysis of gait started with a chronophotography, as can be seen in Fig. 3, allowing for sequential exposures of a runner to be recorded in a single photograph. The first 3D analysis was done in the 1890s in which a film was taken of people wearing light emitting markers, allowing for a more detailed evaluation of each phase of the gait. Although done manually, this technique presaged today's marker-based systems using light reflecting markers [40, 41]. Research in the middle decades of the 20th century added the measurement from gait of energy use, rotation of limbs, and external pressure on the limbs in certain scenarios [40].



*Figure 3: Chronophotography of human motions. Superimposition of several photographs of a man walking and running, late 19th century. [43]*

Current marker-based motion capture systems use specially designed cameras for capturing the light reflecting off of or being emitted from markers positioned on specific predetermined parts of the body. Technology plays a substantial part in current gait analysis practices, both in the capture gait data as well as in the decision making process. More recently, the use of depth cameras introduced a new way of capturing the human body without markers, which also happens to be the technology used for capturing data for experiments in the future chapters.

### 2.1.3 Gait Motion Capture Technology

#### 2.1.3.1 Marker based Motion Capture Systems

Although marker-based motion capture systems are not directly utilised in the experiments undertaken in this work, they are generally regarded as the gold-standard given their sub-millimetre accuracy in tracking the marker [42,43] and have been used for the calibration of marker-less motion capture systems. Marker-based motion capture systems are generally accompanied by a companion processing software with the capability of generating a skeletal representation of the body from the specific markers' positions. The positions are obtained by placing reflective or illuminating markers on the limbs and trunk of the body, which are tracked by a set of (typically) infra-red cameras. There's two versions of marker-based motion capture systems, namely, passive and active marker-based systems. Passive markers reflect the infrared light that is emitted by the cameras, and is captured back by the cameras themselves. Contrarily,

active marker-based systems have markers which emit infrared light which are captured by the camera. The active markers technique enables the system to have different frequencies of light emitted by different markers or having the emissions cycle through the different emitters, thus making it much easier to appoint joint labels to the different markers, but at an additional monetary cost. Passive markers enable the marker based motion capture system to be cheaper than its active marker counterpart, but increases complexity of calibration and assigning joint labels to the different markers. However, marker based motion capture systems in general tend to be more expensive than the alternatives mentioned in this section. The number of cameras, in both active and passive systems, depends on the desired accuracy and also on the size of the room, while the number of markers will depend on the desired accuracy for representing, say, a single joint. Modern systems can calculate the angles and velocities of different joints and as well as the forces on the limbs. Additional tools, like force measuring pressure plates, can also be used [43,44]. Accurate calibration is always required, usually with a special fiducial marker wand, and this can be a time consuming and ad hoc process in some systems.

There are several disadvantages to marker-based systems. Markers must be assigned to different joints in software, which can consume a lot of time, while the system loses accuracy over time unless the cameras are recalibrated frequently. Reflective marker systems are prone to noise in the light, which can only be counteracted by controlling the light level while measurements are taken. In addition, markers can sometimes move slightly with the skin or clothes, making the use of specially designed markers containing bodysuits often advisable. In spite of these shortcomings, the different possibilities of marker placement and high accuracy of marker-based systems qualifies them to be used in clinical environments. A few examples of marker-based systems are Vicon™ [42], OptiTrack™ [45] and Qualisys™ [46], both systems providing a very accurate representation and frequently used in medical environments and research (see Fig. 4), usually up to sub-millimetre accuracy in tracking the markers.

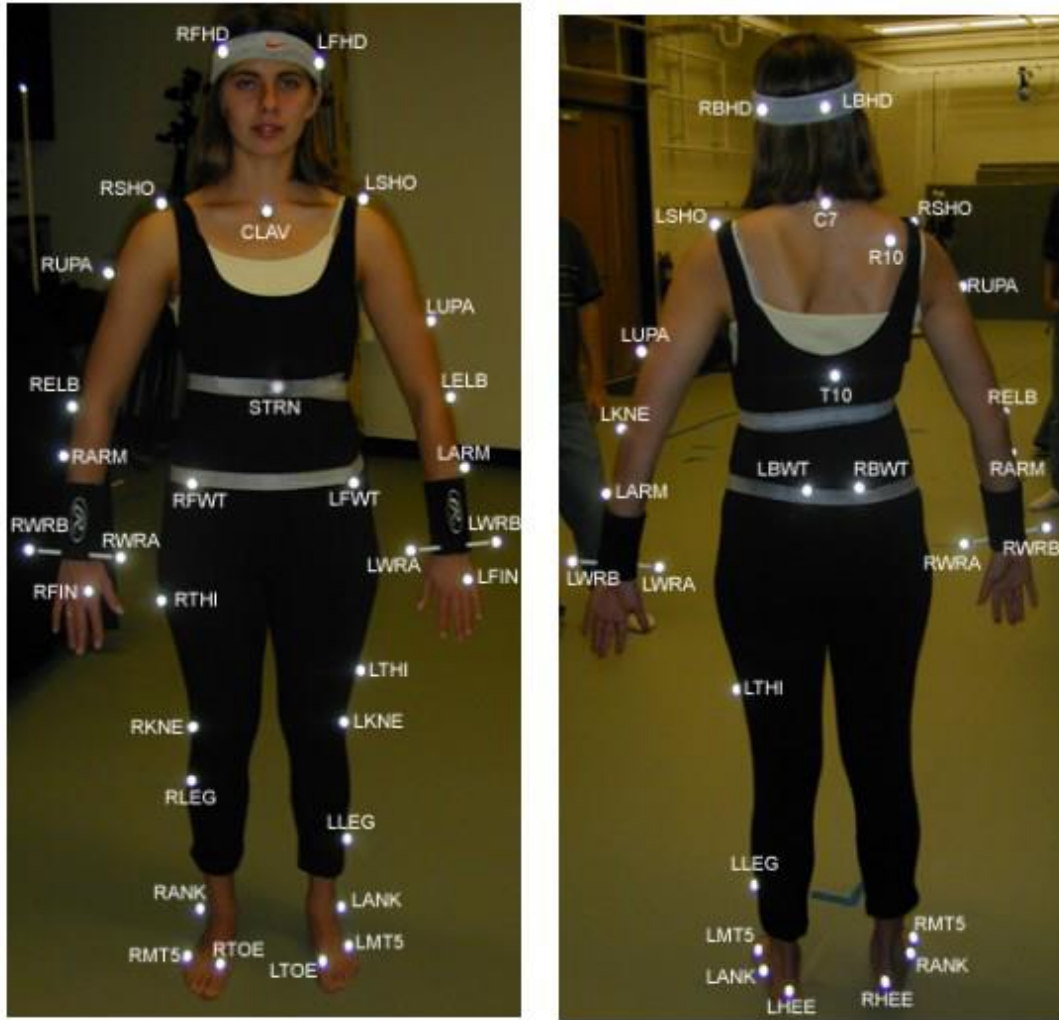


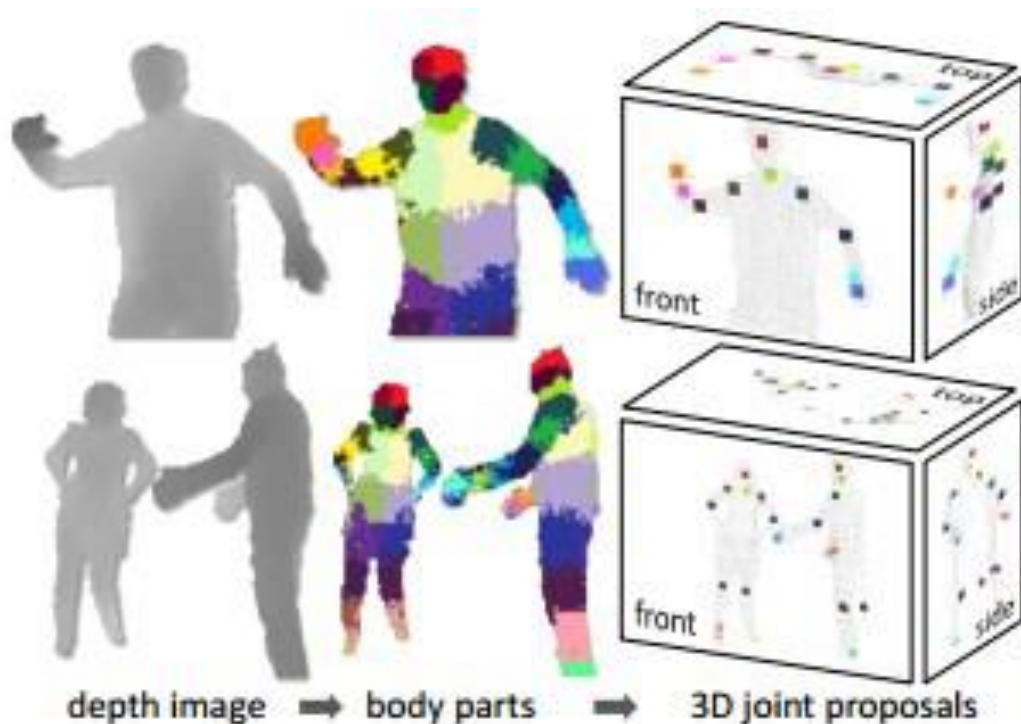
Figure 4: *Vicon™ marker placement. Some Vicon™ configurations require very precise placement of the markers on the body [46]*

### 2.1.3.2 Marker-less Motion Capture Systems

Marker-less motion tracking technologies are more convenient in comparison to the marker based systems. Their setup time is considerably less as they do not require any physical markers to be attached to the body at specific anatomical landmarks. These sensors have the capability mapping their surroundings in all three-dimensions, detecting and isolating human presence and understanding the anatomy of the human body to triangulate positions of body parts on them, using various computer vision algorithms. The Microsoft Kinect™ sensor is the first mass-market sensor developed for marker-less body tracking. The Kinect™ sensor was originally developed to introduce gesture controls into Microsoft Xbox™ Games, but with the built-in depth sensor and 30 fps frame rate was soon adopted for gait analysis [47]. The first version of the sensor used structured-light technology [48] to project and receive a unique pattern of light invisible to the human eye, with the emitter and receiver placed a known distance apart. The distortion caused in the pattern of light received, due to the depth of the reflecting surface in front of the sensor, helps understand the depth of the surfaces around the sensor. The second version of the sensor Kinect™ v2, used a time-of-flight sensor [48] to map the depth of reflecting

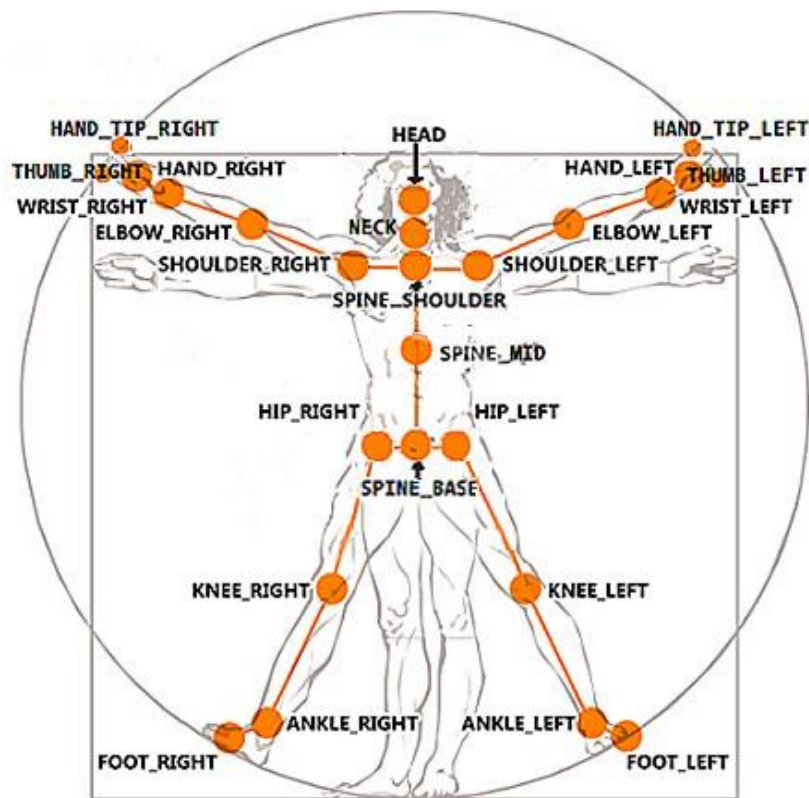
surfaces around the sensor, based on the time taken for the light to reflect back to the receiver. The sensor provides the depth using either structured-light or time-of-flight technology, which is further processed using computer vision algorithms provided as a software library [49] for detecting the presence and isolation of a human body. Successful detection of the human body further utilises the library's capabilities for mapping the limbs, trunk and head according to body position. The data from the Kinect™ can be processed and shown either as a raw feed of the video (RGB), depth (D) stream from the sensor or the processed skeleton extracted from the depth stream [50]. The processed skeleton is represented as X-Y-Z body joint positions of 25 salient joints of the human body, allowing a stick figure to be drawn. Either the RGB-D stream or/and the extracted skeleton can be used for gait analysis.

The extracted skeleton represents a tracked 3D body pose. Each body pose is a composition of various tracked joints arranged in 3D space and represented as x, y and z coordinates in the Euclidean space. These tracked 3D joints correspond to the centre of the joint in real-life, of the person the sensor is tracking. The accuracy and precision of tracking can be tested by examining the extent a joint was placed into the correct body part, and whether the position within the body part was correct over the whole dataset across multiple frames [51]. Notably, the tracking of 3D joints is usually performed internally by the software library that accompanies a marker less-motion tracking sensor. For example, Fig. 5 shows the body parts being used for the detection of the joint positions in the case of the Microsoft Kinect™ sensor, which is the sensor used in this work. Fig. 6 shows the Kinect™ v2 joint names and positions [50] of the tracked joints.



*Figure 5: An illustration of the joint extraction process from the depth image obtained from the Kinect™ sensor. The picture shows the body, then their individual body areas and their joint representations highlighted as multi-coloured segments. This is the internal representation of the body used in extraction of the skeleton by the Kinect™ sensor. The output i.e. 3D joint proposals are used as data source for classification of gait*





*Figure 6: The names, positions and hierarchy of the joints represented in the skeleton extracted from the depth image of the Kinect™ sensor. There are a total of 25 joints that are extracted when including fingers[52]*

The Kinect™ sensor focuses on upper body movement due to its mainly gaming objectives, which is a disadvantage for gait analysis in that the legs and feet are sometimes not as accurate and fluctuate. Nevertheless, the Kinect™ can still be used for gait capturing as it is accurate for the rest of the body, feet accuracy can be improved by post-analysis and averaging during continuous gait cycles improves estimates statistically.

Notably, subsequent experimentation relies on the skeleton pose information provided by the Kinect™ v2 sensor for representing gait. Gait is taken as a sequence of skeletal poses, each represented as a set of 3D positions of 20 salient joints of the human body. The details of the data is elaborated in Chapter 3 (the Methods chapter).

### 2.1.3.3 RGB based video capture

Video-based gait analysis is arguably the most traditional and still widely used form of motion capture. Although video (RGB)-based gait analysis is not presented, it is relevant in understanding the current practice and future of gait analysis. Gait capture through RGB video entails capturing the whole human body within the frame of exposure while the person walks in front of the camera. However, the difficulty in articulating the exact skeleton pose from a video makes it difficult to quantify gait from an RGB video. Moreover the existence of additional cues in video such as body shape, clothing, hair, face provided additional cues which may corrupt the



information that could have been extracted purely from gait. The existence of personally identifiable features such as face, makes it harder to anonymize the data, requiring a manual post-processing step to obscure personally identifiable properties in the interest of data protection and privacy. In addition, the analysis from video largely depends on the training and expertise of the analyser that tends to be an experienced human observer. Humans, while extremely versatile and accurate, are vulnerable to human errors. However, the recent advent of artificial intelligence and machine learning (detailed in Chapter 3) demonstrates potential of disrupting the field of gait analysis from two-dimensional RGB video, through the use of skeleton pose extraction technologies and specific gait analysis techniques.

#### 2.1.3.4 Inertial Measurement Unit (IMU)

While all the above optical methods restrict the physical area for capture of motion, IMUs allow for motion capture in a non-laboratory setting. Although IMUs are not used in this study, it is a promising extension for evaluation of the models in a future study. Inertial measurement units (IMU) are devices that can measure a body's specific force, angular motion and also the orientation of the body. They generally use a combination of accelerometers, gyroscopes and at times, magnetometers. The advent of wearables technology has popularized the use of IMUs for various applications such as fitness tracking (e.g. FitBit™ [52]), activity monitoring and for medical applications, such as event detection and rehabilitation. IMUs generally have a higher sampling rate compared to marker based or marker less systems, providing enough time-series data for analysis of motion with high precision [53]. IMUs also have the advantage of being portable, unlike the other motion capture technologies mentioned, with a 'plug-and-forget' nature of capturing motion data. IMU sensors embedded in smart phones, fitness trackers, smart watches and other proprietary hardware can be used to capture motion for an extended period of time without the wearer's conscious knowledge. The collection of continuous amounts of data helps in fine-tuning the analysis of objectives, such as gait. However, the data captured by the IMU sensors cannot be natively analysed by a human observer and requires extensive post processing and specialised training to be understood by humans. In the case of video, marker-based or marker-less motion capture systems, the motion captured can be visually relayed for analysis by a human observer in a way that is familiar to the observer. However, the signals obtained from IMU sensors cannot be directly understood by human observers, thus requiring additional intervention using computational methods for analysis.

#### 2.1.3.5 Electromyogram (EMG)

Electromyography (EMG) is a diagnostic technique that evaluates the condition of muscles and associated motor neurons. The EMG signal is a biomedical signal that measures electrical currents generated in muscles during its contraction representing neuromuscular activities. The nervous system always controls the muscle activity (contraction/relaxation). Hence, the EMG signal is a complicated signal, which is controlled by the nervous system and is dependent on the anatomical and physiological properties of muscles. The synergistic activation of muscle tissues of the whole body in different stages of a gait cycle generates action currents that flow through the resistive medium of the tissues [54]. The voltage gradients thus produced are read and

recorded as myoelectrical signals. This technique helps understand gait as a function of signals from the motor neurons and the resulting contractions and relaxation of the tissues involved in gait. Studies have demonstrated the correlations between the EMG signals recorded from specific muscles involved in gait (primarily in the lower limbs) and the specific corresponding components in the gait cycle [55–59]. The insights into the neurophysiological mechanisms provided by EMG has its advantages in clinical practice as well [40, 56, 58, 60–62]. Studies have applied the neurophysiological insights in developing intelligence in prosthetics and robotics [56, 63–66]. However, similar to the IMU signals, understanding the EMG signals requires explicit training and cannot be natively analysed by a human observer. Although EMGs are not used in the current study, it's a promising method of gait analysis, given its conduciveness in the creation of advanced limb prosthetics.

Motion capture is the first phase of gait analysis, leading up to the phase of understanding the motion data captured. Historically, the data captured was analysed by trained human observers to spot inconsistencies in gait with limited statistical analysis, relying mostly on visual acquisition of data and analysis using a mental model of the observed gait. However, advancements in computation technology have led to the implementation of artificial gait analysers which are executed directly on the captured motion data. The artificial gait analysers are used both for simplification of the motion data for easy human understanding as well as for end-to-end classification of the captured locomotion, based on the initial motion data. Both approaches i.e. humans and machines, have advantages and disadvantages, which are elaborated upon in the next section.

#### 2.1.4 Gait Analysis through Human Perception

Throughout history, gait analysis has largely been performed by human observation of gait and its perception. Thus, understanding the mechanisms of human perception (HP) of gait becomes imperative to understanding the strengths and weaknesses of the current practice. As it will be mentioned further in this section, humans are naturally good at perceiving gait, however, not without biases. Whether the biases are inherent or learned is not resolved yet and will be investigated briefly in future chapters using computational models. In 1975, Swedish psychophysicist, Johansson Gunnar first demonstrated that human observers are highly sensitive to the structure of biological motion [53, 67]. In his study, the observers were presented with highly simplified representations of human motion using 'point-light' stimuli consisting only of the dynamic locations of a small number of specific parts of the body of a walking person. This was achieved by attaching luminescent balls of light to the salient body joint locations and recording its movement using two-dimensional video cameras. Despite the sparsity of joints, people readily interpret these stimuli as representing human gait. Subsequent research with point-light walkers demonstrated that human observers could extract the identity of familiar people [18–20]. When the walkers are unfamiliar, people could extract certain biologically intrinsic features such as approximate age and gender. Human gait carries an abundance of information beyond such intrinsic features, including the emotional state [9, 10, 68] and the state of health of the individual [21].

Human observers possess the impressive ability to adapt to learning specific types of gait, for purposes ranging from personal identification to clinical evaluation. Given the objective of comparing HP with the computational models, it is useful to consider the capabilities of humans. On the other hand, they have been documented to possess certain biases in perception as well. Studies have noted that an overreliance of HP on specific anatomical features could lead to erroneous judgement in gender classification when presented with an altered visual stimuli [32, 35, 36, 69]. In one study, the same experiment was repeated with the same participants, but by inverting the skeletal pose of the walkers. Almost all of the observers changed their perception of gender to the opposite of what they had predicted earlier for the same walker [32]. Thus proving a lack of robustness due to expectations of anatomical configurations (upright skeleton) and over-simplification of perception. In addition, HP is highly susceptible to fatigue, errors in judgement and mortality. Misclassification of human gait, especially in a clinical setting could lead to commensurate loss in resources and delay in restoration of normal health. Moreover, developing clinical perception is an expensive and time-consuming process that cannot be transferred between humans easily. Thus stressing on the need for an alternate inexpensive perception model that can intelligently perceive the same characteristics from human gait, while being fast and impervious to fatigue and stochastic errors in judgement.

## 2.2 Machine Learning

Advancements in computer technology has provided the opportunity of programming an artificially intelligent gait analyser. The algorithms recruited for the intelligent analysis could be chosen from a vast pool of techniques, ranging from simple statistical tools such as first-order statistics with thresholds, to complex biomimetic algorithms such as perceptron-based artificial neural networks [70]. These algorithms can be used for both classification into categories, and regression of numerical metrics. Thus, the algorithms can be utilised to classify human gait into two or more categories, and also find the probability of it belonging to one or the other class. Broadly the algorithms can be segregated into two paradigms, depending on who does the learning (1) Expert systems, and (2) Machine learning systems.

Expert systems represent a set of algorithms that emulate the decision-making capability of a human expert. They are designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as if-then rules rather than through conventional procedural code. Essentially, the full knowledge of the human expert is digitised and used in decision making. An expert specifies all the steps taken to make the decision, the basis for doing the same and an approach to handle exceptions. A rigid system follows the exact rules from the expert. A flexible system uses the knowledge as an initial guide and uses the expert's guidance to learn, based on the feedback from the expert. In this approach, the human expert does the learning which is then transformed as a digital set of rules for the computer to perform its judgement. The set of algorithms are extremely helpful in the lack of sufficient data, as the training data requirement is minimal. Many projects start with expert systems to validate a new concept and progress towards machine learning approaches.

Machine learning (ML) systems represent a set of algorithms that computer systems use to perform a certain task without using explicit human defined instructions, relying on patterns and inference instead. They build a mathematical model based on sample data, known as “training data” to make judgements without being explicitly programmed to perform the task. Machine learning algorithms are used where it is difficult or infeasible to develop a conventional human generated set of rules for effective performance in the task given the limitations of HP. Additionally, with the availability of sufficient data, machine learning algorithms are employed to learn highly complex nonlinear patterns that are beyond the scope of HP. Although the algorithms do not require explicit instructions, human intervention is a necessary step in training during ‘supervised learning’.

Many people now interact with systems based on machine learning every day, for example in image recognition systems, such as those used on social media; voice recognition systems, used by virtual personal assistants; and recommender systems, such as those used by online retailers. As the field develops further, machine learning shows promise of supporting potentially transformative advances in a range of areas, and the social and economic opportunities which follow are significant. In healthcare, machine learning is developing systems that can help doctors give more accurate or effective diagnosis for certain conditions. In transport, it is supporting the development of autonomous vehicles, and helping to make existing transport networks more efficient. For public services it has the potential to target support more effectively to those in need, or to tailor services to users. And in data analytics, machine learning is helping to make sense of the vast amount of data available to researchers today, offering new insights into biology, physics, medicine, social sciences, and more [83]. From an implementation perspective, machine learning algorithms are a combination of techniques that have been developed so that a machine can understand the nuances in a data and learn to make decisions based on the given training examples. This branch of machine learning is termed as supervised learning [71]. Under supervised learning, a model is generally trained for classification or regression or both [71]. Given the objective of gait classification, the focus shall be on supervised classification task and the ML models that are used for the task. The focus shall be on the evaluation of various ML models and improving them. There are common models that can be used for classification as well as regression, however this study does not focus on them. An overview of the canonical problems in machine learning is presented in Appendix F.

ML is a vast field with active ongoing research, thus this section restricts the scope of ML to the most widely used techniques as of the time of writing this thesis and defines them in the context of the underlying learning mechanism used. The mathematical foundations of the techniques are detailed in the corresponding Appendices.

### 2.2.1 Classification

Classification is a supervised learning task for modelling and predicting categorical variables. Examples include classifying emotional sentiment, music genre, financial fraud or diagnosis of diseases in a person [82, 86, 87, 104, 159, 203]. Many regression algorithms have classification counterparts. The algorithms are usually adapted to predict a class (or class probabilities) instead of real numbers. In this case, multiple ML models will be evaluated on the basis of classification

of gait into two or more categories. As discussed in later chapters, the classification outcomes in case of chapters 4 and 5 is the binary gender (male or female), while in case of chapter 6, it's the state of emotion (between positive, negative or neutral). Different ML models have different approaches to communicating the output. To establish consistency amongst the models, a classification is considered as the category chosen either with the highest probability of outcome, or, the category remained after elimination of all the other candidate categories.

### 2.2.2 Support Vector Machines

Support Vector Machines (SVM) were first introduced by Vapnik in the early 90s [72]. They've found applications in a wide variety of areas. As binary classifiers, SVMs have been used in gait classification [11], face recognition [73], speech recognition [74] and text categorisation [75]. As regressors they've been used in control systems and communications [76], amongst others. In case of classification, SVMs construct a decision rule to classify vectors to one of two (or more) classes based on a training set of vectors whose classification is known a-priori. SVMs do this by implicitly mapping the training data into a higher dimensional feature space. A hyperplane (decision dividing surface) is then constructed in this feature space that bisects the two categories and maximises the margin of separation between itself and those points lying nearest to it (also known as the support vectors). This hyperplane or decision surface can then be used as a basis for classifying vectors of unknown classification. SVMs have the advantage of (1) implementing a form of structural risk minimisation - Essentially attempt to find a compromise between the minimisation of empirical risk and the prevention of overfitting[76], (2) attempting to solve a convex quadratic programming problem. Thus, lacking a non-global minima to get stuck in making the problem readily solvable in a single-shot (as opposed to iterative solving in neural networks) using quadratic programming techniques, and (3) specifying the final trained classifier completely in terms of its support vectors and the chosen kernel function type.

They also have a corresponding regression-based alternative known as the support vector regressors (SVRs), which are a set of non-linear regressors inspired by Vapnik's support vector method for pattern classification. Similar to Vapnik's method, SVRs first map all the data into a higher dimensional feature space using some kernel function (usually). In this higher dimensional feature space they attempt to construct a linear function of position that mimics the relationship between the input and output observed. In other words, they try to find the non-linear map between the points in the higher dimensional feature space and the class they belong to. SVRs are simpler than their corresponding competing methods such as unregularised least-squares [77]. Details of the mathematical formulation of the SVMs are provided in Appendix A.

### 2.2.3 Decision Trees (ID3)

Decision trees (DT) are a popular tool in machine learning and statistics for classification tasks (both classification and regression). For example, the skeleton builder in the marker-less motion capture sensor i.e. the Kinect™ v2 uses decision trees to identify the different joints of the human body in the depth image [78]. The learning mechanism (also known as training phase) in a DT involves two steps, (1) learning a hierarchical, tree-structured partitioning of the input

space, and (2) learning to predict the label within each node. DTs are used extensively in data mining because they

- are computationally fast to train and test
- are well-suited for datasets with mixed attribute types i.e. binary, categorical, numerical
- test with good accuracy and are interpretable

DTs are a different paradigm of machine learning models from support vector machines. They are extremely effective, especially on categorical data. Similar to SVMs, decision trees are applicable as classifiers as well as regressors. However unlike the SVMs, decision trees are white-box classifiers. In other words, the set of rules developed during the training phase of decision trees can be explicitly laid out. While DTs are powerful, they are prone to overfitting on the training dataset and require additional heuristics to provide bounds to their complexity, such as pruning or limiting maximum depth. Intuitively they can be thought of to have a bias-variance trade-off. Where increasing depth of the tree captures increasing complexity within the training dataset. Deep decision trees have low bias but high variance. A decision tree algorithm trained on two different datasets can produce a result with high variance, hence DTs are sometimes termed as unstable learners.

The basic objective in training DTs is to minimise the impurity or entropy of the training set. Intuitively, DTs attempt to minimise ambiguity in classification of a data point into a particular category. A highly impure dataset would possess a high ambiguity of classification. Impurity is generally measured as a quantifiable measure such as (1) Information Entropy, or (2) Gini Index/Gini impurity [79]. Details of the mathematical formulation of the model, its training and inference is provided in Appendix B with a worked- out example for further clarification.

#### 2.2.4 Artificial Neural Networks

An artificial neural network (ANN) is a massively parallel network of interconnected digital processing elements (also known as neurons). The parallel nature of the network enables it to achieve the computation capacity deemed practically impossible for sequential computer models, which includes all the other machine learning models mentioned previously. The function of a single neuron in an ANN is simple, and is often taken as the functional weighted sum of the inputs, passed through a non-linear activation function followed by either a sequence of similar layers of neurons or further processed into an output layer for results. The biological neurons in the brain communicate by means of spike trains (or impulse trains). ANNs are designed to emulate this characteristic which enables them to learn. The perceptron model of the neuron that is implemented in the ANN captures the biological mechanism by encoding the spike rates as the output of the neuron. The ‘weights’ of connection between the neurons represent the strength of the bonding at synapse (represented numerically) or more intuitively the amount of contribution of the output of one neuron to the next. The study and application of artificial neural networks has the dual purpose of using our knowledge about the brain to develop effective and efficient solutions to the problem of human body motion analysis, as well as shedding some light on the possible biological neural mechanisms of the brain that are involved in visual motion analysis.

As mentioned in the previous section, support vector machines operate by fitting a linear hyperplane onto a dataset either in its original form or its projection through a nonlinear transformation function, while decision trees (or random forests [122]) inherently introduce nonlinearity through the minimisation of dataset impurity. However, the key advantage of artificial neural networks is the complexity of nonlinearity that can be mapped by the networks due to its compounded nonlinearity effect of multiple layers. The rich dynamical properties exhibited by some architectures have suggested applications in signal processing, robotics, and control amongst various others [159, 174, 177, 179]. The learning ability in ANNs is inspired by biological systems. The contribution of each neuron varies slowly with time according to the given training set, thus accounting for a stable long term contribution to the final output by each neuron. The ability to learn patterns and map the input stimulus to the output expected has leveraged the application of ANN in medical diagnosis, time series analysis and classification [74, 78, 82, 85, 86, 87, 101-104, 159, 179]. Gait, simply put, can be represented as a time series of multidimensional data, where each dimension represents the temporal evolution of the trajectory of a particular joint moving in three-dimensional space. Certain architectures of ANNs are very well suited for analysing such time series information to encode the spatiotemporal patterns encoded in the data. Similar to SVMs and DTs, ANNs do not require explicit specification of rules of the data to learn the patterns. The way they model a system is by considering the input-output examples of the process (called training examples/sets) and learning a possible nonlinear mapping between the inputs and outputs. The arrangement of neurons in the network is known as the *network architecture*, and it decides the type of data that can be best analysed by the network and its capacity for learning. The complexity of the nonlinear mapping function that can be learned by the network is generally defined by the number of layers in the network and the number of neurons per layer. As the number of layers and neurons increase the learning gets deeper, crediting the network with the term ‘deep learning’ network[80].

#### 2.2.4.1 Recurrent Neural Network (RNN)

The neural network architectures mentioned previously treat every input data point as an independent, self-contained vector with no relation to its previous to next input. However, biological brains are constantly active and correlate new data points with previous data points to make a longitudinal pattern. Humans don’t think from scratch every second. This is most evident during visual perception [81]. Similarly, gait is visually perceived by humans, because of the persistence of memory as the visual stimuli changes through a sequence of human body poses. Had there been no memory (or persistence of thought), each body pose would be treated independent of the others, leaving no room for the concept of gait. Traditional neural networks such as Multi-Layer Perceptron based Artificial Neural Networks (MLPs) [80] and Convolutional Neural Networks (CNNs) [30] suffer from this issue, which the RNN aims to address. RNNs emulate the biological feature of ‘memory’ or ‘persistence of thought’ through the use of looping back mechanisms in the architecture. In theory, RNNs can make use of information in arbitrarily long sequences. A simple RNN architecture is presented in Fig. 7. The unfolding of the RNNs with each time step inherently classifies this architecture as a deep neural network.

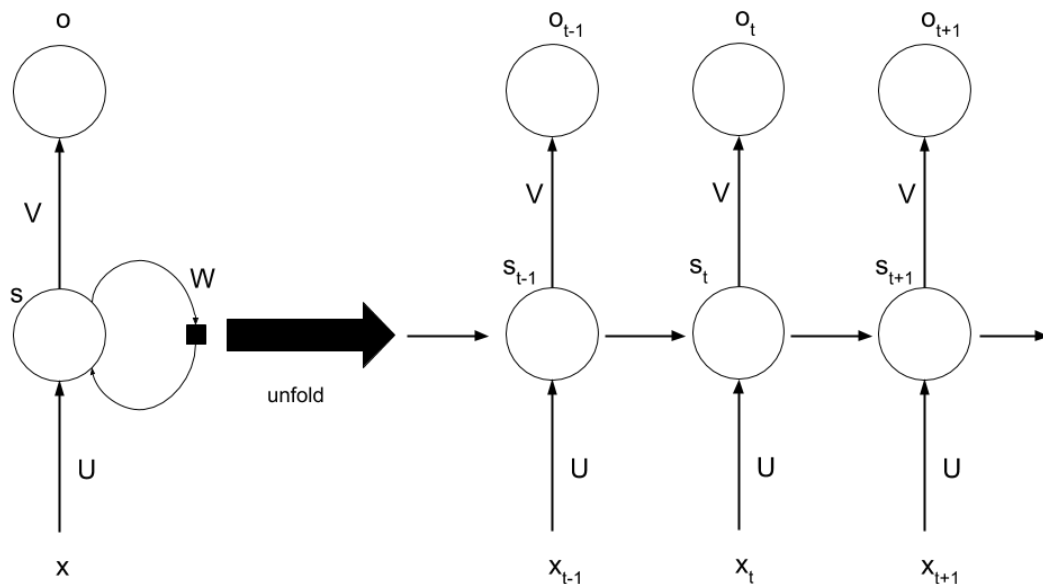


Figure 7: A representation of the ‘unfolding’ of the RNN network with each subsequent input. Given the characteristics of RNN networks, each unfold increases the prominence of the vanishing and exploding gradient problem

Similar to the biological brain, the RNNs achieve this by considering the existing state of the network along with the new input at every step of data point processing. The above figure shows the unfolding (or unrolling) of the network with every subsequent data point. The network essentially combines both the new input it receives as well as its old state (the collection of all its parameters) into a single input vector to process. The formulas that govern the computations are,

- $x_t$  is the input at time step  $t$ . For example,  $x_1$  could be a one-hot vector representation corresponding to the second word of a sentence
- $s_t$  is the hidden state at time step  $t$ . It’s the “memory” of the network.  $s_t$  is calculated based on the previous hidden state and the input at the current step  $s_t = f(U \cdot x_t + W \cdot s_{t-1})$ . The function  $f$  usually is a nonlinearity such as  $\tanh$  or  $ReLU$ .  $s_{-1}$  which is required to calculate the first hidden state is typically either initialised to zero or randomly initialised
- $o_t$  is the output at step  $t$ . For example, if we wanted to predict the next word in a sentence, it would be a factor of various probabilities spanning across the language vocabulary.  

$$o_t = \text{softmax}(V \cdot s_t)$$

Intuitively,

- $s_t$  can be thought of as the memory of the network. It captures the information about the processing of the previous data points in previous time steps. The output,  $o_t$  is calculated solely based on the memory at time  $t$ . However, in practice the function of the RNN depends highly on the implementation. A simple RNN cannot capture information from too many time steps ago. Specific modifications can be made to the simple RNN to improve its long-term temporal pattern capabilities (as mentioned in following sections).
- Unlike simpler and more traditional deep neural networks such as MLP and to an extent CNNs, which use different unshared parameters at each layer/step, the RNN shares the same tuneable parameters across all the ‘unfolded’ layers (as shown in



$U, V$  and  $W$  above). This reflects the self-learning and self-tuning capabilities of the biological human brain and the fact that every time step essentially affects the model as a whole including everything learned up until that time step. The advantage of this approach lies in the greatly reduced number of parameters to be learned.

- Depending on the application, the output at each time step may or may not be essential. Application might focus only on the final output of the RNN and learn the required weights from the final output state.

If we had to draw an analogy with human perception, recurrent neural networks (RNNs) can be thought of as multi-layer perceptron (MLP) networks, with memory added to them. For example, if ANNs are analogous to human perception, MLPs are similar to a memory-less learners. An MLP based learner would be able to learn the correlations between an input and an expected output, but would not remember the previous input provided to it and cannot correlate any previous inputs provided to the current output expected. It can only correlate the immediate input and output. Contrary to this, an RNN based learner would not only be able to correlate the current input, but also all the previous inputs provided to the current output. Essentially, it can correlate data through time to an expected output. RNNs are thus used to find temporal patterns to classify a signal/data sequence.

#### 2.2.4.2 The Vanishing and Exploding Gradients Problem in RNNs

Understanding temporal patterns requires the model (whether biological or machine learning based) to persist the temporal information in time as the relevant information may either be spread across time or have different pockets of relevance in different times. This kind of dependence between sequence data is called long-term dependencies given the distance between the relevant information and the point where it is needed to make a classification. Unfortunately, practically, as distance becomes wider, RNNs lose most of the information making it extremely hard to learn the dependencies. They encounter either a ‘vanishing’ or ‘exploding’ gradient problem. The problem magnifies when the relevant bits of information are spread out at greater distances, especially when the more relevant bits of information towards the beginning of the data sequence, requiring the gradients to be propagated back in time all the way to the initial layer. The gradients being propagated through the deeper layers of the network undergo a sequence of matrix multiplications because of the chain rule. The results of the multiplications shrink exponentially, if they have small values (less than 1) until they vanish, making it impossible for the model to learn from them, leading to the *vanishing gradient problem*. On the other hand, if the values are large (greater than 1), they get larger and eventually increase exponentially to blow up and reach the limits of the highest number that could be processed by the computer, resulting in the crashing of the machine learning model, leading to the *exploding gradient problem*.

The problem of exploding gradients can be dealt with to an extent by clipping the gradients to a predefined threshold, to prevent it from becoming too large. This process doesn’t change the direction of the gradients, but only affects its magnitude. The vanishing gradient problem can be resolved to an extent with the identity RNN architecture [82]. Where the network weights are

initialised to the identity matrix and the activation function are all set to Rectified Linear Units (ReLUs) [160]. The resulting network encourages computations to stay close to the identity function. The effectiveness of this stems from the fact that when the gradients are propagated backwards through time, they remain constants of either a 0 or 1, hence aren't likely to suffer from vanishing gradients. An even more popular and widely used solution is the Long Short-Term Memory architecture (LSTM) [84]; a variant of the regular recurrent network which was designed to make it easy to capture long-term dependencies in sequence data. The standard RNN operates in such a way that the hidden state activation are influenced by the other local activations closest to them, which corresponds to a “short-term memory”, while the network weights are influenced by the computations that take place over entire long sequences, which corresponds to a “long-term memory”. Hence the RNN was redesigned so that it has an activation state that can also act like weights and preserve information over long distances, hence the name “Long Short-Term Memory”.

Extending the human analogy established at the end of the previous section (read 2.2.4.1) the vanishing and exploding gradient problem can be thought of as a human learner without long term memory, who attempts to remember everything or forgets everything too soon. Exploding gradients can be thought of as a learner who tried to remember everything and thus saturates the whole system, reaching the capacity of remembering soon. Vanishing gradients can be thought of as a learner who forgets things too soon and is myopic in their remembrance of previous inputs, essentially making them similar to MLPs but with more tuneable parameters which are not utilised.

#### 2.2.4.3 Long Short Term Memory (LSTM)

The Long Short Term Memory cell/unit is an improvement upon the standard RNN and features heavily in this work. Similar to the RNN, it receives its input from the current time step input  $x_t$  and from the previous time step ‘hidden’ state of the network  $s_{t-1}$  [84]. The main architectural changes are,

- introduction of a memory cell state  $c_t$
- introduction of three sigmoid gates- forget gate  $\sigma_f(t)$ , update gate  $\sigma_u(t)$  and the output gate  $\sigma_o(t)$
- the ability to remove or add information to the memory cell state

The cell state and gates are learnable i.e. the LSTM learns the relevant temporal patterns in the training phase. Thus, it also learns the differences between different temporal patterns that lead to a certain classification, allowing it to identify subtle changes in patterns, which may be indiscernible from the aggregated gait metrics. This architecture is inspired from the memory and learning capabilities of the biological brain model as well. While the RNN introduced the concept of memory or thought persistence in the neural network, the problem of quickly forgetting everything, thus not being able to learn anything (also known as the vanishing gradient problem) or trying to remember everything leading to a crash of the model (known as exploding gradient) became apparent. The introduction of the memory cell state is similar to the concept of long term memory hypothesised in biological brains [84], to which information can be written and read from based on how relevant it is. The three various gates are used for regulation of information flow to and from the cell state, having parallels in relevance of information and

biological attention to the information and being receptive to new information. Thus LSTM aims to emulate an intelligent learning system that is capable of extracting relevant patterns across long time spans of information while being receptive and open enough to change itself while learning from new patterns. The LSTM cell has featured heavily in recent research about classification of time-series based data, such as speech recognition [85], financial price prediction [86], music genre classification [87] and language translation [87,88] to name a few. Representation of human gait as a multidimensional time series signal allows it to be used as input data for the LSTM cell. The biomimetic emulation of memory and the underlying functional emulation of basic human biology makes it an excellent candidate for a biomimetic model (BM) to analyse gait. While there are other computational models that mimic biology more closely [89,90], there is a trade-off between the complexity of the model, leading to closer biomimicry and the practicality of training the model to be used as a classifier. Generally, an increase in complexity demands an increase in the size of training dataset and a higher computational budget for training and testing. There are other BM that have been proposed in literature are mentioned in section 4.2.1. However, the number of tuneable parameters in those models are orders of magnitude higher than the ones in the LSTM, requiring a training dataset and computational power that is considerably higher than what is available in a typical high-end computational device. Thus, given the practicality of the problem to be addressed, the more complex models are not evaluated here.

Extending the human analogy mentioned in the previous section (read 2.2.4.2), LSTMs provide the advantage of deciding whether to remember a certain sequence or not, based on their relevance to the final expected output. LSTM based learners decouple their memory (cell state) from provided inputs and output using gates. The gates are responsible for learning whether to let a certain sequence of inputs affect the memory, based on their relevance to improving classification. This allows the learner to learn extended temporal patterns without forgetting the past inputs or reaching capacity trying to remember everything. Thus overcoming the problems found in RNNs.

## 2.2.5 Model Comparison

Models, like the support vector machines (SVM) [77] learn to linearly separate the training input features using a hyperplane to find the best class for a new test input feature set. In case of a non-linearly separable dataset, SVMs utilise the non-linear projection functions, also known as kernels [77, 91, 92] to derive linearly separable projections of the input. While the SVMs are more suited to classifying numerical data, Decision Trees (DT) [28] are more versatile and can classify categorical data as well. DTs view all inputs as categories and learn to use the most relevant features to reduce ambiguity in class, while developing cumulative branches of rules to minimise ambiguity. The inner nodes of the DT represent binary rule sets, while the leaf nodes represent the classes. The appropriate combination of input features and hyper-parameters of the ML models can lead to optimal classification performance. Performance here is defined as the percentage of predicted classes matching the actual classes in the test set. However, in case of SVMs and DTs, the extraction of relevant features from multi-dimensional temporal signals like human gait requires human intervention and domain expertise, requiring a reinvention of features

for a new gait classification objective. Biomimetic algorithms, of which artificial neural networks (ANN) are prominent [93], are inspired by the intricate complex connections of billions of neurons in the human brain and attempt to mimic the transference and modification of information as it flows through layers of neurons. ANNs make it possible for the learning models to learn highly non-linear relationships between the inputs and the output objectives, improving classification performance dramatically. Deep learning (DL) models-based on ANNs take it a step further and eliminate the need for any human intervention by designing specialised ANN models, to learn the features directly from the raw input and utilise the learned features in classification. Thus generating an end-to-end solution for any classification objective. DL models have also been shown to outperform all other ML models in classification. Memory-based DL models, like the Recurrent Neural Networks and Long Short Term Memory (LSTM)-based networks have the additional ability to understand sequential data [27,84] while learning the optimum features for classification, which makes them highly suitable for processing multi-dimensional temporal data like human gait. Memoryless DL models like the deep feedforward neural network (FFNN) can learn features from static input directly and have been shown to outperform SVMs, DTs and shallow ANNs [80,94,95]. Interestingly, the process of learning the requisite features from input and utilising the features learned in classification is similar to how HP operates as well. ANNs mimic a memoryless analytical brain, while LSTMs mimic an analytical brain with both short and long term memory. SVMs and DTs represent purely statistical learning models with no correspondence with the human brain. We classify ANNs and LSTMs as biomimetic, and SVMs and DTs as non-biomimetic learning models.

The objective of developing an artificial gait analyser favours the use of machine learning approaches as opposed to expert system-based techniques, given the capability of the former set of algorithms to learn highly complex nonlinear relationships for maximising performance. However, human observers have shown themselves to be accurate and versatile gait analysers, thus the incentive to adhere to human-like learning and inference mechanisms.

### 2.2.6 Backpropagation and Genetic Algorithms

The training phase of the ANNs (including RNNs, LSTMs) improves the ANNs' ability for classification (or regression) by iteratively adjusting its hyper-parameters, such that, the output of the ANN closely matches the desired label of corresponding input in the training dataset. The adjustments to the ANNs are made either using (1) backpropagation (BP), or (2) genetic algorithms (GA). BP works by calculating the error deltas of the weights to be adjusted to minimise the error value in each iteration, by back-propagating the deltas to change the weights [70]. GA use a fitness function [70, 96] to determine the best performing models from the population and cross-breed them from the mating pool, usually with some crossover and mutation. Studies have compared both BP and GA on numerous parameters including time taken for training, size of training dataset, ease-of-use, outcome accuracy of classification and regression [97–100].

Studies have employed GA to successfully classify human body movements, especially for clinical diagnosis [101–104]. The ease-of-use, low requirement of training dataset, explainability

and robustness to over-fitting have helped classify clinical conditions such as Parkinson's [22,101–104].

BP can be thought of being analogous to 'learning from mistakes' in human perception. BP was originally proposed as a learning- mechanism for multi-layer perceptron networks [93]. Although it is generally agreed that BP cannot directly be implemented in real neurons, there is implicit recognition that neural mechanisms must somehow implement them [105]. Studies have reported the success of using BP in models of human learning at a molar level [105–107], while other studies have emphasized its failures [108–110]. The parallels between BP and human behavioural learning makes it the learning process of choice for our experiments that compare HP and BM models. Thus, for the experiments that utilise BM models, a variation of BP referred to as 'backpropagation through time' (BPTT) [111] is used to back-propagate the errors not only in space but also in time for the BP models. Appendix D.IV provides the mathematical details of the implementation of backpropagation for the ANNs.

## 2.3 Related Work

Gait classification is a well-researched application of gait analysis [9, 10, 18, 32, 67,112–115]. The ability to perceptually organise point-light displays (PLD) into the percept of a specific human action has long served as a demonstration that humans are adept at recognising the actions of their conspecifics. In 1973, Swedish Psychologist Gunnar Johansson published his results on experiments performed on motion perception of a multitude of directions in 3D space [67,112]. His experiments demonstrated the inferred depth and direction of motion from 2D point light animations of joints (without the need for the form aspect of the whole body) of a walking person from walking durations of a quarter of a second, establishing the powerful visual perceptual capabilities of the human mind and its reliance on optical flow to deduce motion. Subsequently, a meta-analysis of experiments was conducted examining gender classification capabilities of humans from point-light displays, obtaining 66% correct from side view and 71% for other views, with an optimal classification of ~79% correct[9,10]. To obtain estimates of perception efficiency, observers were presented with multiple repetitions of knocking, waving and lifting movements and evaluated the observations based on gender and affect classification of the model [10,116]. Efficiency was expressed as a squared ratio of human sensitivity to artificial neural network (ANN) sensitivity. The gender classification showed a proportion correct of 0.51 and an efficiency of 0.27%. A nuanced analysis of accuracy of gender classification was performed with confidence level of the observer of the PLD, showing ~85% correct at the highest confidence level[115]. Another study limited the temporal parameter of the duration of the stimulus to a minimum of 1.6 sec and a maximum of 2.7 sec of dynamic PLD to fully inform the observer of the gender of the walking model [32]. The conclusions from the above work informed the limits and variations of the study in this thesis in terms of human capabilities in gait analysis. A lack of available literature on gender classification efficiency from gait with human observers was also noted.

### 2.3.1 Automation of Gait Classification

Automation of gait classification has been an ongoing research topic for more than two decades, with gender classification being the most common objective for evaluation of the proposed classification techniques. There has been an increase in the number of applications of machine learning (ML) and deep learning (DL) models in the classification of gender from gait towards the latter part of the decade. Support vector machines, decision trees and feedforward artificial neural networks being some of the most common classification models, with support vector machines being used most frequently [1–5, 11,117–119]. While the performance outcomes of the models are significantly higher than chance and supersede human performance in most cases, there is a marked dissimilarity between the way HP and ML models perceive gait. As mentioned in the Introduction, performance here is defined as the percentage of predicted classes matching the actual classes in the test set. Humans have the ability to visually observe the human body, isolate the skeletal dynamics and compare the motion to established notion of gait. However, the machine learning models (SVM, DT, ANN) operate on static data points. To conform to the input requirements of the models, human gait needs to be represented as a static set of numerical values. Multiple studies have explored the static representation in conjunction with different classification models with encouraging results. A brief of the related literature to classification gender from human gait is presented in Table 1.

Most studies have utilised canonical static descriptors of gait, used for clinical applications; spatiotemporal gait metrics such as stride length, joint angles and displacements, gait cycle time etc. as features for gait classification. The main research in such studies is focused mainly on evaluating the machine learning models for classification accuracy. In the case of the intrinsic feature of gait, a relatively smaller number of studies have proposed novel features to be extracted from gait such as, Gait Energy Images (GEI)[118] and Gait Energy Volume (GEV) representations of gait [119]. The novel features have shown the most promise in gender classification from gait. There has also been a shift of motion capture technology from two-dimensional RGB sensors to more sophisticated 3D motion capture technology, especially the marker-less technologies in recent literature [1, 2,119]. The shift in motion capture technology has led to dramatic improvements in classification technology, especially in gender classification from skeletal motion. However, the study of the potential of gait analysis using 3D sensors is at an early stage compared to the maturity of the 2D sensors.

2D or 3D silhouette-based gait videos (such as GEI and GEV) are often model-based to fit a *stick-figure* to focus on the limb motion trajectories. For all purposes, model-free silhouettes are not considered as they provide additional external cues for easy classification of gender (hair, clothes etc), which contaminates the generalisability of the ML model to conditions other than what it's been trained for. Lee and Grimson employ features from dynamic human silhouettes, by dividing the silhouette into seven regions and an ellipse is fit to each region [120]. The centroid, the aspect ratio of the major and minor axes of the ellipse, and the orientation of the major axis of the ellipse are taken as features. The method achieves a correct classification rate (CCR) of 84.5% in a database of 24 subjects (10 males and 14 females). Similarly, ellipse based features were used, but combined with multi-view features to improve performance [5]. Another

study extracted static features as a sequence set of 2D stick figures and utilised an SVM and ANN as classifiers of choice [11]. The CCR is 95% for SVM and 92% for the ANN for 100 subjects (84 males and 16 females). 26 hand-crafted features including cycle time, mean and standard deviations of joint angles, their displacements and their corresponding time derivatives were provided as inputs to the ANN. Different from these methods, a new method was proposed which recognises gender from static full body images, using Histogram of Oriented Gradients (HOG) as feature and Adaboost [121] and Random Forest Algorithms[122] as classifiers [4]. A CCR of 75% was achieved for 888 images. Notably, the concept of Gait Energy Image (GEI) was introduced [118] as a static representation of the dynamics of gait for individual person identification. Utilising this technique, CCR of ~88% was obtained in gender classification [3]. As GEI is model-free, the most important features in classification are hair style and chest, which are external cues and not inherent in the dynamic of gait itself.

The launch of affordable RGB-D cameras made it possible to extract the 3D depth map of a walking person, without any contact with the person's body, making the sensor suitable for biometric and clinical gait analysis. An extended version of the GEI to 3D using binary voxel volumes analogous to the 2D silhouettes, known as Gait Energy Volume (GEV) was proposed as a feature [119]. The study spatially aligned the voxels corresponding to the lower body averaged over a gait cycle. It was demonstrated that using 3D depth-based features along with traditional 2D silhouette features led to a 6% rise in accuracy performance in gender classification from gait [2]. The study of the potential of gait analysis using 3D sensors is at an early stage compared to the maturity of 2D sensors.

Although gender classification from gait has been studied extensively, most studies focus on human perception and machine perception in isolation, leading to a lack of literature encompassing HP, BM and NBM perceptions in the same study and evaluated using the same dataset. Among the notable works comparing HP and ML, Davis and Gao [1] use PLDs in an automatic method. The method proposed uses an adaptive three-mode Principal Component Analysis (PCA) to extract features from PLDs, the CCR in a 40 subjects database (20 males and 20 females) is 95%. They also recruited 15 observers to recognise gender by visually observing the PLDs with a CCR of 69%. The exceptionally high performance of the automatic method could be attributed to the human intervention and the explicit hand-crafting of the static features (through PCA), which the human observers lack, leading to a bias in the perception of gender from gait between the learning models. While the learning models have been studied separately and evaluated for their performance at their best, studies exploring their learning efficiencies, in terms of the amount of information required to reach a certain level of performance has not been published to the best of our knowledge.

In terms of the extrinsic feature of gait, the ability of human observers to understand emotion from gait is well studied. In 1890, James et al. believed that walking alongside another person and mimicking or even observing their gait would reveal what the walker is feeling[123,124]. In the 1900s, the idea of being able to ascertain one's character from their style of walking was further solidified [9, 10, 17, 68, 124–128]. Human gait conveys affects similarly to voice or facial expressions [123]. Understanding emotional affect is an active area of study [17,126–128]

and human observers have been able to identify emotion from biological motion [9, 10, 68, 128]. The ability of the observers in being able to extract the relevant features for emotion, has been studied in gestures such as knocking, waving, lifting [9,10], dancing[68] and gait [12,13,17,33,34]. Ability of human observers to extract emotion from gait was pioneered by Montepare et al. in 1987 [34] by utilising human observers to categorise neck-down videos of walkers into emotion categories of happiness, sadness, anger and pride. While most research on expression of emotions has focused on the human face as the medium of expression, there has been a rising interest in understanding the nuances of human gait and its correlation with emotion expressed by the walker [37, 68, 116, 128–130]. Examples of important features for determination of emotion can be divided into two categories, (1) Postural features, such as head inclination and elbow flexion, and (2) movement kinematics, such as velocity and acceleration (jerk of limbs), especially arm movements [116,131]. Thus, emotion classification acts as an ideal testing criteria for the ability of the BM to learn the relevant features for classification of a short-lived gait feature, providing support for its application in tasks such as clinical diagnosis and crowd sentiment monitoring.

The use of static representations as features mentioned above for use in support vector machines, random forests, decision trees and multilayer perceptron neural networks, reduces the extensive dynamics of skeletal motion for a particular objective, making it less generalisable for objectives other than the one intended for the feature extraction. In general, a lack of research establishing parallels between human and artificial gait classifiers was observed, as of the time of conducting this research. Humans are still the best learners known and there is much to learn given the efficiency and accuracy of human learning and observation.



<b>Gait Data Source Type</b>	<b>Extracted Feature</b>	<b>Size of Dataset</b>	<b>Classification Model Type</b>	<b>Performance Output</b>	<b>Authors</b>
2D silhouette	Ellipse fit body regions	24 (10 M, 14 F)	Calculated body part ratios	85%	Lee and Grimson [120]
2D stick figure	Shoulder focus	20 (7 M, 13 F)	Human observers	76%	Saunders et. al. [5]
2D stick figure	Motion signatures	100 (84 M, 16 F)	SVM	95%	Yoo et. al. [11]
2D stick figure	Motion signatures	100 (84 M, 16 F)	ANN	92%	Yoo et. al. [11]
Full body images	Raw Images	888 (600 M, 288 F)	AdaBoost	72%	Cao et. al. [4]
Full body images	Raw Images	888 (600 M, 288 F)	Random Forest	73%	Cao et. al. [4]
Full body images	Edge Map	888 (600 M, 288 F)	Random Forest	73%	Cao et. al. [4]
Full body images	HOG	888 (600 M, 288 F)	Random Forest	73%	Cao et. al. [4]
Full body images	Part Based Gender Recognition (PBGR)	888 (600 M, 288 F)	PBGR	76%	Cao et. al. [4]
2D silhouette	Static images	124 (93 M, 31 F)	Human Observers	95%	Yu et. al. [3]
2D silhouette	Static images	124 (93 M, 31 F)	SVM	88%	Yu et. al. [3]
Point Light Displays	Motion trajectories	40 (20 M, 20 F)	PCA	95%	Davis and Gao [1]
Point Light Displays (PLD)	PLD animation	40 (20 M, 20 F)	Human Observer	69%	Davis and Gao [1]

*Table 1: Brief of the related literature, classification of gender from human gait*

## Chapter 3 Methods

With the aim of comparing human perception (HP) of gait with machine perception (through the use of computational models), the models are divided into two categories, namely, biomimetic (BM) and non-biomimetic (NBM). As mentioned previously, the biomimetic category of ML models includes models that aim to emulate or mimic biological systems in neuroscience. Notably, there is a difference in being structurally and functionally biomimetic. The definition of biomimetic with respect to the categorisation of models is limited to being structurally biomimetic. Examples of this category include all the perceptron-based artificial neural networks. The extent to which it's functionally biomimetic shall be evaluated in the next chapter. On the other hand, non-biomimetic models (NBM) include the models which are purely statistical in nature and do not attempt to emulate any biological or neurological behaviour. Examples of this category include hyperplane based classifiers (such as SVMs) or information entropy based classifiers (such as DTs).

The experimental configuration for comparing HP, BM and NBM learning paradigms, can broadly be divided into:

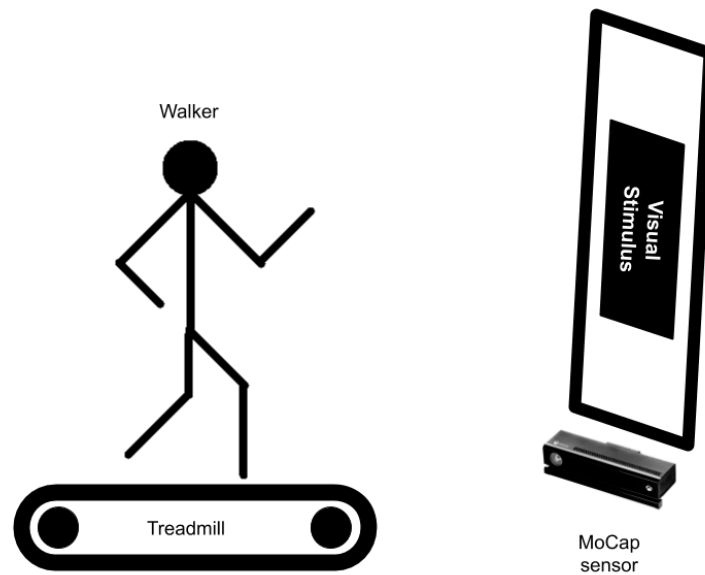
- Data collection- hardware and software setup for motion capture of walkers
- Stimuli development- developing visual stimulus for the psychophysical experiments, data processing and feature extraction for evaluating the BM and NBM
- Model development- developing the BM and NBM with specific hyper-parameters using open source libraries
- Data analysis- the statistical tests used to address the hypothesis and research questions

This chapter provides the common practices used across all the chapters. The result of data collection described in this chapter is utilised throughout the thesis. It is important to note that this chapter does not provide the specific implementation details, but rather a context of the experiments undertaken. The specific changes made to the data and models in each experiment are detailed in the bespoke experiments' chapters themselves.

### 3.1 Data Collection

The hardware setup of the system consists of three parts:

- A treadmill with side railings to hold on to for support
- A motion capture sensor for capturing the biomechanics of the walking person
- A 50 inch LED screen mounted vertically (portrait mode) in front of the treadmill for visually inducing a particular emotion in the walker



*Figure 8: An illustration of the experiment setup for data collection, denoting the treadmill, walking subject, the 72 inch stimulus screen and the computer-controlled motion capture sensor. Sensor was kept at a vertical distance of 1 m and horizontal distance of 1.8 m from the center of the treadmill*

Fig. 8 shows the hardware setup for capturing the walkers' data. The setup was utilised for data collection of both the intrinsic and extrinsic feature types of the experiments. As described in the Introduction chapter, gender is taken as an exemplar intrinsic feature while emotion is taken as an exemplar extrinsic feature. The Microsoft Kinect™ v2 [50] sensor was taken as the motion capture sensor of choice for all the data captures, given its accuracy in tracking the skeleton, portability and affordability.

The Kinect™ camera features both a standard RGB camera, capable of capturing at a resolution of 1920x1080 at 30 frames per second, and a Time-of-Flight depth camera which is capable of capturing the depth at a resolution of 512x424 at 24 frames per second. Notably, for further experimentation, only the depth camera was considered for extraction of the skeletal pose. The RGB camera was not utilised. To measure depth, the Kinect™ emits infrared light and uses its Time-of-Flight camera to measure the time it takes for the reflected pulse to be received, using this data to calculate the depth of the subject [132]. According to the Kinect™'s datasheet, the optimum distance between the sensor and the user is approximately 1.8 m [133]. Thus the distance between the sensor and the midpoint of the treadmill was kept at 1.8 m. The walkers were suggested to remain at the center as far as possible. Both the depth and the RGB cameras run at 30 frames per second and the field of view is 43° vertically and 57° horizontally, with a vertical tilt range of  $\pm 27^\circ$ . This range of coverage is more than sufficient for treadmill-based applications, although the available space between the camera and the user will depend on the experiment space's layout. It is important to note that the input received from the RGB cameras was not utilised in any experiments. The skeleton tracking information extracted from the depth camera was the only raw source of information for the experiments. The sensor is equipped with a time-of-flight depth camera which is used to create a 3-dimensional depth map. The Kinect™ for Windows v2 library utilises the depth map to extract the skeletal joint coordinates as a 3D offset (in millimetres) from the position of the sensor (taken as origin) [134]. When compared

with the state-of-art optical motion tracking methods (such as Vicon™ [42]), the anatomical landmarks from the Kinect™ generated point clouds can be measured with high test-retest reliability, and the differences in the inter-class coefficient correlation between Kinect™ and Vicon™ are <0.16 [135,136]. Essentially, the 3D spatial coordinates of the body joint centres calculated by both the Kinect™ system and the state-of-art Vicon™ system are highly correlated. Further works showed that both systems can effectively capture >90% variance in full-body segment movements during exergaming [137]. The validity of biological motion captured using the Kinect™ v2 sensor is established with human observers through reflexive attentional orientation and extraction of emotional information from the upright and inverted PLD [138].

It is important to note that all further experimentation utilises the above mentioned skeletal pose representation derived from the depth camera of the Kinect™ v2 sensor to represent gait. Gait is represented a sequence of skeletal poses of the human subject, captured at 24 frames (poses) per second. Each skeletal pose is described as a set of 3D Cartesian coordinates of 20 salient joints of the human body. Notably, the RGB camera video and raw depth information from the sensor was not used in the experiments. Although gait was captured as a temporal evolution of the 3D joint coordinates, certain experiments required further feature extraction to adhere to the constraints of the ML models. Specifically in the case of NBM models, which operate on static data. The following table (Table 1) briefs the features extracted from raw skeletal gait data, the details of which are presented in the corresponding experiment chapters

<b>Type of Feature</b>	<b>ML Model Type(s)</b>	<b>Classification Objective(s)</b>	<b>Sections (Chapters)</b>
3D Joint Positions of Veridical Skeleton	LSTM	Gender Classification, Emotion Classification	4.2, 4.5 (4), 5.1 (5), 6.1, 6.2 (6)
3D Joint Velocities of Veridical Skeleton	LSTM	Gender Classification, Emotion Classification	4.4, 4.5 (4), 5.4 (5), 6.1, 6.2 (6)
3D Joint Accelerations of Veridical Skeleton	LSTM	Gender Classification, Emotion Classification	5.4 (5), 6.1, 6.2 (6)
3D Joint Positions of Modified Skeleton	LSTM	Gender Classification, Emotion Classification	5.2, 5.4 (5), 6.1 (6)
3D Joint Velocities of Modified Skeleton	LSTM	Gender Classification, Emotion Classification	5.4 (5), 6.1 (6)
3D Joint Accelerations of Modified Skeleton	LSTM	Gender Classification, Emotion Classification	5.4 (5), 6.1 (6)
2D Joint Positions of Veridical Skeleton	LSTM	Gender Classification	4.5 (4)

2D Joint Velocities of Veridical Skeleton	LSTM	Gender Classification	4.5 (4)
First Order Statistics	SVM, RDF	Gender Classification	4.3 (4)
Clinical Gait Metrics	SVM, RDF	Gender Classification	4.3 (4)

*Table 2: Variations of extracted features used in different experiments, adhering to different model constraints*

Human gait data was collected in two phases, in the UK, at the University of York and in the US, at Florida Atlantic University. In the case of UK, data was collected at the LIVE Lab (Room S008) in the Department of Electronic Engineering including various Bachelors and Masters of Engineering students, including myself. In the case of US, data was collected in the Psychology Department, by students of Prof. Terrence Barnhardt; David Bickham and Belen Wertheimer. The protocol of data collection in the case of US was designed by me, while in the case of UK, it was a group effort by the students of the LIVE Lab. In both countries, I setup the infrastructure (including software), provided training to the students who would perform majority of the data collection, collected the first few subjects with the students to demonstrate the procedure and monitored the sanctity of collected data in the subsequent subjects. In the case of US, the monitoring of data collection was performed through online cloud synchronisation of the captured data, which I could access remotely. The data collected in both UK and US were combined and utilised for the purposes of gender classification (intrinsic feature), whereas only the data collected in the US was utilised for classification of emotion (extrinsic feature). The following table (Table 2)

<b>Site of Data Collection</b>	<b>Number of Subjects in the Dataset</b>	<b>Classification Objective for the Dataset</b>	<b>Chapter(s) which Utilise the Dataset</b>
University of York, UK	41 (26 M, 15 F)	Gender Classification	4, 5
Florida Atlantic University, US	22 (6 M, 16 F)	Emotion Classification	6

*Table 3: Details of Data Collected, including their purpose and utilisation*

### 3.1.1 Intrinsic Feature: Gender

Forty one consenting healthy adults (26 male, 15 female) between the ages of 18 and 50 years old were recorded walking on the treadmill. Participants volunteered and received credit towards a participation grade for their class. Written consent was obtained from all the participants taking part in the data collection process. Gait data was recorded as spatiotemporal three-dimensional joint trajectories for 20 tracked joints of the body. The tracked points on the walker's skeleton included the head, neck, shoulders, elbows, wrists, fingertips, mid spine, back, hips, knees, hips, ankles and toes. The collection of the joint positions formed a static frame. Data was captured at

24 frames per second, each frame represented by 60 numbers (3D coordinates of 20 joints) and a corresponding timestamp of capture of the frame. Data was recorded for 6 sessions per participant. Each session consisted of a minute of walking on the treadmill at a self-selected speed followed by a minute's rest. The joints were extracted utilising the Kinect™ for Windows v2 library provided by Microsoft. The details of the analysis of the data and its use in evaluating the biomimetic and non-BM is provided in Chapter 6. Table 1 provides the aggregated attributes of the walkers in this data collection process. This dataset is utilised in experiments conducted in Chapter 4 and 5.

	Height (cm)	Weight (kg)	Age (years)
Male	176.23 ± 32.43	80.49 ± 2.86	26 (26.06 ± 6.42)
Female	128.56 ± 23.51	73.3 ± 4.59	21 ( 21.29 ± 1.23)

*Table 4: Description of the walking subjects taking part in the stimulus set. Both the HP and BM models are evaluated using this dataset. Age is displayed as median (mean ± standard deviation).*

### 3.1.2 Extrinsic Feature: Emotion

Twenty-two subjects (6 males, 16 females) consenting healthy adults participated in the study. All the participants were undergraduate students who received credit towards a participation grade for their class. Appropriate consent forms were signed and anonymity maintained. The subjects walked on a treadmill set to a self-selected speed, approximately 1.5 meters from a 55-inch Samsung Plasma Display screen mounted in landscape orientation on a TV mount, as shown in Fig. 23. Subjects walked for approximately 25 minutes on the treadmill, without any rest in between. During the treadmill walk, subjects were asked to focus on the screen ahead and were shown a sequence of photographs to induce certain emotions. Each subject was shown coloured photographs designed to induce, positive, negative and neutral emotions. The visual stimulus size was 43 degrees wide and 24 degrees long for the whole photograph, without any zero (black) padding. A degree here is defined as the subtended angle at the nodal point of the eye. The sequence of showing the photographs was randomised for each trial and the gap between showing the visual stimulus for invoking different emotions was about 5 minutes of walking without any stimulus. The photographs were provided by The Center for the Study of Emotion and Attention, University of Florida, through IAPS (International Affective Picture System) database [139]. The IAPS provides normative ratings of emotion (pleasure, arousal, dominance) for a set of colour photographs that provide a set of normative emotional stimuli for experimental investigations of emotion and attention, solely for use in academic, not-for-profit research at recognised degree-granting educational institutions. Gait information was captured during the subject's exposure to each of three emotions inducing visual stimuli, separately. Similar to experiments in previous chapters, gait was recorded as spatiotemporal three-dimensional joint trajectories for 20 tracked joints of the body. The tracked points on the walker's skeleton included the head, neck, shoulders, elbows, wrists, fingertips, mid spine, back, hips, knees, hips, ankles and toes. The collection of the joint positions formed a static frame.

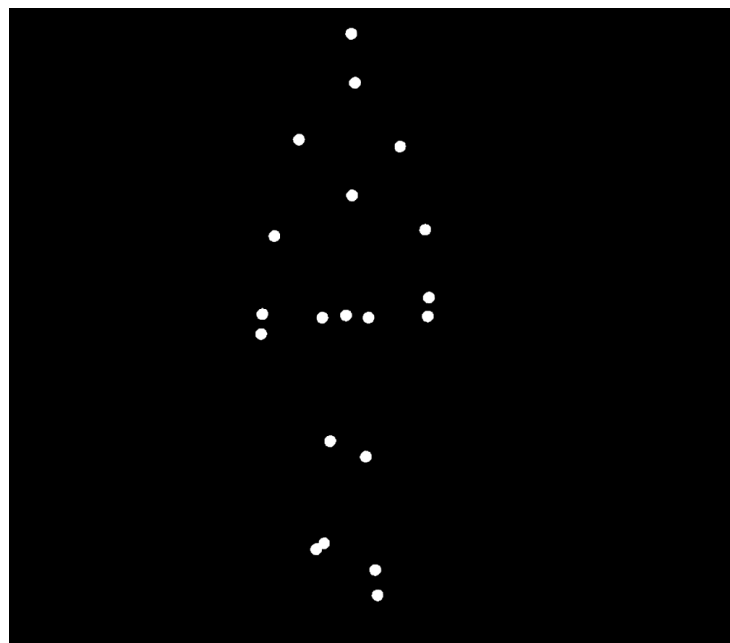
Data was captured at 24 frames per second, each frame represented by 60 numbers (3D coordinates of 20 joints) and a corresponding timestamp of capture of the frame. The analysis of the data collected is detailed in Chapter 6. Notably, this dataset is utilised for experimentation in Chapter 6 only.

## 3.2 Stimuli Development

Each of the HPs, BM and NBM have a bespoke requirements for the type of data they operate on. Thus, multiple data processing pipelines were developed to manipulate the skeleton motion data captured to appropriately format the data to be used as stimuli for the various models.

### 3.2.1 Psychophysical Experiments

The multidimensional time series data of the temporal evolution of the skeletal joint coordinates was converted into a moving point light display (PLD) by processing the raw joint coordinates in the MATLAB[140] programming environment. The MATLAB program was developed by William Hahn and Elan Barenholtz at Florida Atlantic University, who also ran the subjects for the psychophysics experiments. Each of the skeletal poses was converted into a frame in the PLD by developing 20 white dots on a black background for each of the 20 skeletal joints. The position of the white dot matched the two-dimensional position of the corresponding skeletal joint, omitting the depth information. The timestamps of the original data capture were utilised for determining the frame rate in the converted PLD. Fig. 9 shows an example of the conversion of one of the skeletal poses into a frame in the PLD.



*Figure 9: Conversion of one skeletal pose into a frame in the point light animation to be used as visual stimulus for the psychophysics study*

### 3.2.2 Computational Experiments

The data was processed separately for the NBM and BM, according to the specific models' operational requirements. In the case of the NBM, various spatiotemporal features were extracted from the multidimensional time series information of gait. The skeletal motion data was segregated into gait cycles, where the initiation and termination of the gait cycle is marked by the left foot coming in contact with the treadmill surface, concluding its swing phase. For each of the gait cycles, the following feature metrics were calculated:

- Swing/Stance Ratio- the ratio of the amount of time spent by each leg in swing and stance phase
- Single/Double Support Ratio- the ratio of the amount of time when the walker had one foot on the treadmill surface to both feet being on the treadmill surface
- Stride Length- the distance covered by either foot from toe off (entering swing phase) to initial contact (entering stance phase), on the treadmill surface
- Cadence- the number of gait cycles completed per minute

In the case of BM, no features were extracted to provide a close parallel to the input provided to human perception experiments. However, the spatial smoothing of the joints was performed to reduce the jitteriness of the raw motion capture using a 12 Hz Butterworth low-pass filter. All the programming developed for the experiments were developed in Python 3.5.2 using Anaconda 4.2.0 (64-bit)[78,141] as the Integrated Development Environment (IDE) of choice. The gait cycle feature extraction algorithms utilised the NumPy version 1.16.4 library[142], while the normalisations and low-pass filter smoothing were performed using the SciPy version 1.1.0 library[143].

### 3.3 Model Development

BM and NBM were developed using the open source Python libraries, Scikit-Learn v0.18.2 [144] and Tensorflow v1.13 [145], in Python 3.5.2. The support vector machines (SVM) along with the various kernel initialisations, decision trees (DTs) and the random forest initialisations from the decision trees were implemented in the Scikit-learn library. The artificial neural networks along with the various architectures (feed-forward, recurrent and LSTM variation) were implemented using the Tensorflow library. In addition to being widely used, the libraries were chosen for their ability to accelerate execution by leveraging the hardware. Although scikit-learn is largely written in Python, the core algorithms are written in Cython[146], an extremely efficient wrapper for C functions to achieve performance. For example, SVMs are implemented in the Cython wrapper around LIBSVM [146,147]. Its implementation negates the argument of Python being slow in training machine learning algorithms. Similarly, Tensorflow represents the artificial neural network as a set of multidimensional arrays (known as tensors) and can run on multiple CPUs and GPUs to leverage the parallel computation capabilities of multi-threading. Tensorflow leverages the CUDA technology [148] to execute operations on the tensors in parallel, to accelerate the process of learning. The implementation of all variations of artificial neural networks were developed natively in Python.



### 3.4 Data Analysis

In case of the psychophysics experiments (evaluating for the intrinsic feature), the human observers were given a forced binary choice question of inferring the gender from the point light animations. Each observer had a corresponding accuracy of identifying the gender of the walker. Similarly, in the case of the computational models, each model was treated as an ‘independent observer’. For BM, a different random initialisation of the network weights (before training) implied a different independent observer. As each observer was assessed for gender classification performance with different variations of stimulus exposure duration, the output data collected contains the gender classification results per observer. The detailed breakdown of the results are provided after each corresponding experiment in chapters 4, 5 and 6. A normality test is performed on the results using the Shapiro-Wilk test, given the limited number of observers/models. Following which, a two-sided student’s t-test, specifically a one-sample t-test, is used for calculating statistical significance of equality of the expected mean value of a sample population and a known mean value. The null hypothesis in the above is that the expected value (mean) of a sample of independent observation is equal to the known population mean. Normality of the data is ensured using Shapiro-Wilk normality test before its parametric analyses. In case of the intrinsic feature i.e. binary gender of male or female, the known population mean is taken at 50% (chance performance) whereas in the case of extrinsic feature i.e. emotions (positive, negative and neutral) the known mean is 33.33%. A two-sided one sample t-test was used to test the sample mean for both positive and negative differences from the known performance mean. Details of the extrinsic feature data analysis is detailed in Chapter 6. A one-way ANOVA [149] test is used to compare the performance between models for different exposure durations. The results of a single model was evaluated using a one-sample t-test because the ANOVA test requires at least two sample population for analysis. However, for experiments which resulted in two independent sample populations, the ANOVA test was used to determine significance in difference of the means, as it provides a higher statistical significance compared to a two-tailed t-test. The results of the experiments of each of the chapters are presented as values that include the statistical test value themselves (t-statistic and F-value) as well as the corresponding p-value of statistical significance. It is important to note that the p-values mentioned are rounded to the nearest significant digit to make a point about the significance of rejecting the particular null hypothesis. For example, if the p-value is 0.009, then it is mentioned as  $p < 0.01$ , rather than  $p < 0.05$  (as quoted in standard literature) to highlight their significance.

In addition, performances are presented in aggregated mean and standard deviation formats wherever relevant. Implementations for determining all of the above statistics can be found in the SciPy v1.1.0 and NumPy v1.16.4 library.

## Chapter 4 Biological, biomimetic and non-biomimetic learning

In the endeavour of developing a high performing, robust and versatile artificial gait classifier, inclusion of human perception (HP) as a gait classifier is essential, given the accurate and versatile nature of the biological perception.

The definitions of biological, biomimetic and non-biomimetic were mentioned briefly in the methods chapter. However, to avoid any confusion, they are reiterated as follows,

- Biological- this category includes the human observers (HP) only, given their biological learning nature
- Biomimetic (BM) - this category includes the models that aim to emulate or mimic biological systems in neuroscience. This includes the perceptron-based artificial neural network models. Notably, the models included in this category are structurally biomimetic. Their extent of functional biomimicry with respect to gender classification from gait is the emphasis of this chapter.
- Non-biomimetic (NBM) - this category includes models that are purely statistical and do not attempt to mimic any biological system. This includes hyperplane-based models, such as SVMs and models based on reduction of information entropy, such as the DTs.

Notably, the definitions of ‘biomimetic’ and ‘non-biomimetic’ are limited with regards to their biological inspiration in terms of learning paradigms and does not include definitions pertaining to robotics or other variations.

Inclusion of HP in this study also serves the purpose of filling the gap in the literature about evaluation of human and machine-based learning systems on the same gait dataset, as most studies focus on either human or machine-based gait analysis separately. In the case of humans, studies have shown a significantly higher than chance performance for HP in case of the objective of gender and personnel identification of the walker [7, 10, 18–20,115,127]. Most studies testing the validity of artificial gait classifiers have used gender classification as the test objective [7, 9, 10, 20, 32,115,127,150], owing to the abundance of literature about H’s ability to distinguish gender while acting as a good proxy for gait classification into binary classes. Maintaining the theme, this chapter uses gender classification as a testbed for evaluating HP, BM and NBM, on the same stimulus set. In addition to serving the practical objective of developing an artificial gait classifier, comparing humans with the various different machine learning algorithms, each possessing a different approach to learning, would help structure our understanding of HP and learning mechanism as well.

Biological learning systems, such as HP, perceive gait motion through visual observation and extract the necessary information for a given objective, utilising either biological conditioning (for gender classification) or specialised learned training (for clinical gait disorder classification). Although highly efficient, the knowledge and perception learned by the biological brain is not easily replicated or transferred. The learning period to develop the necessary perception is highly

variable. The learned perception itself is volatile and susceptible to fatigue, illness and mortality. Advancements in machine learning technology has recently enabled us to develop machine-based perception and apply it to real world problems. Requiring a fraction of the time and resources, machine perception is impervious to the shortcomings of HP. In this study, we explore a variety of machine learning algorithms and compare them with humans, in the specific task of correctly perceiving gender from human gait. The specific objective was chosen given the human brain's impressive ability to correctly identify gender from human gait. However, the learning paradigms could potentially be used in other gait analyses, including clinical gait analysis, with minimal modifications. However justifying the same requires further experimentation. Our study was conducted in two parts, 1) Psychophysical experiments for evaluating HP, 2) Computational experiments for evaluating machine perception. Both variations used the same dataset as input. This study emphasises on maintaining unbiased comparison conditions by avoiding heavy human induced modifications to the barebones 'off-the-shelf' computational models. One such modification could be by combining two or more different architectures of neural networks (such as CNNs and RNNs), thus creating ambiguity in contribution between either architecture. Appropriate pre-processing is performed based on the type of input expected by the models while maintaining high fidelity to the information content of gait in the inputs provided. The dataset used in all the experiments in this chapter is the same as described in section 3.1.1. This chapter focuses on a broad comparison of machine learning models with human perception on the task of gender identification from gait. The objectives of this chapter include (1) ascertaining similarities between the machine learning models and human perception, (2) deriving a baseline accuracy of classification of gender from gait. The gait dataset used for this chapter is a combination of the data collected in the US and the UK, detailed in the methods chapter (section 3.1.1). The combined dataset contains skeletal gait representations of forty one consenting healthy adults (26 male, 15 female) between the ages of 18 and 50 years.

## 4.1 Experiment 1: Biological Models

Humans are the most versatile learners known in history. They have been known to reach the extent of neuroplasticity to alter existing biological nervous connections to achieve a certain objective [151], making those extremely powerful learning systems. Human vision in conjunction with their mental perception forms a complete sensory pathway to learn from, with the vision, acting as the data source and perception representing the learning model for inference from the visual data. In terms of gait perception, humans have been shown to understand multiple characteristics of gait through visual perception alone. Studies have shown humans to require no longer than 2 complete gait cycles to correctly identify gender from human gait [32, 151]. In terms of duration, this translates to less than 2.7 seconds of walking animation. Although humans can detect and decipher biological motion from point light animation of walking human figures within 200 msec, at least 1.6 seconds of stimulus is required for significantly above chance performance. The current experiment aims to establish gender classification performance in humans as a function of increasing duration of stimulus exposure. In other words, the experiment aims to test humans for gender classification accuracy and

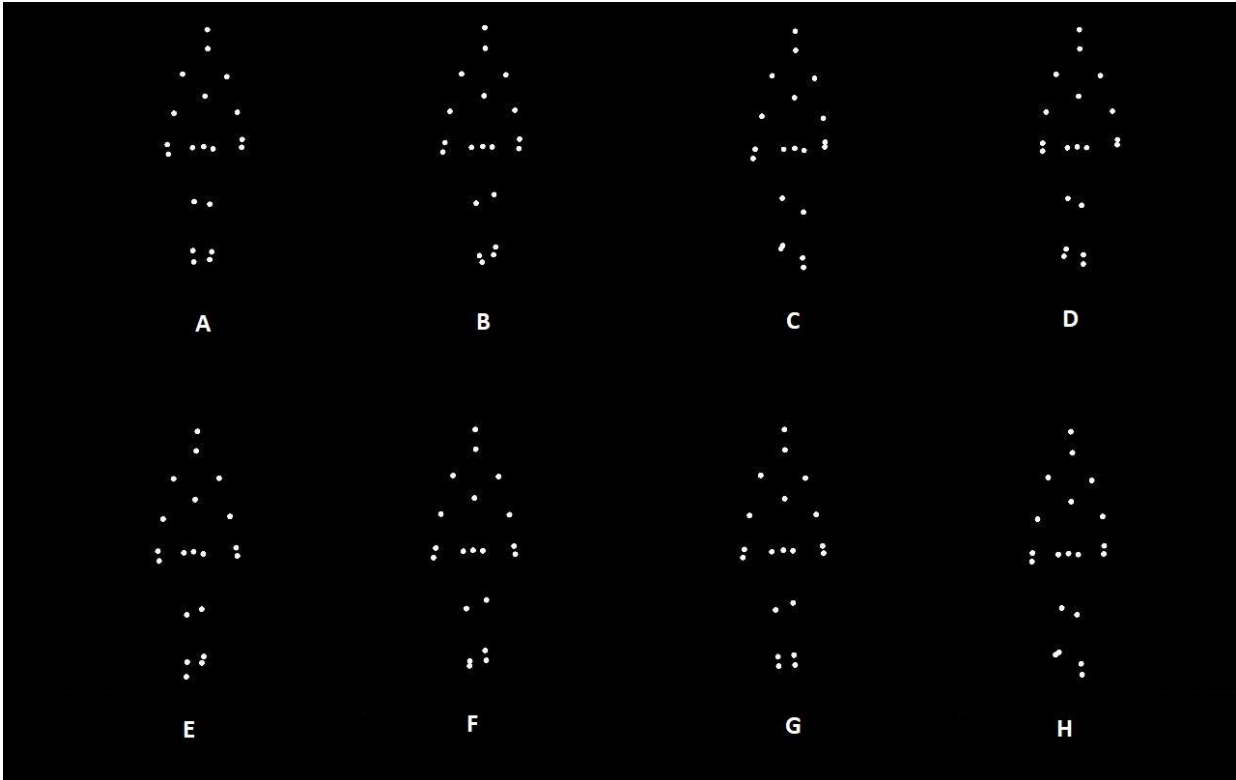
sensitivity with different durations of stimulus with each stimulus ending at a different stage of the gait cycle.

## 4.1.1 Experiment Setup

### 4.1.1.1 Stimuli

A PC-compatible computer monitor with a high performance raster graphics system displayed stimuli on an Iiyama ProLite B2283HS colour monitor (1920x1080 resolution, 60 Hz refresh rate). Human figures were defined by 20 circular white dots of 5 pixel radius overlaid on a black background, located on the head, neck, shoulders, elbows, wrists, fingertips, back, spine, hips, knees, ankles and toes. None of the dots were occluded by other subjective parts of the figure. Animated sequences were developed by placing the dots at the three-dimensional trajectory of each of the 20 tracked joints, and temporally sampling the coordinates to produce 24 static frames per second, as shown in Fig. 10.

The stimulus size was  $6^\circ$  wide and  $8^\circ$  long for the whole frame, including zero (black) padding. A degree here is defined as the subtended angle at the nodal point of the eye. The actual walking clip was  $2.5^\circ$  long and  $4^\circ$  wide. When the static frames were played in quick succession, a vivid impression of a walking person emerged. There was no progressive component to the walking animation, thus the human figure appeared to walk on an unseen treadmill with the walking direction oriented towards the observer. None were notably over- or underweight, as shown in Table 1. The x and z component were sampled to display the walker in the coronal plane. The coronal plane was chosen over the sagittal plane or axial plane to emphasise lateral sway and maximise the provision of dynamic cues to the observer [32,127]. The recorded gait sequences were converted into an animation sequence in the same fashion to be presented as visual stimuli. Animation playback was normalised for size [127] and occurred at veridical speed with linear interpolation of joint trajectories between frames. The veridical speed was determined based on the timestamps attached with each recorded frame. The observers were seated in a well-lit room in front of the monitor and had access to a standard computer mouse for interaction. The randomly chosen walker stimuli were presented for exposure durations of 0.4, 1.5, 2.5 and 3.8 sec, followed by an on-screen prompt in the form of two buttons, a 2-alternative-forced-choice (2AFC) paradigm, requesting the observer's perceptual classification of gender through a mouse click with screen buttons labelled as either 'male' or 'female'. The order of all the stimuli presented was randomised. Following the response from the observer, the next stimulus was presented. A total of 200 walking clips were shown per observer per exposure duration and the responses recorded for each.



*Figure 10: Point Light Displays (PLD) of a walking stimulus, labelled A to H, at eight different stages of a gait cycle. In psychophysics experiments, the eight different body poses shown here are an example of the sequence of frames shown in the animation. Left to right, top to bottom show each of the frames that are shown one after the other at a rate of 24 frames per second to create the illusion of motion. Here, the body poses labelled A to H are played in quick succession to give the illusion of walking motion*

#### 4.1.1.2 Biological Model (HP)

Twenty one (15 female, 6 male) healthy observers with age ranging from 20 to 43 years old, participated in the experiment. All had some experience of biological motion displays, although none had been required to make judgements about gender. All subjects were naïve to the purposes of the experiment. Appropriate consent forms were obtained from the observers to ensure proper ethical compliance with approved Florida Atlantic University's Internal Review Board (IRB).

#### 4.1.2 Results

Using HP as the classification model achieved an accuracy of 63% of all the trials across all exposure durations, which was significantly greater than chance performance of 50% ( $t_{20} = 7.8$ ,  $p < 0.001$ ). Correct classification at 0.4 sec, which consisted of approximately a quarter of a step cycle was above chance at 60%, ( $t_{20} = 3.7$ ,  $p < 0.01$ ), which was in disagreement with a previous study [32]. This could be attributed to the presentation of the stimulus in the coronal plane as opposed to the sagittal plane [7], leading to higher emphasis on the dynamic cues. Performance at 1.5 sec was 65.6% ( $t_{20} = 3.8$ ,  $p < 0.005$ ), which is higher than the performance at 2.5 sec of 61.1% ( $t_{20} = 4.8$ ,  $p < 0.001$ ). One study explains this anomalous phenomenon due to an additional partial step at 2.5

sec by highlighting the preferred perception of velocity over positional cues, where sensitivity to gender classification decreases mid-swing in the gait cycle [127]. HP was able to discriminate gender with highest accuracy at 3.8 sec with 64.7%, ( $t_{20} = 3.4$ ,  $p < 0.01$ ). Details of the results are shown in Table 2. Overall, HP is consistent with other perception studies [32, 67, 127], thus providing a reliable baseline for comparison with the biomimetic performance on the same stimulus set.

Model\Stimulus Duration	0.4 sec	1.5 sec	2.5 sec	3.8 sec
Biological (HP)	60 ( $p < 0.01$ )	66 ( $p < 0.005$ )	61 ( $p < 0.001$ )	65 ( $p < 0.05$ )

*Table 5: Gender classification accuracy in % as a function of exposure duration of the stimulus*

## 4.2 Experiment 2: Biomimetic Models

The artificial gait classifier should not only be able to classify gait for the objectives tested in this experiment, but also be versatile enough to classify a variety of given gait classification or regression objectives. Human observation has been known to achieve this amount of versatility and performance, which shall be utilised as a conceptual benchmark for evaluating the computational models. Artificial neural networks attempt to emulate the connectionist approach to learning in humans [152]. The connectionist approach attempts to mimic various aspects of the human brain. Convolutional neural networks (CNN) mimic the layered approach of the vision through implementation of a series of convolutions mimicking the layers V1 to V5 in the human visual cortex [153–157]. Recurrent neural networks (RNN) attempt to mimic the temporal pattern processing capabilities of the human brain by emulating biological memory. However, RNNs suffer from a technical problem, rendering their memories extremely short-term [158]. An improvement on the RNNs known as Long Short Term Memory (LSTM) cells, better emulates human memory and its temporal pattern capabilities by developing additional learnable gates to remember or forget information, enabling it to process temporal information over long periods of time. The learning gates inherent in the network parallel the short and long term memory of the human brain, enabling the network to remember the relevant temporal pattern while ignoring patterns that don't contribute toward the classification objective. This experiment aims to present an LSTM network with the temporal evolution of joint trajectories during human gait and train it for gender classification to evaluate for resemblance with HP. Although there exist variations of the LSTM architecture, such as BiLSTM [159] and GRU [82] networks, they are modifications of the basic concept of introduction of multiple memory gates into the network, thus the experiment considers a standard LSTM without specialised modifications. Successful conformity between the biological and the model would encourage further exploration into the biomimetic approach to improve the BM model to superhuman performance levels.

### 4.2.1 Choice of Biomimetic Model (BM)

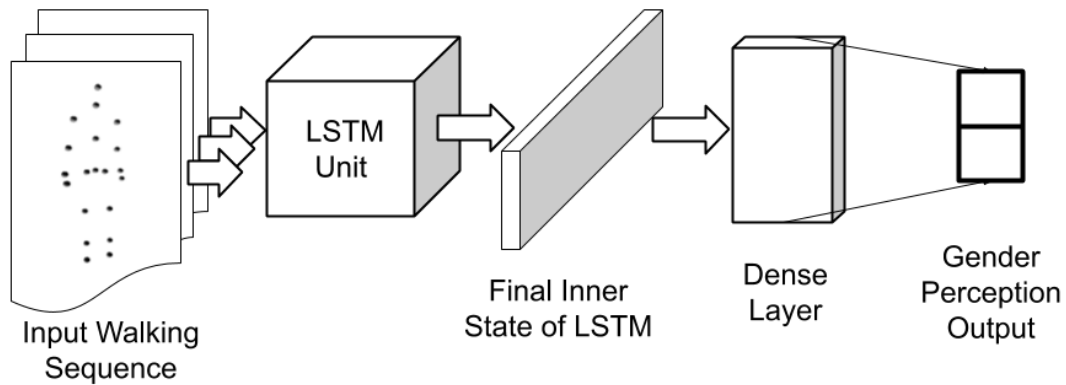
While numerous models and approaches are possible [27,70,89,90,93] the following criteria were applied for the selection of biomimetic BM models: (1) modelling is based on biological principles and an understanding of neuroscience in a connectionist approach, (2) processing requires minimal human-assisted, hand-crafted feature design, (3) models are capable of processing arbitrarily long data sequences and (4) models are practical enough to train and classify on the available dataset.

Previous works have proposed a computational biological model for motion perception through the use of feed-forward and recurrent ANNs that aim to emulate the two-fold neural pathway [89, 90]. The study [89] assumes that biological movement classification is based on a feedforward architecture of learned prototypical patterns. The proposed model develops corresponding models for global and local processing modules to provide a one-to-one correspondence to the hypothesised perceptual modules. Specific global and local features are extracted from motion (e.g. optical flow) to parallel the modification of information in the neural pathway. The two pathways emulate the form and motion pathways used by biological snapshot and pattern neurons. They were able to get high accuracy in identifying walking, running and limping using the computational models. One of their conclusions included the lack of need for modelling attention mechanisms for motion perception. Further studies [89,90] present a neural model that simulates receptive fields for images of the static human body, as found by neuroimaging studies and temporally integrates their responses by leaky integrator neurons. The model performance was compared to psychophysical experiments in terms of ability to categorise point-light displays into human walking and scrambled motion, with significant similarities in performance as they showed using template matching. Despite the perceptual correspondence, the practical applications of the model for automated gait classification is subject to the availability of an extensive training data and requirement for high computational capability, given the large number of tuneable parameters in the models. Moreover, the aforementioned model requires explicit pre-processing of gait information by extracting hand-crafted optical flow features. The limited dataset and computational capability therefore dictates a more practical alternative that still meets the mimetic criterion. The models aim to mimic the human perceptual pathway very closely, discouraging non-human alterations to be performed on them to increase the performance to ‘superhuman’ levels. Computationally, traditional RNNs used in the previous studies suffer from the vanishing and exploding gradient problem, making them ineffective in meeting the criterion of being able to process long sequences [158]. On the other hand, Long Short Term Memory (LSTM) cells, a variety of RNNs, introduce additional gates in the network that regulate the flow of information into short and long term memory, thus enabling them to remember relevant temporal patterns over long periods of time [84]. LSTM cells also mimic the memory capability in human learning more closely. In particular, their ability to learn multidimensional time series representations captures the dynamic joint trajectories in gait from point light animations as those used by HP. Additionally, there is no restriction on the model for the provision of structural information for processing. For the purposes of the current experiments, we consider LSTM models operating on sequences of PLD motion to be BM model

machine learning model under evaluation. The alterations are inspired from HP literature to further test the biomimetic nature of the model, in the next chapter.

## 4.2.2 Experiment Setup

### 4.2.2.1 Biomimetic model (BM)



*Figure 11: Implementation of the LSTM network architecture for processing gait sequences. Gait is provided as a temporal sequence of 3D body poses, to the LSTM and the final inner cell state of the LSTM is fully connected with the binary output layer*

A standard LSTM cell model consisting of 128 hidden states was designed as shown in Fig. 11. The cell state weights were initialized as a random normal distribution. The final cell state was ReLU activated [160] and connected to an affine output layer, which represented the one-hot labelled gender identity of the walker during training. The ‘Male’ and ‘Female’ labels are represented as [1, 0] and [0, 1] respectively. During testing, the output layer represented the classification values. The error of classification was evaluated using a cross-entropy function [161] for updating the weights using an Adam optimiser [162] based on the error differentials and a learning rate of 0.001. The most probable output was taken as the class label during classification. 10 different LSTM models were developed by randomising the initial weight matrix before training. Each of these could be argued to represent an independent HP undergoing training. The reason for the independence was also for the comparable statistics as followed in standard machine learning literature.

#### 4.2.2.2 Data Input.

The three-dimensional trajectories of each of the 20 tracked joints were concatenated to form a vector representation of a static frame with a cardinality of 60, representing the location of the head, neck, shoulders, elbows, wrists, fingertips, mid-back, hips, knees, ankles and toes. Gait input to the model consisted of a sequence of vector representations of subsequent static frames, sampled at 24 frames per second. Joint trajectories were size normalised [127] and standardised with a zero mean and unit standard deviation. Model training sessions included, initialization of the model weights, classification of the output probabilities based on the gait input, propagation of the classification



error and updating the network weights. Model training was executed in batches of 50 and repeated for 100 epochs. Input sequence durations mirrored the exposure durations in the corresponding HP experiment and varied incrementally for 10 durations from 0.4 sec to 3.8 sec in steps of 0.4 sec (10 static frames). 10-fold cross validation was carried out to ensure model generalizability and a total of 250 gender classifications were obtained per input sequence duration. The models trained per session per duration were stored locally for future analyses.

#### 4.2.2.3 Model Training and Testing

Each model training sessions included:

- initialization of the model weights,
- classification of the output probabilities based on the gait input,
- choosing the higher probability output as the predicted class and converting it into a one-hot encoding,
- calculating the error by comparing it with the actual one-hot label,
- back propagating the error deltas, and
- updating the network weights using the Adam optimiser.

Model training was executed in batches of 50 and repeated for 100 epochs to ensure stable error function values before backpropagation and update of weight values. Input sequence durations mirrored the exposure durations in the corresponding HP experiment and varied incrementally from 0.4 sec to 3.8 sec in steps of 0.4 sec. A 10-fold cross validation was carried out to ensure model generalizability. The model was generated and trained ‘from scratch’ for each fold. A total of 250 gender classifications were obtained per input sequence duration. Results were obtained as accuracy of gender classification for each gait duration.

#### 4.2.3 Results

BM correctly classified 76% of all the gait inputs presented across all the input durations ( $t_9 = 9.2$ ,  $p < 0.001$ ). Chance performance remains the same at 50%. Correct classification at about a quarter of a step cycle at 0.4 sec was 71.5%, ( $t_9 = 5$ ,  $p < 0.001$ ), higher than the same with HP ( $F_{1,29} = 3.6$ ,  $p < 0.05$ ). The difference in performance indicates a higher inference capacity from a limited amount of available data. Here, performance is defined as the correct classification of gender as a percentage of the total test samples classified. The inference performance increases slightly with an increase in the amount of information available from 0.4 to 3.8 sec ( $t_{9,9} = 2$ ,  $p < 0.1$ ). At 3.8 sec, the model correctly classified gender with 81.2% accuracy ( $t_9 = 9.6$ ,  $p < 0.001$ ), considerably higher than HP ( $F_{1,29} = 9$ ,  $p < 0.01$ ). Generalizing across all the input (or exposure) durations, the classified gender with a significantly higher accuracy than the HP ( $F_{1,29} = 39.9$ ,  $p < 0.05$ ). Details of results obtained for the BM have been presented in Table 3 with the corresponding trend plotted in Fig. 12. As shown in the figure, mean performance peaks temporarily at 1.6 sec (about halfway completion of one gait step) with 79% accuracy ( $t_9 = 10.1$ ,  $p < 0.05$ ), suggesting a dependence on dynamic and velocity cues similar to HP at 1.5 sec. Notably, performance (percentage of test data correctly classified) at all durations was above chance. It may also be noteworthy that while HP was exposed to the stimulus for a specific amount of time, whereas the BM was exposed to a specific number of frames of the stimulus. All data was captured at about 24 frames per second,

essentially translating a single frame into 0.04 seconds duration. Table 3 shows the exposure duration in a number of frames as well as in terms of number of frames, for a more accurate representation of the stimulus. However, Fig. 12 shows gender classification as a function of exposure duration in time only for a more intuitive visual representation.

In summary, the BM performed significantly better than chance in gender classification from 3D moving point representations of human gait. Similar to human perception, there was a significant increase in gender classification accuracy for gait information duration of 0.4 seconds compared with duration of 3.8 seconds. The increased gender sensitivity at 1.6 seconds could be attributed to an inherent sensitivity to dynamic velocity-based cues in LSTM networks for gender classification, similar to HP. One could argue that the presentation of the stimulus as a forward facing (towards the camera) skeleton could potentially limit real-world applications. However, the availability of 3D data could be leveraged to apply a simple preprocessing rotational step to the skeleton to correct for any misalignment in global skeletal configuration. Having received the results from both HPs and BM, it would be interesting to compare them to the results obtained from computational models that do not emulate basic human biology and rely on a more statistical approach to learning and classification. Such models, are termed as NBM.

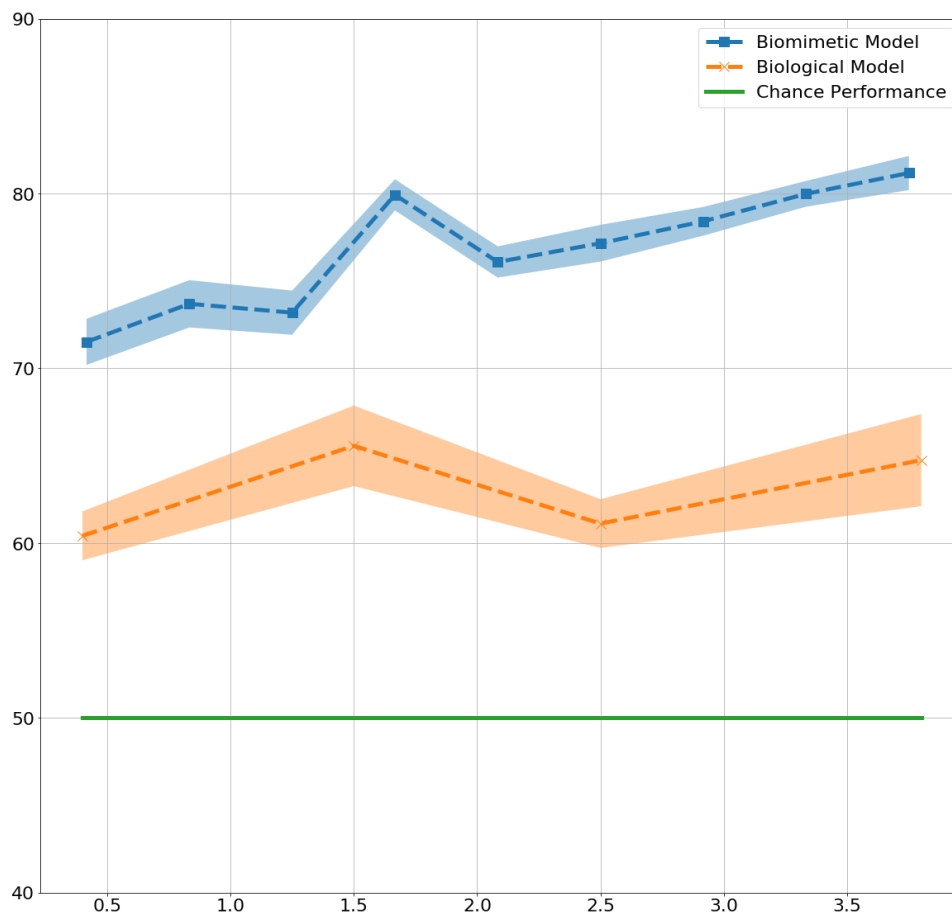


Figure 12: Correct gender classification accuracy (in %, mean and standard error) by the BM as a function of exposure duration in seconds.

Stimulus Duration / Model	0.4 sec	1.5 sec	2.5 sec	3.8 sec
Biological (HP)	60 (p<0.01)	65.6 (p<0.005)	61 (p<0.001)	64.7 (p<0.05)
Biomimetic (BM)	71.5 (p<0.001)	73.1 (p<0.001)	77.1 (p<0.001)	81.2 (p<0.001)

Table 6: Correct gender classification (in %) as a function of exposure duration of the stimulus (in terms of frames as well as seconds) with  $p < 0.001$  (using two-tailed  $t$ -test compared with chance) for all the durations.

### 4.3 Experiment 3: Non-Biomimetic Model (NBM)

Most machine learning-based automated gait analysis literature utilises NBM such as support vector machines (SVM) [77] and random decision forests (RDF) [122]. The NBM are capable of analysing static data only, unlike HP and LSTMs which can analyse sequences of data. Their reliance on the principles of linear separability and information entropy to create rules for classification, resembles expert systems. These models require a static vector representation of features and a corresponding class label for training. For testing, they require a static vector of features to predict the output class. In the case of gait data, they would require a static representation of the dynamic spatiotemporal gait data for classification. Thus, for the purpose of this experiment gait data was represented statically in two forms, (1) as first-order statistics of the temporal signals, and (2) extracted metrics used in a clinical setting to describe gait for diagnosis and rehabilitation monitoring. In this experiment, we evaluate the SVMs and RDFs on the two static representations of gait data for resemblance with HP and BM.

#### 4.3.1 Experiment Setup

##### 4.3.1.1 Non-biomimetic Model (NBM)

SVMs are designed to linearly separate a set of vectors to achieve maximum classification accuracy. Their training phases help find the most optimal hyperplane for linear separation of sets of numerical feature vectors into binary classes. SVMs operate on numerical feature vectors and cannot implicitly handle categorical data. Decision trees, on the other hand, learn to classify data by choosing optimal split conditions of attributes to minimise ambiguity in classification. RDFs are a collection of decision trees that have been trained on randomly chosen subsets of features from the original dataset. The final verdict of predicted class in an RDF is taken as a majority vote of the consistent decision trees. The advantage of the RDF over SVMs lies in its ability to classify categorical data, in addition to numerical data. The SVMs used in this experiment vary in their kernel function that help project the feature set into an additional dimension to attain better linear separability. The SVM kernels include:

- Linear, referred to as *SVM-Linear*,
- Radial basis function with gamma as 0.99, referred to as *SVM-RBF*, and
- Sigmoid, referred to as *SVM-Sigmoid*

The RDFs consisted of 10 randomly generated ID3 Decision Trees[28] developed with a minimum requirement of 2 samples for splitting and a maximum of 3 features for splitting consideration.

#### 4.3.1.2 Data Input

As mentioned earlier, gait data was processed for its static representation in two forms, (1) first-order statistics for the time series data, and (2) relevant clinical gait metrics, as follows:

- Gait data, originally represented as a temporal sequence of vectors, was described with four first-order statistics attributes, namely, minimum, maximum, mean and standard deviation of each dimension of the multi-dimensional time series signal. Given that the data was represented as a time series of 20 triplets (60 values per frame), the static representation consisted of 240 numerical values for every gait sequence. The temporal sequence duration of the signal varied from 0.4 sec to 3.8 sec in steps of 0.4 sec (10 frames). The resulting static dataset was normalised between [-1, 1] and standardised to have a mean of zero and unit standard deviation. The final processed dataset consisted of 250 features (of 240 values each) for each gait duration, with 9 gait durations in total (from 0.4 to 3.8 seconds).
- The study was conducted in conjunction with the Cambridge University Hospitals, thus the gait metrics utilised for clinical gait analysis in the gait analysis laboratory were mirrored as static representations of gait. 12 clinically relevant spatiotemporal gait metrics, including, stride length, cadence, single-double support, stance-swing phase ratio, speed of walking and knee flexion for each leg during stance and swing were provided as input feature sets to the static learning models. All the features were standardised to have a mean value of zero and a standard deviation of one. The features were further normalised to lie within the [-1, 1] range for uniformity and to discourage the models from learning the gender from the structural information and to rely solely on the gait dynamics. The final dataset consisted of 250 features (of 12 numerical values each) for each of the gait duration, with 9 gait durations in total.

Notably the spatiotemporal gait metrics were extracted per gait cycle by synchronising the recorded skeletal gait with the template of normative gait, as described in section 2.1. A gait cycle starts with the left foot toe off (lifting off of the ground), going into swing phase, subsequently entering stance phase with heel touch, finally followed by toe off again. The cycle from toe off to the subsequent toe off of the left foot was considered a single gait cycle and the metrics extracted for each of such cycles.

#### 4.3.1.3 Model Training and Testing

Every training and testing session involved the development of SVMs (with various kernels) and RDFs ‘from scratch’, trained with a chosen section of the dataset and tested with the held-out dataset. A 10-fold cross validation training-testing protocol was followed, ensuring no overlap of participants between training and testing datasets. Thus, the testing dataset consisted of people not seen before by the models during training, maintaining generalizability of the models.

The cycle of model development, training and testing was started afresh for each session and for each type of data i.e. separately for first-order statistics and clinical gait metrics.

### 4.3.2 Results

The results obtained through operations upon (1) first-order statistics, and (2) clinical gait metrics were as follows:

- In case of the first-order statistic feature sets, at 0.4 sec, the RDFs and SVMs of all the kernels were able to identify gender with significantly better than chance performance with:
  - RDF at 75%,  $t_9 = 9$ ,  $p < 0.001$ ,
  - SVM-Linear at 84%,  $t_9 = 12$ ,  $p < 0.001$ ,
  - SVM-RBF at 78%,  $t_9 = 10$ ,  $p < 0.001$ , and
  - SVM-Sigmoid at 68%,  $t_9 = 4$ ,  $p < 0.01$ , as shown in Fig. 13.
- There was no significant difference in performance in the NBM between 0.4 sec and 3.8 sec of duration, as shown in Fig. 13. In contrast, HP and LSTMs had demonstrated a substantial increase in performance with increase in temporal data availability. Gender sensitivity remained similar across all the durations of gait input, demonstrating a dynamic cue agnostic learning mechanism. The performance outputs have been detailed in Table 4.
- In case of the clinical gait metrics, all the NBM performed close to chance performance. The best performing out of the cohort was the SVM with radial basis function (SVM-RBF) with a gender classification accuracy of 59% ( $t_9 = 1.8$ ,  $0.1 < p < 0.2$ ), followed by the RDF classifier with an accuracy of 59% ( $t_9 = 1.5$ ,  $0.1 < p < 0.2$ ). The statistical significance of the results is yet to be established with the collection of more data. However, the motivation for testing the NBM further is reduced based on the results.

Duration /Model	0.4 sec	1.5 sec	2.5 sec	3.75 sec
SVM-Linear	83.8 ( $p < 0.001$ )	83.5 ( $p < 0.001$ )	82.8 ( $p < 0.001$ )	82.5 ( $p < 0.001$ )
SVM-RBF	78 ( $p < 0.001$ )	78 ( $p < 0.001$ )	78.5 ( $p < 0.001$ )	77.9 ( $p < 0.001$ )
SVM-Sigmoid	68 ( $p < 0.01$ )	68.1 ( $p < 0.01$ )	68.2 ( $p < 0.01$ )	68.2 ( $p < 0.01$ )
RDF	74.9 ( $p < 0.001$ )	74.6 ( $p < 0.001$ )	74.2 ( $p < 0.001$ )	73.5 ( $p < 0.001$ )

Table 7: Correct gender classification accuracy (in %) by the NBM as a function of the duration of stimulus exposure (in number of frames). Statistical significance ( $p$ -value) is obtained using two-tailed  $t$ -test compared with chance (50%).

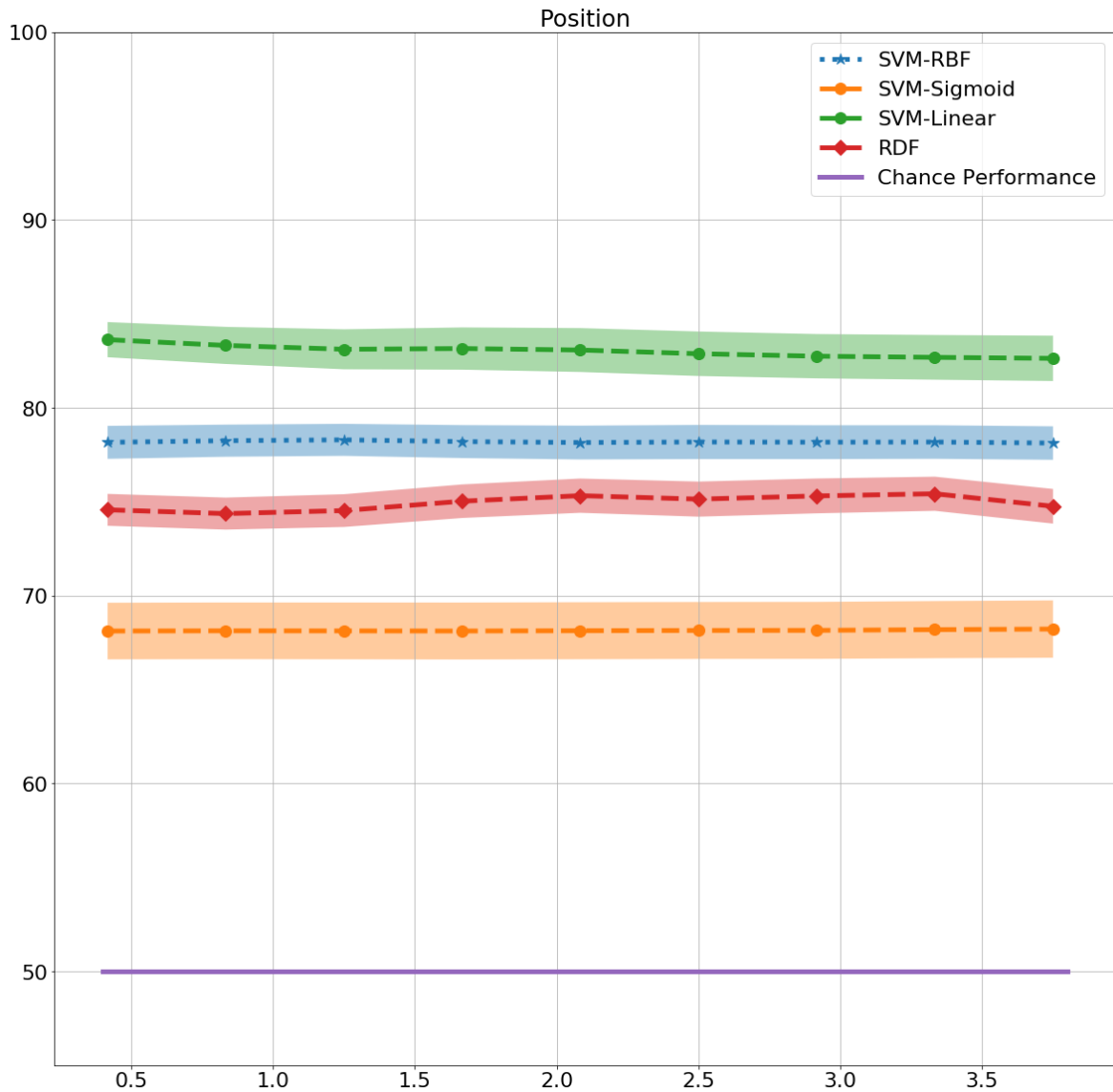


Figure 13: Correct gender classification accuracy by NBM (in %, mean and standard error) as a function of duration of gait data (in seconds) used to generate the static representation.

#### 4.4 Experiment 4: Biomimetic (BM) and Non-Biomimetic Models (NBM) with Velocity Cues

Humans are known to rely on dynamic velocity-based cues when determining gender from gait, through (1) increase in performance between the coronal and sagittal point dot representations of the walking subject, and (2) change in gender classification sensitivity at different stages of the gait cycle. This experiment focuses on training the BM and NBM on velocity cues exclusively, to evaluate for the presence of this human-like characteristics. A conformity to the sensitivity profile obtained in Experiment. 2 establishes a common trait shared between human and machine-based models.

## 4.4.1 Experiment Setup: Biomimetic Model (BM)

### 4.4.1.1 Biomimetic model (BM)

The architecture of the BM follows the same architecture as described in Experiment. 2. The difference in this experiment lies in the data for training and testing rather than the model itself. The weights initialisation, training regime and classification protocols are kept the same.

### 4.4.1.2 Biomimetic Data Input

In Experiment 2, the BM was exposed to data that represented the joints as triplets of values that represented the three-dimensional position of the joint relative to the sensor position. For this experiment, the model is exposed with triplets of values, but representing velocities of the joints as opposed to the position. Temporal derivatives of the gait of the walkers were used for generating corresponding velocities of the joints. The positional data was smoothed with a 5-frame moving average filter before calculating the derivatives for adjacent frames. The training and testing protocol was followed as mentioned in Experiment 2. Similarly, a 10-fold cross validation was performed to ensure generalisability of the model, ensuring that the walking participants included in the test set were different from the participants in the training set.

## 4.4.2 Results

### 4.4.2.1 Biomimetic model (BM)

The biomimetic LSTM model trained with three-dimensional velocity (LSTM - Velocity) achieved an overall accuracy of 81% ( $t_9 = 9.4$ ,  $p < 0.001$ ), significantly better than the HP ( $F_{1,18} = 82$ ,  $p < 0.001$ ) for all durations as well as the LSTM model trained with three-dimensional positions (LSTM - Position) of joint trajectories, ( $F_{1,18} = 5.6$ ,  $p < 0.05$ ). The model achieves its highest accuracy at 2.8 seconds of exposure with an accuracy of 83%, ( $t_9 = 9.4$ ,  $p < 0.001$ ). As shown in Fig. 14, the lack of velocity-based cues is noticed at 2.2 seconds in LSTM - Velocity, corresponding to the similar lack dynamic cues at the same time in LSTM - Position, demonstrating the dependence of the LSTM network on velocity while determining gender.

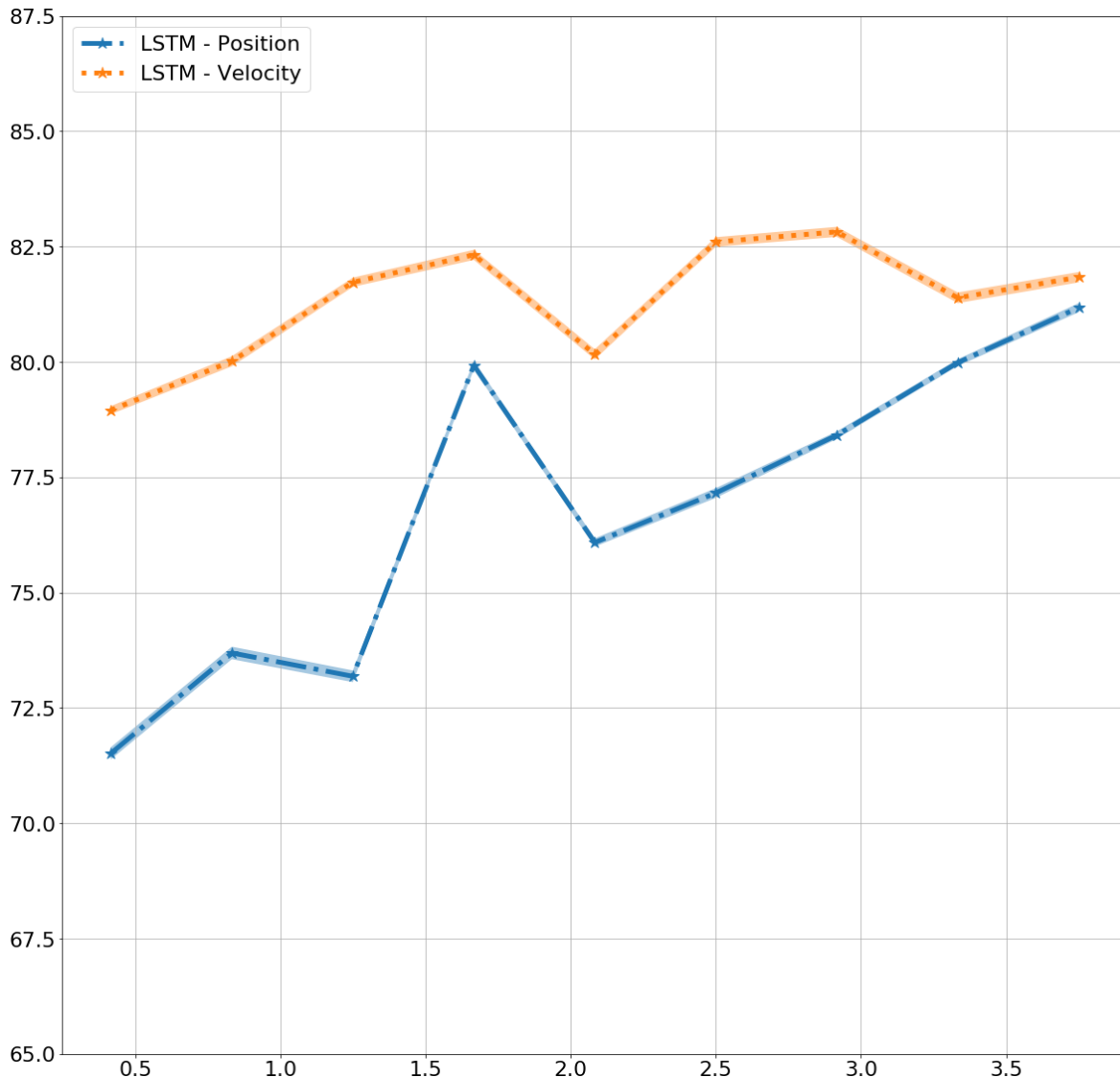


Figure 14: Correct gender classification (% , mean and standard error) by LSTM models trained with position and velocity as a function of exposure duration in seconds.

#### 4.4.3 Method: Non-Biomimetic Model (NBM)

##### 4.4.3.1 Non-Biomimetic Model (NBM)

The non-biomimetic models namely the SVM-Linear, SVM-RBF, SVM-Sigmoid and the RDF remain the same as the previous experiment, however the training data provided is a static representation of the joint velocities trajectories as opposed to the positional trajectories.

##### 4.4.3.2 Non-Biomimetic Data Input

The positional joint trajectories of 20 tracked joints are smooth using a 5-frame moving average filter followed by a temporal derivative of the smoothed signal to obtain the three-dimensional velocities of the joint trajectories. The result is represented as four first-order statistical attributes, namely, minimum, maximum, mean and standard deviation of the signal. The temporal duration of the signal varies from 0.4 sec to 3.8 sec in steps of 0.4 sec (10 frames). The static representation is normalised between  $[-1, 1]$  and standardised to have zero mean and unit



standard deviation. The resulting dataset is used for training and testing of the NBM models with a 10-fold cross validation of the walkers to ensure generalisability.

#### 4.4.4 Results

##### 4.4.4.1 Non-biomimetic models (NBM)

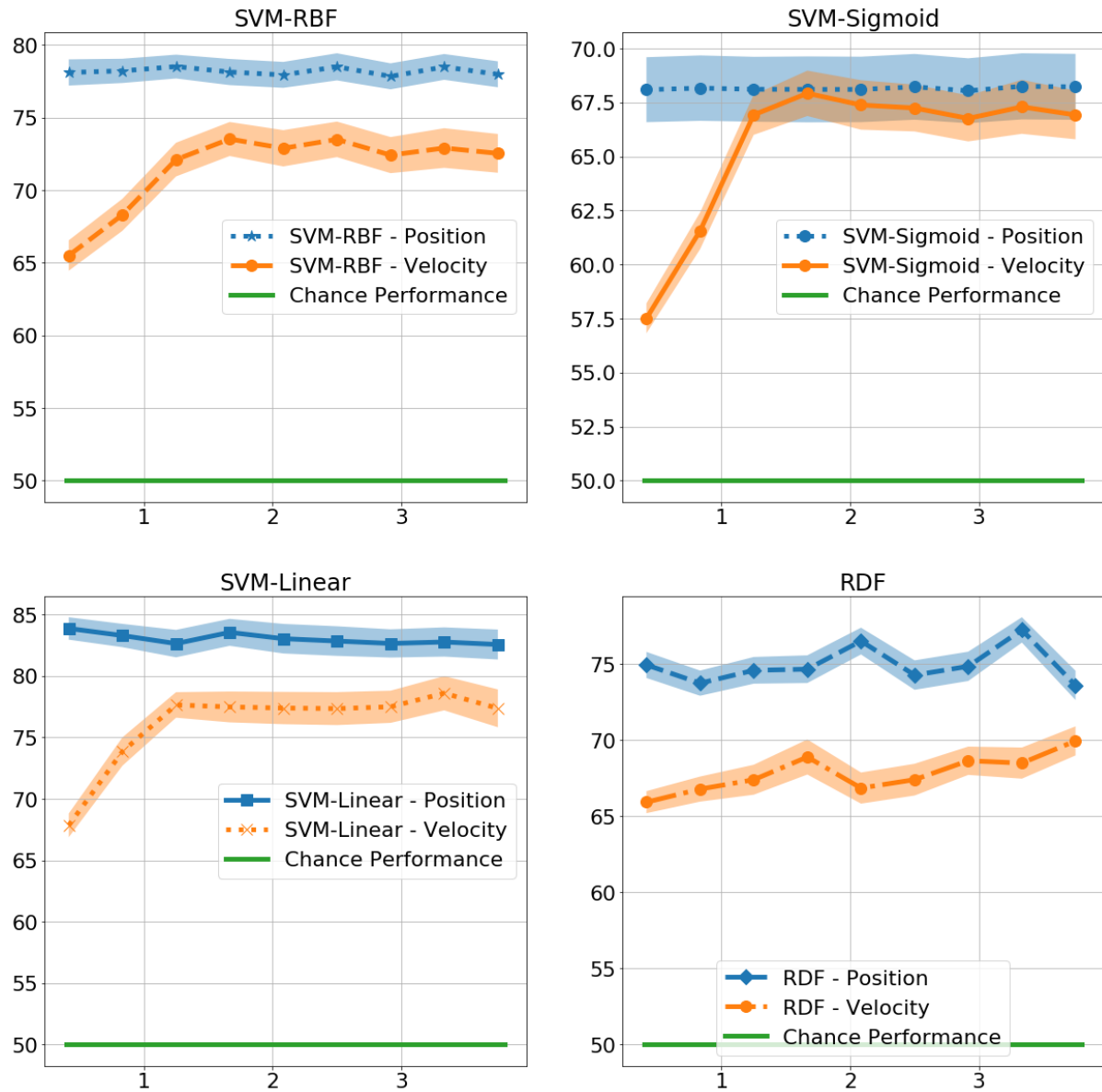


Figure 15: Correct gender classification accuracy (in %, mean and standard error) of NBM trained with position and velocity information, as a function of duration of gait considered for extracting the static representation.

Model\Gait Data	Position	Velocity	F-Test, $F_{1,19}$
SVM - Linear	83%	76%	37 ( $p<0.001$ )
SVM - RBF	78%	72%	52 ( $p<0.001$ )
SVM - Sigmoid	68%	65%	5 ( $p<0.05$ )
RDF	75%	68%	149 ( $p<0.001$ )

Table 8: Correct gender classification accuracy (in %) by NBM trained with position and velocity gait information.

Fig. 15 represents the gender classification performance of the NBM trained with the first-order statistics of the velocity profiles of the joints, as a function of the duration of gait data utilised to derive the static representations. As observed, the performance of the NBM actually significantly upon training with velocity data, compared with the position data (details provided in Table 5). SVM-Sigmoid demonstrates a significant increase in performance with increasing duration of gait data provided, ( $F_{1,18}= 4.5$ ,  $p<0.05$ ) with the best accuracy of 67%. Additionally, the change in performance is not statistically significant in other SVM and RDF models. This behaviour goes against the expected biological behaviour and the results demonstrated by the LSTMs, denoting a loss of performance in a form of data which is biologically more conducive to gender classification.

## 4.5 Experiment 5: Biomimetic Models with Two-dimensional Input

LSTM networks have demonstrated their biomimetic nature not only architecturally but also behaviourally, through close resemblance with the HP, making it conducive to draw parallels between the learning mechanisms. However, one could argue a gap in parallelism of comparison as the HP observed the moving point animations on a two-dimensional screen for gender classification while the LSTMs were provided with three-dimensional motion information. This experiment trains the LSTMs with two-dimensional gait information to form a direct comparison with the HP, without assuming any depth inference capabilities of HP.

### 4.5.1 Method

#### 4.5.1.1 Biomimetic model (BM)

The LSTM model architecture is similar to the previous experiment. However the number of inputs are reduced by a third, owing to the loss of the z-components of the joint trajectories. The weights initialisation, training and testing regime and class classification protocols are maintained. The only difference is brought about by the difference, or lack thereof, of the data input being provided.

#### 4.5.1.2 Data Input

Experiment 2 represented gait as a temporal evolution of the positional joint trajectories. The three-dimensional trajectories of each of the 20 tracked joints were concatenated to form a vector

representation of a static frame with 60 values (20 triplets). This experiment maintains the same data style but hides the z-component, modifying the vector representation to have a 40 values (20 doublets, each doublet representing the X and Y coordinate of the joint trajectory). As mentioned in previous experiments, the gait information is varied from 0.4 sec to 3.8 sec in steps of roughly 0.4 sec (10 frames). The resulting data is normalised between [-1, 1] and standardised to have zero mean and unit standard deviation.

#### 4.5.2 Results

As shown in Fig. 16, there is no statistically significant difference in the outcomes of the LSTMs trained with three- and two-dimensional joint trajectories, in corresponding position and velocity information. In other words, omission of depth information didn't impact the performance significantly. However, the difference between the models trained with corresponding 3D and 2D values is significant across all durations ( $F_{1,18} = 6$ ,  $p < 0.05$ ) for LSTM - 2D Position and LSTM - 2D Velocity, with accuracy of 76% and 80% respectively. Notably, the performance of the models trained with 3D and 2D velocities are significantly higher than the models trained with the corresponding position representations. The two-dimensional models also demonstrate the same unique gender sensitivity trait possessed by HP and BM trained with three-dimensional representations of gait, further supporting the close resemblance of the LSTM models with HP.

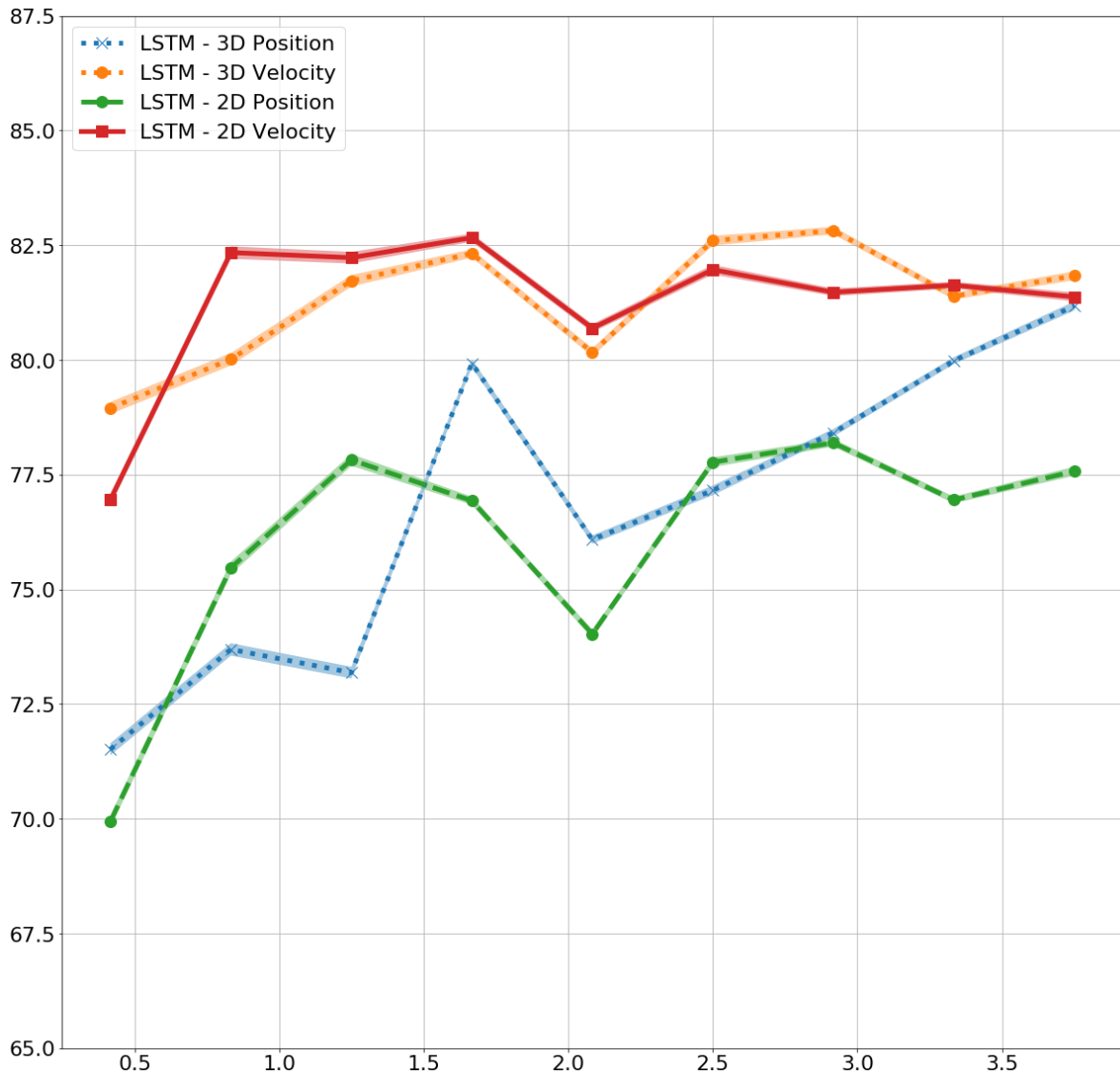


Figure 16: Comparison of correct gender classification accuracy (in %, mean and standard error) using biomimetic LSTM models trained in three- and two- dimensional position and velocity representations of the joint trajectories

## 4.6 Discussion

As shown in Experiment 1, HP was able to identify gender from gait with significantly above chance performance from moving dots presentations of joints, while conforming to existing HP literature. There is a significant increase in gender classification performance between 0.4 sec and 3.8 sec of stimulus exposure duration, further establishing H's ability to learn from temporal patterns. The increased gender sensitivity at 1.5 seconds is attributed to the prevalence of dynamic, velocity-based cues at the phase of the step cycle corresponding to that time [7], demonstrating the preference of humans towards dynamic velocity-based cues compared to structural position-based cues for gender classification. Without the preference for dynamic cues, one would have expected a non-negative monotonically increasing performance with additional temporal information. Results from the psychophysics study encouraged its comparison with results obtained using the dataset with computational models.

Experiment 2 showed that the BM that were trained on the first-order statistical representations of the temporal signals performed significantly better than chance and the corresponding models trained on clinical gait metrics performed at or below chance performance. Notably, there was no significant change in performance with increasing duration of exposure of gait input, unlike both HP and BM. Thus, denoting a deviation from human-like models. In addition, there was no change in sensitivity to gender classification with increasing availability of information at different phases of the step cycle either. Both the above characteristics have been observed in HPs and LSTMs, suggesting a dissimilarity of the NBM from a common connectionist learning mechanism shared between HP and LSTMs. HP also possess the trait of being sensitive to dynamic velocity-based cues for gender classification. While NBM demonstrate their ability to classify gait with impressive accuracy, their non-compliance with human-like behaviour reduces their chances of consideration for the artificial gait classifier in terms of:

- Need for static representations of gait, leading to loss of temporal data which may not necessarily translate to non-gender-based classification objectives
- Need for human intervention (similar to an expert system) to engineer the necessary features given a different gait classification objective
- Agnosticism to additional temporal data, leading to deviation from human-like perception which learns from increasing presence of data
- Lack of change of gender classification sensitivity at different stages of the gait cycle

However, humans are also known to be sensitive to dynamic cues of gait. Thus, in order to detect the existence of the biological trait of sensitivity to dynamic, velocity-based cues in the artificial classifiers, the Experiment 3 trains the models on velocity cues exclusively, to evaluate the change in performance. Experiment 3 also aims to establish the degree of conformity between HP and different approaches to learning in computational models. A significant increase in percentage correct gender classification, based on velocity-based cues (compared with position-based cues) established a human-like behaviour in the computational models.

Experiment 4 showed that training BM and NBM with joint velocities produced highly differing results when compared to the corresponding models trained with joint positional trajectories. The result of the NBM goes against the notionally established human dependence on dynamic velocity-based cues for gender classification. However, the results obtained through the biomimetic LSTM networks not only conforms to the expected biological behaviour, but also demonstrates the shared gender sensitivity trend observed at different phases of the walk cycle between the HP and BM. However, one could argue that the provision of three-dimensional gait information to the BM but the two-dimensional screen-based input to the HP could cause a gap in direct comparison between the models, making it difficult to draw parallels between the learning mechanisms. Towards this objective, Experiment 5 trained the biomimetic LSTM model and tested on two-dimensional gait information by omitting the depth information, which the HP had to infer from the screen.

As demonstrated in Experiment 5, the loss of depth information didn't cause any significant change in performance accuracy; all the traits that were shared with HP and corresponding BM trained with three-dimensional gait information were maintained, further supporting the close resemblance of the artificial model with HP. In case of the ML models, ten randomly initialised

(untrained) models undergo training and serve as ten independent artificial observers of gait. Thus, in case of intrinsic feature (gender) classification from gait, taking a confidence level of 95%, sample size of 10 and a worst case classification accuracy percentage of 50%, provides a confidence interval of 30.99%. In the case of intrinsic feature classification HP, taking a confidence level of 95%, sample size of 21 and a worst case classification accuracy percentage of 50%, provides a confidence interval of 21.39%.

In summary, biomimetic models share the following characteristics with HP, which the non-biomimetic models either do not share or have the opposite trait of: (1) Increase in gender classification performance with increasing temporal availability of gait information, (2) Preference towards dynamic velocity-based cues as opposed to structural position-based cues for gender classification, leading to higher performance in the former data type, and (3) Unique trend of gender sensitivity during different phases of the walk cycle. The results support the existence of a memory-based generic learning system in HP of biological motion. Additionally, the close resemblance in performance between HP and BM confirms the ability of the computational learning models in emulating human learning dynamics. This result is further established by humans' ability to acquire knowledge on gait classification which are not innate, for clinical and security identification purposes.

From an application standpoint, the results encourage the potential of using BM for gait classification. Although the paper uses gender classification as the gait classification objective, the resemblance with HP widens the scope of application to non-gender related classification tasks as well. The results could also be further improved through a more conservative approach to cross-validation such as the leave-one-out cross validation. The improvements to be resulted from modifying the BM and training configurations is yet to be documented. Interestingly, the availability of 3D skeletal data delivers itself to test the effect of rotation of the skeleton and its correlation with gender identifiability for HP. The same should not affect the performance of the biomimetic as a simple pre-processing rotation step could correct for any misalignment in global skeletal configuration. While treadmill walking may not be congruent with over ground walking, the experiment establishes the ability of the machine-based models in extraction of relevant features from spatiotemporal skeletal data, given an objective which the model possesses no prior knowledge of.

The NBM on the other hand, require explicitly hand crafted features which may not be applicable to a generic classification task. As shown in Experiment 3, the clinical gait metrics representation of gait didn't possess information about gender, while the static four attribute representation possessed enough information about gender for a better than chance performance. The transferability of the same features towards a new objective (such as person identification) is brought into question and requires further experiments.

## 4.7 Conclusions

Humans are highly adept at classification of gait for a multitude of objectives, from gender classification to clinical diagnosis, while relying on a common learning mechanism of

spatiotemporal perception of gait. This thesis focuses on exploring underlying human learning mechanisms by training a number of machine learning models, each possessing a different approach to learning, and comparing results in the gait classification task of gender classification on the same stimulus set. Thus, the motivation for the analysis is two-fold. Firstly, to understand the underlying learning system in HP in making spatiotemporal classification decisions. Secondly, to find an artificial model to classify gait while maintaining a close resemblance with humans, to ensure generalisability across multiple gait classification tasks. Results indicate the existence of a generic memory-based biological learning system for spatiotemporal classification tasks that benefits from training. This result encourages the potential of utilising the biomimetic memory-based LSTM model, as the model of choice for a given gait classification task. NBM, while demonstrating high performance in the gender classification task, don't follow the traits shown by HP, making them less reliable in performing a generic gait classification task. In addition, the need for generating static hand-crafted features from human gait brings the possibility of transference of static features across classification tasks into question. Notably, the learning efficiency of LSTM models also ensures that the performance is not significantly impacted even in the case of the absence of depth information, further supporting the biomimetic memory-based model as an effective artificial gait classifier.

## Chapter 5 Biological and Biomimetic Perception of Gait

Results from Experiments 2 and 5 in the previous chapter have demonstrated the commonalities between the biomimetic model (BM) and human perception (HP), supporting the possibility of application of the BM model as a versatile, high performing gait classifier. HP and BM possess commonalities such as,

- Increase in performance accuracy with increasing visual stimulus exposure durations
- Higher sensitivity to dynamic cues compared to static cues, and,
- Qualitatively similar profile of performance in gender classification accuracy at different stages of the gait cycle. Notably, the profile is maintained in the absence of depth information as well.

A significant body of research has investigated the information that can be extracted from human gait. Johansson[67,112,113] first introduced the point light display (PLD) animation technique, in which points of light attached to the limb joints of walkers were recorded in the dark, a type of stimulus that came to be known as ‘biological motion’. This basic paradigm led to extensive study of human visual perception of these dynamic patterns [9, 10, 32, 115, 117]. HP are not only able to distinguish biological motion from sparse noise, but can also recognise the identity of walkers [18,19] and generic attributes from strategically placed sparse PLDs, such as gender [7,20,32,127,150] emotion [9,10,68] and walking direction [36,69]. Human perception and learning, however versatile, has been documented to possess biases which affect its gait classification performance. The ability to perceive certain properties from biological motion has been found to be highly dependent on whether the stimulus conforms to a canonical, upright, viewpoint; when a point-light walker is inverted vertically, gender classification performance drops to below chance levels[32,35], a phenomenon known as the inversion effect[36]. This phenomenon has been well documented in other fields of literature as well, including facial expression classification [163–165]. The mechanism underlying inversion effect remains ambiguous, with reasons ranging from an innate neural structure which predisposes HP, to a learned perception due to heavily biased learning datasets, leading to fixed notions of anatomy [35,36,69]. The detrimental effect of inversion of visual stimulus on gender classification could be attributed to the disruption of the global configural perception of human gait [36], which may possess either an innate component or a learned component or a combination of the two. Assuming gender to be a global configural feature, one can argue for its learned nature [36,69,166–168] . However, in the case of gender from gait, HP are known to rely on anatomical differences in the skeletal structures between men and women for correct gender classification. Correct classification performance is maximised in an upright walker orientation in the coronal plane [32,127]. This specificity has been attributed to the differences in body structure and dynamics of the hips and shoulders between males and females [32,127]. In addition, new born babies also show predisposition of attention to localised features of gait [169]. If one was to argue about gender classification being a local feature, based on attention to the shoulders and hips as shown by previous studies experiments [32,127] suggest local motion perception to be an innate feature of HP [69]. This ambiguity is addressed through a series of experiments on the BM model in this chapter. Considering the BM no prior assumption of an anthropomorphic



structure of the human skeleton, results similar to HP experiment might suggest the origin of inversion effect as a learned behaviour rather than an innate one.

This chapter aims to address the questions mentioned above, through a series of experiments. Experiment 1 tests for the presence of human-like susceptibility to inversion in the BM, by training and testing multiple instances of the BM model, each randomly initialised (before training). The objective is to demonstrate the presence of inversion vulnerability in the modes. Significant presence of the said effect would indicate that the inversion effect is an emergent property from learning the association between gender and the global configuration aspects of PLD-based biological motion. Additionally, a second motivation of this chapter is the development of a practical, high-performing artificial gait classifier that overcomes the observed limitations in HP. Thus Experiment 2 tests the dependence of the model on structural cues by removing them and forcing the model to learn the association between gender and motion cues only. The hypothesis being, if the models' gender classification accuracy is being hindered due to the presence of structural cues (the way it does in humans), removing them should improve the accuracy in upright stimuli and robustness to inversion. A further extension of the hypothesis is tested in Experiment 3 by providing the model with spatiotemporally modified gait. The hypothesis in this case being, if the BM is functionally similar to humans, then provision of explicit motion based cues in addition to the removal of structural cues should improve gender classification accuracy in upright skeleton and increase robustness to inversion even further. The dataset used in all the experiments in this chapter is the same as that used in the previous chapter (Chapter 4), also described in section 3.1.1.

As a notable corollary to the above experiments, the results of the experiments has the potential of also contributing towards the understanding of HP. In order to assess whether generic learning mechanisms could provide a plausible account for the inversion effect in humans, we assess the performance and robustness to inversion of several machine-learning models of varying degrees of similarity to theorised mechanisms used by humans in biological motion perception. If machine-learning models that are biomimetic show similar vulnerability to inversion, this could suggest that human performance depended on similar generic learning mechanisms, rather than needing to appeal to specialised modules.

## 5.1 Experiment 1: Inversion Effect

The inversion effect is an extensively studied phenomenon in HP. The effect has been studied through multiple input methods including face inversion [163,165,169,170] and biological motion inversion. When biological motion is presented upside-down, perception is strongly impaired [36,114,164]. The effect seemed to occur irrespective of the experimental task and affected the detection of a point-light walker [164,166]. In case of gender classification from gait, when presented with the vertically inverted stimuli, humans performed significantly below chance with performance varying from 37% to 41%, with significantly higher classification confidence when responding incorrectly. In most cases, humans changed their classification of gender for the same walker when presented with the inverted stimuli [32]. While Barclays et al. maintains the global coherent shape of the walker, it has invited criticism from subsequent works because of the

synthetic nature of the stimulus [167] which seems to omit local motion information. Given the importance of local motion,[35] utilised motion-captured data on human walkers to test for the inversion effect on gender perception, resulting in a chance or near chance performance on the inverted stimuli. Retaining the same theme, this experiment evaluates the BM on metrics of accuracy and classification confidence when predicting inverted gait inputs, along with introducing the metric *classification inversion probability*.

## 5.1.1 Experiment Setup

### 5.1.1.1 Data Input

During training, the BM model was provided with the same dataset as mentioned in Experiment 1 of Chapter 4. Along with the dataset, the training and testing protocols were replicated as well. The model was trained with a sequence of frames of body poses, collectively representing human gait. Each frame represented the body pose as a collection of three-dimensional trajectory of 20 tracked joints of the skeletal system. The testing criteria, however, was different from previous chapters. The test dataset was generated by vertically mirroring the three-dimensional joint trajectories of the walkers on a horizontal plane. Essentially, the  $y$  component of the trajectory was mirrored while maintaining the values of the  $x$  and  $z$  coordinates in the generated dataset. This resulted in a mirrored centre-of-mass motion as well. The joint trajectories were further processed through size normalisation followed by standardisation with a zero mean and unit standard deviation. The most probable output was taken as the class label during classification and the absolute difference in the classification values between the output nodes was regarded as classification confidence. *Classification Inversion Probability* here is defined as the ratio between the number of walkers with opposing gender classifications between upright and inverted orientation to the total number of walkers.

### 5.1.1.2 Biomimetic Model (BM)

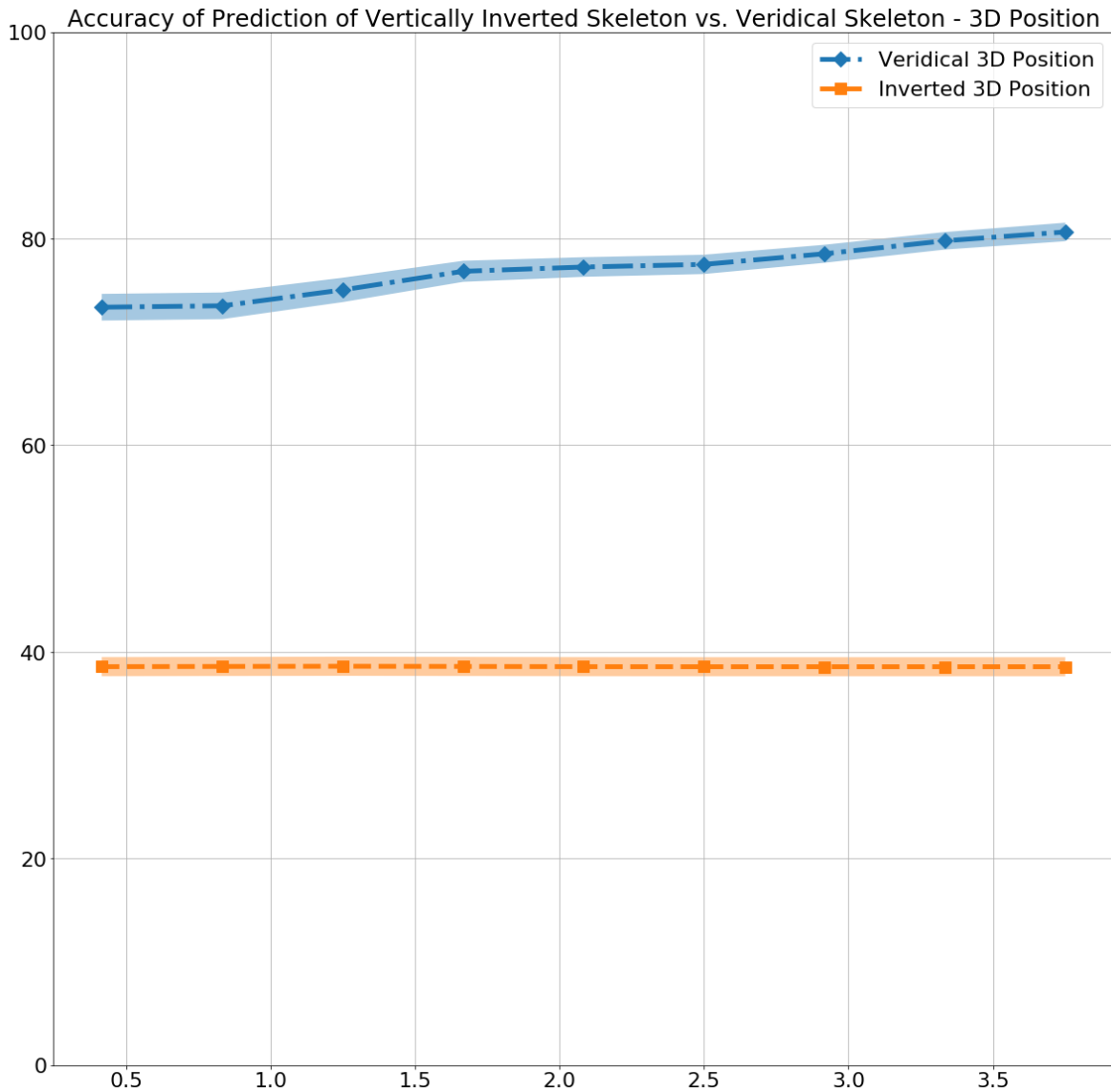
Similar to Chapter 4, the model consisted of 128 randomly initialised hidden states. The final cell state was ReLU activated [160] and connected to an affine one-hot labelled output layer, where the ‘Male’ and ‘Female’ genders were encoded as [1,0] and [0,1] respectively. The model weights were updated using an Adam optimiser [162] based on the error differentials and a learning rate of 0.001. The most probable output was taken as the class label during classification. 10 different BM models were developed by randomising the initial weight matrix before training, thus developing 10 different BM. Each of the BM could be argued to represent an independent HP undergoing training. The reason for the independent BM was also for the comparable statistics as followed in standard machine learning literature.

## 5.1.2 Results

Across all input durations, the classification performance for the BM was below chance at 37% ( $t_9 = -3.7$ ,  $p < 0.005$ ). Performance across different durations remained stable without any significant difference. Overall, classification confidence levels for correct and incorrectly identified genders showed no significant difference ( $F_{1,18} = 0.03$ ,  $0.75 < p < 1.0$ ). However, at

durations above 3.6 sec, classification confidence for incorrectly identified genders was higher than correct responses, with the difference in confidence levels being considerably below zero ( $t_9 = -2.7, p < 0.05$ ). The probability for inversion of gender classification remained close to chance performance of 50% overall. A higher probability for inversion was observed at 3.6 and 3.7 sec duration at 60%, with scope for further investigation for statistical significance. Fig. 17 visually compares the classification accuracy obtained from the model as a function of exposure duration to stimuli, both for veridical (upright) stimuli as well as vertically inverted stimuli.

Interestingly the BM classification accuracy of the inverted walkers does not change significantly with increase in exposure duration and is similar to the accuracies reported in case of HP. For during shorter durations, classification confidence levels overlap for incorrect and correct classifications. The tendency of bias for both the HP as well as BM could be attributed to an over-reliance on the hip and shoulder motions as a result of the skeletal structural differences between men and women [32, 67, 127]. The objective of the next experiment is to condition the BM explicitly on dynamics of human gait using the gender-neutral structure of the walker, to observe the change in inversion effect and gender classification performance. However, evaluation of HP trained purely on dynamic motion is unattainable as humans need to see the anthropomorphic structure to derive the motion. The BM, however can be developed and trained on synthetically generated gender-neutral structure of the walker for learning discriminating features solely from the dynamic cues. Thus, the next experiment attempts to evaluate the contribution of skeleton structure-based cues to the perception of gender from gait in the inverted stimulus. If the BM functionally emulated HP, then lack of structural cues should reduce the effect of inversion of the stimulus, and thus, increase correct gender classification accuracy. However, a reduction (or no significant change) in accuracy would argue for a deviation of the BM from HP.



*Figure 17: Gender classification accuracy (in %, mean and standard error) of the BM when tested on veridical (upright skeleton) stimuli compared with (vertically) inverted stimuli. In both cases, the model is trained on veridical stimuli.*

## 5.2 Experiment 2: Contribution of Structural Cues for Biomimetic Models

A gender-neutral posture was generated by averaging postures across all the participants of the study [127]. The resulting walker possesses a generic anthropomorphic posture within the variance of the participating subjects, which is then used for a human observation study. While this approach is perfectly suited for human observers, given their a priori assumption of an anthropomorphic model, the BM provides us with greater flexibility of experimentation because of their ability to learn a generic spatiotemporal stimulus. This allows for a higher range of postural modifications which can generalise beyond the variance of postures available in the dataset at hand. In this experiment, a new BM is developed for comparison with the model used in previous experiments and evaluated for gender classification accuracy and robustness to

vertical inversion of the walker. For the purpose of this experiment, the model trained with veridical walkers shall be referred to as  $BM_1$ , and the new model as  $BM_2$ .  $BM_2$  is trained on a gait input sequence synthesised by modifying the structure of the walkers in the existing data to reflect a gender-neutral body structure.

## 5.2.1 Method

### 5.2.1.1 Biomimetic Model

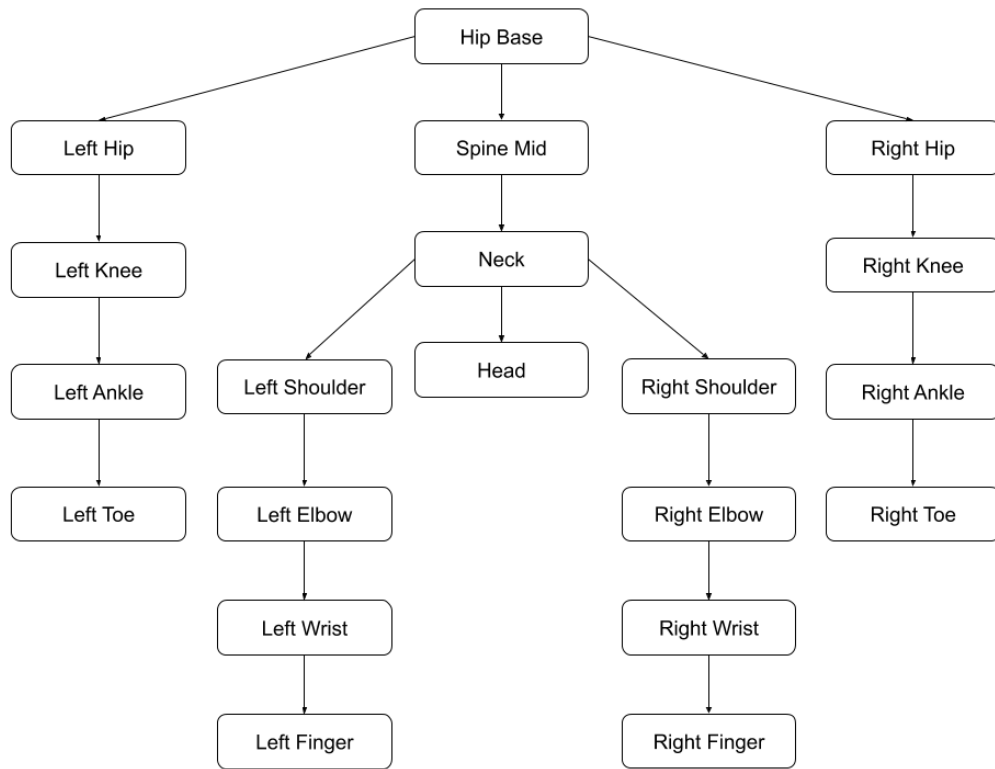
The model architecture was the same as that of the previous experiments. The training and testing of  $BM_2$  obeyed the same protocols as followed by the previous experiments, including development of 10 LSTMs with randomly generated initial weights, providing 9 degrees of freedom during evaluation of results.

### 5.2.1.2 Data Input

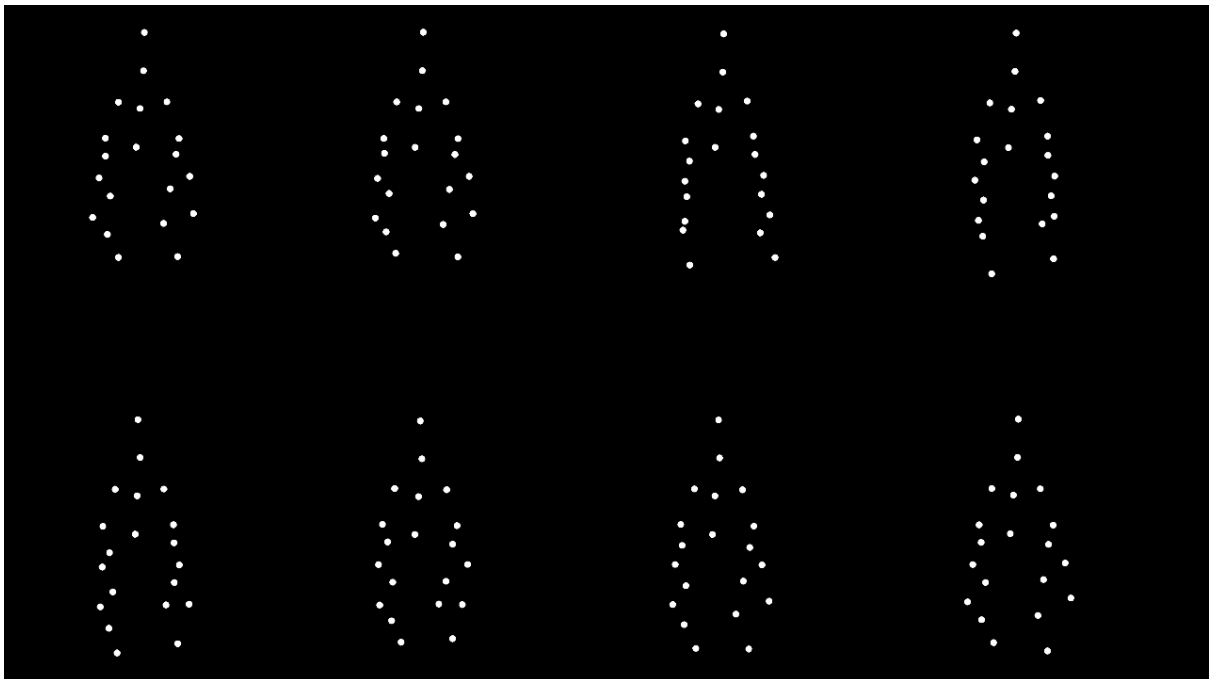
The input dataset is generated by changing the limb lengths of each of the 19 limbs connecting the 20 joints to have unit length. Fig. 17 describes the joint dependency tree of the human body as a hierarchy of attached joints, where each node in a subsequent layer of the tree is dependent on its parent node. The new three-dimensional joint trajectory of the gender-neutral walker is determined by adjusting the magnitude of the limb vector to a unit and calculating the new trajectory of the vector in the direction of the limb. The hip base is taken as the reference joint for calculating the new joint trajectories. The new joint trajectories are determined using the following steps:

$$\begin{aligned} L &= |i_{pos} - x_{pos}| \\ \hat{D} &= (i_{pos} - x_{pos}) / L \\ i'_{pos} &= x_{pos} + \hat{D} \end{aligned}$$

where,  $L$  is the limb length between parent joint  $x$  and dependent joint  $i$ , with their trajectories represented by  $x_{pos}$  and  $i_{pos}$ .  $\hat{D}$  is the unit vector in the direction of the limb vector and  $i'_{pos}$  is the new three-dimensional trajectory of the dependent joint  $i$  after the structural correction. The process is repeated for each of the 19 limbs of the body. The result of structural corrections in the static frames is demonstrated in Fig. 18. The model is trained and tested with the new dataset similar to Experiment 1 and 2.



*Figure 18: Joint dependency tree of the human body representing the parent and child joints originating from the hip base*



*Figure 19: Point light display (PLD) of a walker with unit limb lengths at different stages of the gait cycle. This stimulus dataset is used for training and testing to evaluate for its dependence on structural cues of the walker.*

### 5.2.2 Results

BM<sub>2</sub> performed at 75% accuracy in gender classification across all the input durations ( $t_9 = 7$ ,  $p < 0.001$ ). There was no significant difference in gender classification between the models BM<sub>1</sub> and BM<sub>2</sub> ( $F_{1,18} = 0.004$ ,  $0.75 < p < 1.0$ ), signifying alternative dynamic gender discriminatory cues available for the model to learn to the same degree of classification in case of vertically inverted walkers, BM<sub>2</sub> performed slightly above chance performance at 53% ( $t_9 = 2$ ,  $p < 0.1$ ). However, there was a significant improvement in gender classification of vertically inverted walkers between the two BM across all durations ( $F_{1,18} = 138$ ,  $p < 0.001$ ), suggesting a high contribution of structural cues towards the bias leading to poor performance in BM<sub>2</sub> and increase in robustness with dynamic cues in BM<sub>1</sub>. The difference in classification confidence values between correct and incorrect responses was however not significant.

Reduction in inversion effect in BM<sub>2</sub> further supported the robustness of learning gender discriminating cues from dynamic information. However, model performance remained indistinguishable. Providing dynamic and velocity cues to HP has led to an improvement in gender classification [127]. The next experiment leverages this result to perform an extensive evaluation of dynamic cues of joint velocities and acceleration derived from veridical and structurally corrected walkers with the aim of establishing the best model training and testing strategy for accuracy, robustness and generalisability to non-gender related gait classification tasks. An increase in gender classification sensitivity with dynamic cues would demonstrate the existence of a common neural module shared between HP and BM, introducing the possibility of understanding the human brain better through biomimetic systems. As a practical outcome, the result would establish the best model training and testing strategy for accuracy, robustness and generalisability to non-gender related gait classification tasks as well.

## 5.3 Experiment 3: Spatiotemporal Pre-processing Strategies for Biomimetic Models

This experiment aims to evaluate the BM performance on datasets that have been synthetically generated through the application of various spatiotemporal pre-processing steps on the standard walkers' dataset. It may be noteworthy to quickly recap that the synthetic modifications were implemented to isolate or enhance spatial and temporal characteristics of the skeleton during gait. The modifications are chosen based on known HP preferences and limitations. If BMs shared the functional similarities with HP, then the models trained with the modified version of gait are expected to improve the performance outcome both in terms of performance accuracy and by overcoming human-like limitations. The modifications are associated separately with spatial and temporal characteristics of gait i.e. skeletal structure of the walker, and dynamic motion trajectories of the joints of the skeleton. Succinctly, the spatial modification includes, (1) walkers' veridical structure, and (2) structurally modified walkers to have unit limb lengths, as mentioned in Experiment 1. Temporal modification includes, (1) position, (2) velocity, and (3) acceleration of the three-dimensional joint trajectories. The objective of the analysis is to

establish a strategy of choosing the appropriate model and pre-processing steps with a given threshold and priority of performance measures.

### 5.3.1 Method

#### 5.3.1.1 Biomimetic Model

The biomimetic architecture for building the framework of the model, training and testing protocol is the same as the previous experiments, however, differences in the spatiotemporal pre-processing steps would result in six different trained models. The variations of BM trained on various synthetically generated datasets, requires a formalised method of referring to the models. Thus, the nomenclature of the models is established according to the following pre-processing steps:

<b><i>SPATIAL\TEMPORAL</i></b>	<b>Position</b>	<b>Velocity</b>	<b>Acceleration</b>
<b>Standard Structure</b>	$BM_{pos}$	$BM_{vel}$	$BM_{acc}$
<b>Unit Limb Length Structure (gender-neutral)</b>	$BM_{pos,ull}$	$BM_{vel,ull}$	$BM_{acc,ull}$

Some of the models mentioned in the above cohort are covered in previous experiments.  $BM_{pos}$  refers to  $BM_1$ , while  $BM_{pos,ull}$  refers to the  $BM_2$  model in the second experiment.

#### 5.3.1.2 Data Input

Temporal derivatives of the gait of the standard walkers and gender-neutral skeleton walkers were used for generating the corresponding velocity and acceleration values from the position joint trajectories. The positional data was smoothed using a 5-frame moving average filter before calculating the derivatives for the subsequent frames to reduce high frequency artefacts. The data underwent normalisation and standardisation following the guidelines from the previous experiments before training and testing of the models. The procedure was replicated to generate the data for the vertically inverted walkers, by mirroring the joint trajectories on a horizontal plane passing through the centre-of-mass of the body, as described in Experiment 1. Subsequently, temporal derivatives of the trajectories provide the corresponding velocity and acceleration of the joints. The resulting datasets can be enumerated as follows:

- Position, velocity and acceleration of joints of walkers with the original skeletal data,
- Position, velocity and acceleration of joints of walkers with gender-neutral skeletal data,
- Vertically inverted data for the datasets mentioned above in 2.

The trained models are tested on the upright walkers' data for accuracy and on inverted walkers for robustness to viewpoint dependency. The corresponding classification confidence values were stored for future analysis.



## 5.3.2 Results

### 5.3.2.1 Upright (Right-Side-Up) Skeleton

The  $BM_{acc,ull}$  model trained on joint accelerations of walkers with unit limb lengths, possessed the highest overall gender classification accuracy of 82% ( $t_9 = 9.4$ ,  $p < 0.001$ , across all durations), with the performance reaching 87% at 3.75 sec ( $t_9 = 12$ ,  $p < 0.001$ ). The performance of the model increased with increasing input duration available from 0.4 to 3.75 sec, for the upright walker orientation ( $F_{1,18} = 5.1$ ,  $p < 0.05$ ). Results from the  $BM_{acc,ull}$  model further demonstrates the presence of gender specific, distinct and robust features in the dynamics, specifically acceleration, of the joint motion trajectories of the walker with unit limb length structure. Detailed results of the models have been presented in Table 6 and Fig. 20, with  $BM_{acc,ull}$  closely followed by  $BM_{vel,ull}$  with an accuracy of 82% ( $t_9 = 12$ ,  $p < 0.001$ ), on the upright orientation. A relatively significant improvement in their classification accuracy on the upright orientation ( $F_{1,18} = 3.5$ ,  $p < 0.1$ ) was observed. Structural and temporal processing achieved higher accuracy and robustness.

<b>Stimulus Exposure Duration / Biomimetic Model</b>	<b>0.4 sec</b>	<b>1.5 sec</b>	<b>2.5 sec</b>	<b>3.75 sec</b>
$BM_{pos}$	71	73	77	81
$BM_{vel}$	79	81	82	82
$BM_{acc}$	76	77	79	79
$BM_{pos,ull}$	73	79	76	77
$BM_{vel,ull}$	78	82	83	83
$BM_{acc,ull}$	78	83	86	87

*Table 9: Correct gender classification accuracy (in %) by the BM across all the stimulus exposure durations (in seconds).*

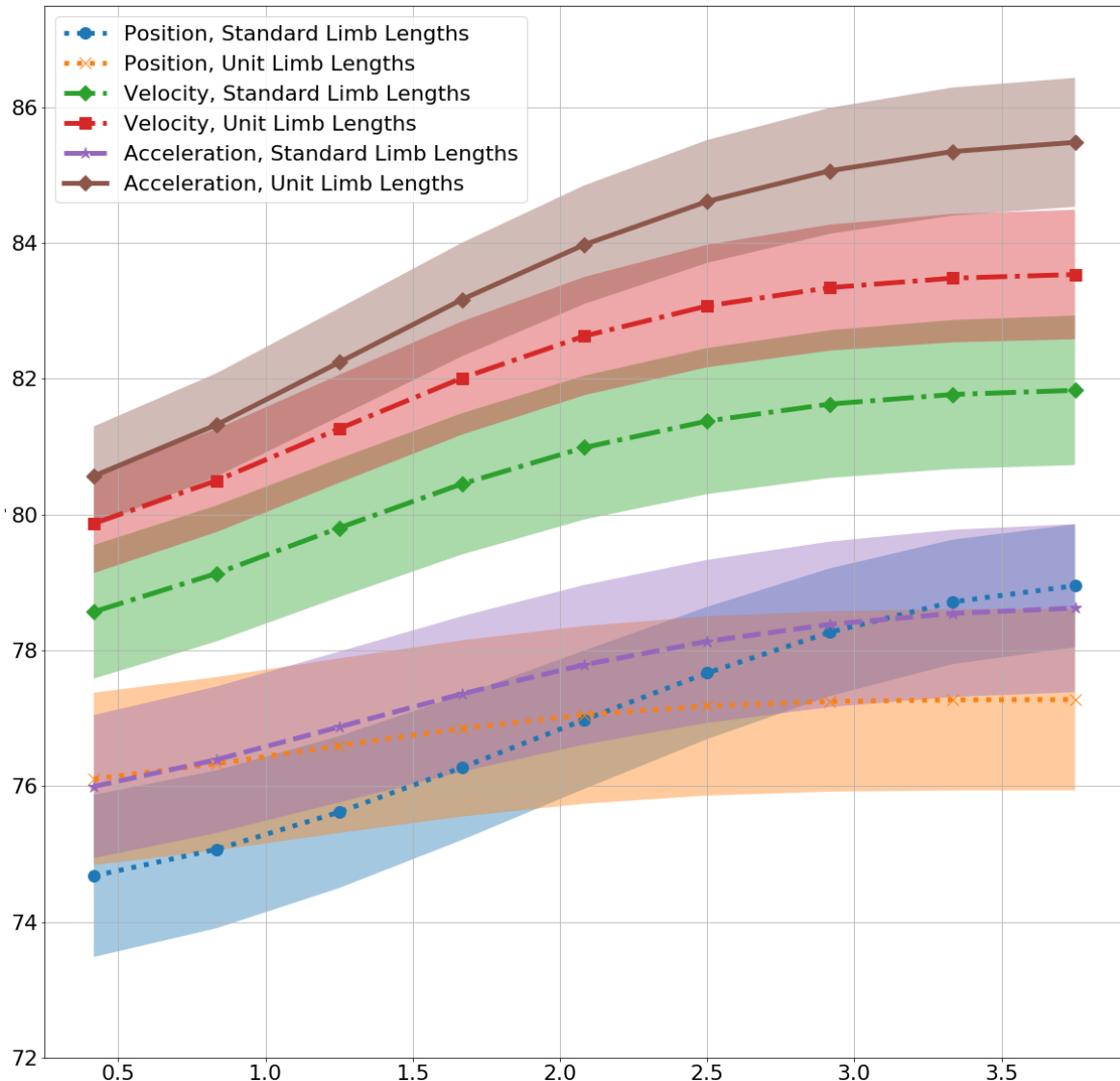


Figure 20: Correct gender classification accuracy (in %, mean and standard error) of all the BM variations as a function of exposure duration (in seconds). The values are filtered through a one-dimensional Gaussian filter

### 5.3.2.2 Vertically Inverted Skeleton

In case of vertically inverted stimuli, there was a significant improvement in classification accuracy in the models trained with the spatiotemporally modified gait, compared to the models trained with veridical gait. Coincidentally, as with the corresponding upright skeleton tests,  $BM_{acc,ull}$  had the highest performance in classification in vertically inverted walkers, with an accuracy of 81.6% across all durations ( $t_9 = 12$ ,  $p < 0.001$ ) supporting the argument of robustness of dynamic cues as opposed to structural cues. The accuracy of the model increased with increasing exposure duration available from 0.4 to 3.8 sec for the inverted skeleton orientation ( $F_{1,18} = 20$ ,  $p < 0.001$ ). Fig. 21 shows the visual trend of classification accuracy with increase in exposure duration to vertically inverted stimuli in all the variation of the models. The difference in the accuracies is most notable at the highest exposure duration of 3.75 seconds (90 frames), where  $BM_{acc,ull}$  performs at the highest classification accuracy of 81.6%, which is significantly higher than the accuracy of 37% obtained by  $BM_{pos}$  ( $F_{1,18} = 32.23$ ,  $p < 0.05$ ). Notably, the

difference in accuracies are not significant within the models trained with corresponding velocity and acceleration of gait. For instance, at 3.75 seconds exposure duration,  $BM_{vel}$  and  $BM_{acc}$  are not significantly different from each other in terms of accuracy ( $F_{1,18} = 0.00$ ,  $p > 0.99$ ) and  $BM_{vel,ull}$  and  $BM_{acc,ull}$  are not significantly different from each other ( $F_{1,18} = 0.11$ ,  $p > 0.7$ ). However  $BM_{vel}$  and  $BM_{vel,ull}$  differ significantly ( $F_{1,18} = 13.77$ ,  $p < 0.05$ ), similarly  $BM_{acc}$  and  $BM_{acc,ull}$  differ significantly as well ( $F_{1,18} = 26.92$ ,  $p < 0.05$ ). This provides the impression that each subsequent layer of spatiotemporal modification i.e. from spatial modification and temporal modification results in an increase in gender classification accuracy. The same can also be visually inferred from Fig. 21. Notably, the increase in classification accuracy through temporal modification is higher than the corresponding increase through spatial modification. For instance, at 3.75 second exposure duration,  $BM_{pos,ull}$  has a mean accuracy of 55.4% while  $BM_{vel}$  has a mean accuracy of 62.6%, which is significantly higher than the former ( $F_{1,18} = 5.45$ ,  $p < 0.05$ ). Additionally,  $BM_{acc,ull}$  has the lowest overall Classification Inversion Probability of 0.12 across all durations, as shown in Table 7. The models trained on veridical walker body structures,  $BM_{acc}$  and  $BM_{vel}$  performed similarly on the inverted walkers with an overall accuracy of 61%,  $p < 0.001$ . Thus, structural and temporal processing achieved higher accuracy and robustness. Table 8 compares the classification of accuracies between all the pairs of BM and shows the difference in the accuracies using one-way ANOVA.

<b>Biomimetic Model</b>	<b>Classification Inversion Probability</b>
$BM_{pos}$	0.52
$BM_{vel}$	0.38
$BM_{acc}$	0.46
$BM_{pos,ull}$	0.36
$BM_{vel,ull}$	0.12
$BM_{acc,ull}$	0.12

*Table 10: Average Classification Inversion Probability (fraction of test stimulus where the gender classification is inverted on inversion of stimulus) of the different variations of the BM*

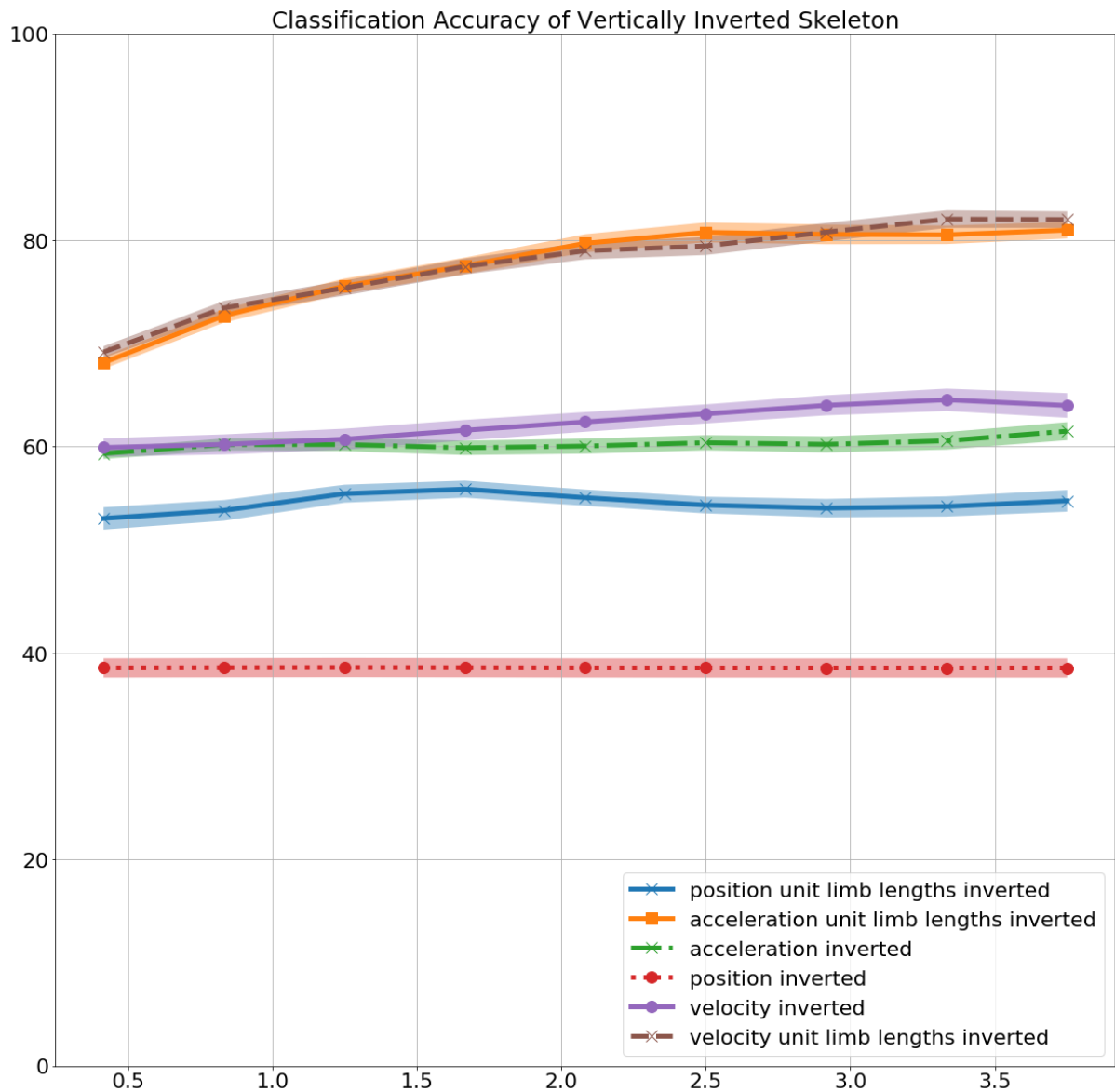


Figure 21: Gender classification accuracy (in %, mean and standard error) in case of vertically inverted stimuli using BM trained with a variety of spatiotemporally modified gait. The x-axis represents the exposure duration of the stimuli for the model, in seconds

	BM <sub>pos</sub>	BM <sub>vel</sub>	BM <sub>acc</sub>	BM <sub>pos,ull</sub>	BM <sub>vel,ull</sub>	BM <sub>acc,ull</sub>
BM <sub>pos</sub>	p=1.0	p<0.05	p<0.05	p<0.05	p<0.05	p<0.05
BM <sub>vel</sub>		p=1.0	p=0.99	p<0.05	p<0.05	p<0.05
BM <sub>acc</sub>			p=1.0	p=0.15	p<0.05	p<0.05
BM <sub>pos,ull</sub>				p=1.0	p=0.23	p<0.05
BM <sub>vel,ull</sub>					p=1.0	p=0.93
BM <sub>acc,ull</sub>						p=1.0

Table 11: One-way ANOVA results on BM pairs ( $p$ -values). Null hypothesis assumes no significant difference in the means of the classification accuracies of the BM model pairs at 3.75 sec duration of exposure to vertically inverted stimuli

## 5.4 Experiment 4: Evaluating the Spatiotemporal Pre-processing Strategies for Biomimetic Models on 2D Gait Data

One can argue for the difference in modes of evaluation of the presence of the inversion effect in the BM from the corresponding human experiments reported in literature. In literature, the inversion effect in human observers was detection through the vertical inversion of a walker on a two-dimensional screen. The information available to the observers in the experiments lacked depth information, but the corresponding information provided to the BM was three-dimensional and thus included depth. This discrepancy in stimuli could lead to a lack of similar evaluation criteria. However, upon further experimentation, the BM exhibited the inversion effect in the absence of depth information as well. Generalising the argument further, one can argue for the successful operation of the proposed approach to the presence of depth information, which was unavailable to the human observers. To address this question this experiment is conducted to evaluate the BM on the same criteria as earlier on gait information that lacks depth information ( $x$  and  $y$  coordinates of joints only) to closely resemble the HP counterpart of the experiment. The objectives of this experiment are (1) evaluate the efficacy of the proposed spatiotemporal modifications to improving the BM on 2D data, and (2) evaluate for the presence of inversion effect in the model on 2D data and evaluate for a change in gender classification accuracy on vertically stimuli after the proposed spatiotemporal modification.

Similar to Experiment 3, an experiment was conducted to train and test the BM, but on 2D data. The setup of the experiment was the same as that of Experiment 3. The only difference was in terms of the data which the BM operated on. Unlike the previous experiment, this experiment trained the BM on gait data represented as a spatiotemporal evolution of the 2D ( $x$  and  $y$ ) coordinates of the 20 salient joints of the skeleton. As with Experiment 3, the model was trained on upright skeletal coordinates and was tested on the vertically inverted skeletal coordinates for the inversion effect. Six different variations of the models were trained, based on six variations

of spatiotemporal modifications (as mentioned in the methods of Experiment 3). The rest of the experiment protocols remained the same.

## 5.4.1 Results

### 5.4.1.1 Upright (right-side-up) Skeleton

Similar to its 3D counterpart in Experiment 3,  $BM_{acc,ull}$  classified gender with the highest mean accuracy of 86.82% at 3.75 second exposure duration, significantly higher than chance ( $t_9 = 13.1$ ,  $p < 0.05$ ) more accurate than  $BM_{pos}$  ( $F_{1,18} = 4.34$ ,  $p = 0.05$ ). The other four variations of the BM, namely  $BM_{vel}$ ,  $BM_{vel,ull}$ ,  $BM_{pos,ull}$  and  $BM_{acc}$  while performing significantly higher than chance were not significantly different from . Thus, similar to Experiment 3, the combination of all the spatiotemporal modifications results in the most accurate gender classification by a significant margin. The results are visually represented as a function of exposure duration (in seconds) in Fig. 22.

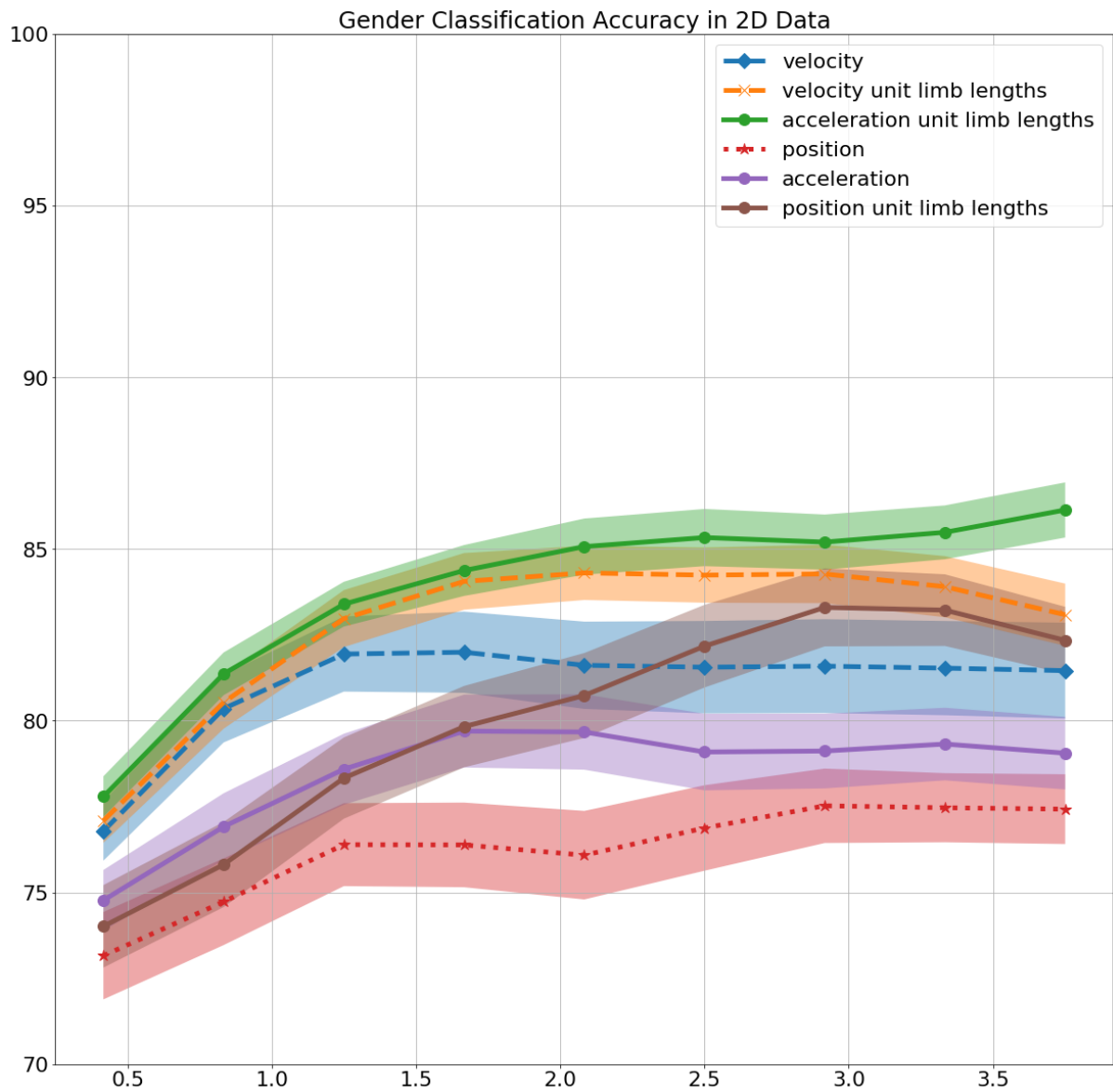


Figure 22: Gender classification accuracy (in %, mean and standard error) as a function of exposure duration (seconds) using BM trained using six varieties of spatiotemporal modifications on 2D gait

### 5.4.1.2 Vertically Inverted Skeleton

The results are visually represented in Fig. 23.

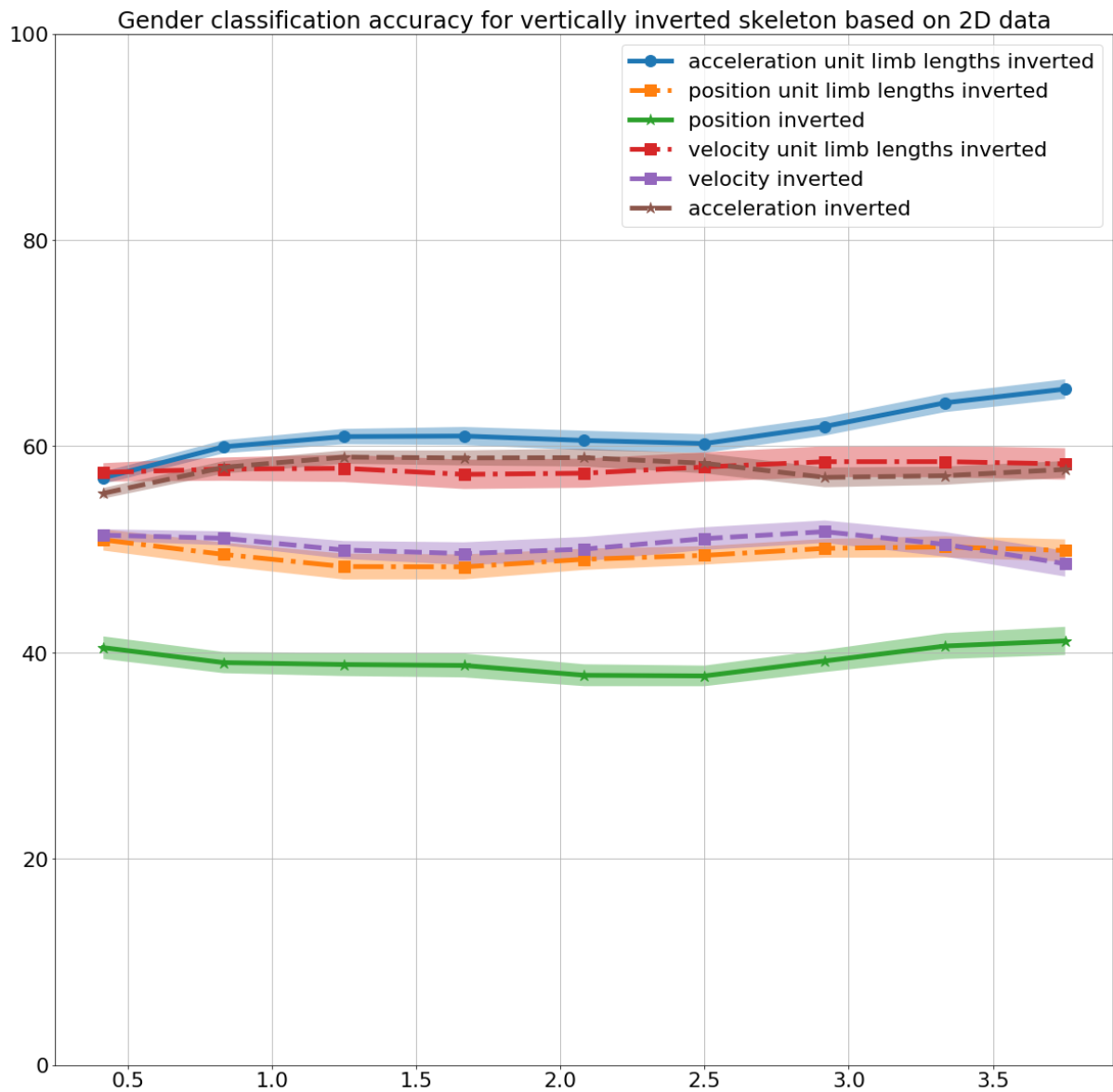


Figure 23: Gender classification accuracy of vertically inverted skeleton (in %, mean and standard error) using BM trained and tested on two-dimensional gait data (without depth information). Classification accuracy (%) as a function of exposure duration (seconds) of gait stimuli

As confirmed visually in Fig. 23, the model exhibits inversion effect in the absence of depth information as well with a mean classification of 41.07% at a 3.75 second exposure duration, below chance performance ( $t_9 = -1.9, 0.05 < p < 0.1$ ). While the performance is not significantly below chance, there is a significant difference in classification accuracy on upright skeleton and inverted skeleton with a mean gender classification accuracy of the former at 77.57% and the latter at 41.07% ( $F_{1,18} = 40.93, p < 0.05$ ). An approximately 36% reduction in classification accuracy can be attributed to the inversion effect, which, as shown can be exhibited in the lack of



depth information as well. Notably, the spatiotemporal modifications on veridical gait improved classification accuracy on the models trained on the 2D representation of gait as well. For model comparison the same model nomenclature as that of the previous experiment is followed here as well. Similar to the previous experiment  $BM_{acc,ull}$  classified gender in the inverted stimuli with the highest mean accuracy of 66.29% ( $t_9 = 4.90$ ,  $p < 0.05$ ). Interestingly, unlike the previous 3D counterpart, the accuracy of  $BM_{vel}$  and  $BM_{acc}$  are significantly different from each other ( $F_{1,18} = 4.97$ ,  $p < 0.05$ ), while  $BM_{vel}$  and  $BM_{vel,ull}$  are not significantly different from each other. This suggests that the hypothesis of temporal modifications resulting in a higher increase in accuracy, compared to spatial changes, does not hold in case of models trained on 2D gait data. Nonetheless, any combination of spatiotemporal modifications result in a significantly higher gender classification accuracy in the models.

In summary, gender classification accuracy and robustness increased with every additional spatiotemporal processing step, revealing more readily available gender discriminating cues with additional pre-processing. The lack of structural influence on the joint trajectories could exaggerate the behavioural differences expressed through motion between male and female walkers leading to a robustness in the inversion of the walker. The BM also demonstrated capacity for self-learning the relevant cues without the need for hand-crafted features. The significant improvement in performance between the best performing BM and the biological model ( $F_{1,29} = 113$ ,  $p < 0.001$ ), is not surprising given the difference between  $BM_{pos}$  and biological perception, as shown in Experiment 1. The difference in performance accuracy between  $BM_{pos}$  and  $BM_{acc,ull}$  is significant ( $F_{1,18} = 13$ ,  $p < 0.001$ ), given the basic nature of the spatiotemporal pre-processing.

However, the difference in robustness between  $BM_{acc,ull}$  and  $BM_{pos}$  is significant, ( $F_{1,18} = 734$ ,  $p < 0.0001$ ). Considering the similarity in inversion effect between  $BM_{pos}$  and humans, the difference in robustness could be extended towards the perceptual bias of humans as well. Notably, when trained with the 2D variant of data (withholding depth information of the joints), no significant difference in performance profile was found while exhibiting similar biomimetic properties.

## 5.5 Discussion

The psychophysical results are consistent with previous studies in visual perception [8,67,113,127,171], suggesting a plausible biological baseline for comparing human performance with the BM. Experiment 1 reveals that a common inversion effect is shared between humans and BM in the misclassification of inverted stimuli. The emergence of the inversion effect in the BM as a result of learning the association between gender classification and biological motion supports the hypothesis of the inversion effect being a learned perception. The commonality also suggests that BM learning may operate using similar global mechanisms as humans in processing biological motion. That the BM tested exhibit the same characteristic after training (even when initialised with random weights) suggests that the behaviour emerges due to convergence towards a set of weights that are optimised for gender classification, and

provides additional support for global configural perception as a learned process, as opposed to local perception as an innate one [167].

Another way to look at the same phenomenon is through the top-down (background knowledge based) and bottom-up (data driven) approach to perception. The BM do not possess a preconceived anthropomorphic notion of the human body in motion but rather build their conception through observation of individual joint motions that are then correlated with gender. This resembles a data-driven, bottom-up approach to motion perception, whereas humans may create percepts from biological motion using both top-down and bottom up approaches [164].

Experiments 2 and 3 were directed at understanding and improving the BM to become more robust than HP with the aim of developing a high-performing gait classifier. The absence of any preconceived anthropomorphics in the model is utilised in Experiment 2 using a gender neutral representation of the body structure in order to mitigate the inversion effect and to increase classification performance to progressively higher levels. The observation that humans use dynamics in preference to positional cues in identifying gender from gait [8] was leveraged in Experiment 3 to train the BM on the velocity or acceleration of the joints. Higher temporal derivatives increased the accuracy, robustness and efficiency of the models. The increase in performance, despite the potential loss of relevant information (through omission of structural cues), highlights the redundant nature of the information extracted by both humans and the machine learning models. However, the advantage of machine models lie in their lack of restriction on the type of information they can work with, which is not necessarily the case in humans. For example, acceleration of foot trajectories is perceptually important [168] but may be difficult to provide in an isolated manner to improve performance. Machine models, on the other hand, allow for testing on arbitrary modifications to biological motion vectors, whether it is to improve model performance or to extend understanding of HP. Interestingly, Fig. 20 shows a visual increase in the standard error band in accuracies in  $BM_{acc,ull}$  with increase in time duration of exposure, while the opposite is shown in the case of  $BM_{pos}$ . Understanding the underlying cause of this phenomenon calls for further experiments, which, although is outside the scope of this thesis, but is an interesting avenue of exploration.

## 5.6 Conclusions

Gender classification from human gait was used to evaluate differences and commonalities between human and artificial and learners. A biomimetic machine learning model that does not require hand-crafted features is shown to resemble in operation certain functional aspects of human performance, while additional modifications guided by HP are shown to exceed it. The results provide support for a generic, rather than a pre-tuned, learning system in human visual perception, and one that is based on global, rather than local, configural processing. The approach may allow for robust gait classification in other applications, such as for the diagnosis of clinical movement disorders. Other attributes of the walker such as age, weight, emotional state or other personality traits could be treated in a similar way. However additional data would be needed to represent such attributes. Given an extended database, it is straightforward and analogous to the gender classification problem to train and test the model for other attributes that may be represented in walking patterns.

Dominance of dynamic information, presence of an inversion effect and increase in performance with increase in data availability are some of the aspects of HP that are shared by the machine models. Results demonstrate the potential of BM as an automatic gait classifier with performance significantly exceeding human observation. As a clinical application, the trained BM could be used as an assistive technology for clinicians in medical gait analysis. This is reserved as future work.

As with any predictive model, the BM could possess biases of their own; however, the significantly higher performance and difference in known biases allows the opportunity for the models to be utilised either in isolation or together to compensate for each other's biases.

# Chapter 6 Extrinsic Feature Classification Using Biomimetic Models: Emotion Classification from Human Gait

In the previous chapters, the biomimetic machine learning model was able to classify gender of a human walker with an accuracy that was significantly more than human observers' performance. While the ability of the model was noteworthy, the classification outcome i.e. gender as a gait feature, is intrinsic to the walking subject. Generally, one's gender remains the same throughout one's lifetime. The proposed paradigm to improve the model was applicable in the particular classification outcome. However, the generalisability of the approach would be highlighted if it is applicable in a classification outcome that is conceptually different from the ones that has been evaluated on so far. Thus, this chapter focuses on the classification of gait features that are extrinsic and short-lived, using the same approach evaluated earlier. The inclusion criteria for the extrinsic gait feature are, (1) the availability of established human perception (HP) literature demonstrating the ability of humans in classifying the extrinsic feature from gait, and, (2) the potential of the feature to be externally induced and removed, possibly in a matter of a few minutes. Adhering to the above, the expression of the state of human emotion through gait is taken as a proxy for classification of the extrinsic gait feature. Extrinsic feature classification is a considerably harder problem to classify in comparison to gender, given the variety of ways in which humans display emotion and the lack of substantial studies correlating gait and authentic emotional exhibition. This chapter acts as a preliminary evaluation of the proposed paradigm and marks the beginning of a more detailed study.

Previous studies of emotional state classification from gait have suggested an influence of heuristically chosen local features, such as head inclination, arm swing or heavy-footedness [12–15, 17, 129]. However, the problem with such predetermined feature extraction is that, in principle, a very large number of such local features can be defined. Testing all the possible pre-processed features would require a lot of time and resources, essentially violating the constraint that has been raised in the previous chapters i.e. minimisation of the expert human intervention and the need to handcraft features. Thus, in addition to evaluating the viability of the proposed paradigm, the experiments in this chapter also evaluate the ability of the biomimetic model (BM) in learning the necessary skeletal gait. Additionally, to maintain the sanctity of correspondence with HP, any global dimension reduction of gait was also avoided [33]. The gait dataset used for this chapter was collected in the US as detailed in the methods chapter (section 3.1.2). The dataset contains skeletal gait representations of twenty-two subjects (6 males, 16 females) consenting healthy adults, with visually induced emotions using images from the IAPS dataset.

## 6.1 Experiment 1: Classification of Emotion from Human Gait Using Biomimetic Models

In the experiment design, we emphasise and address the problem of authenticity of emotions being expressed through gait. Most previous literature used professional actors for motion

capture of gait, while the actors portrayed various emotions [12, 13, 17, 33, 34]. Few studies hired students with some acting experience and instructed them to feel the emotions while they walked [33,34], while other studies invoked the emotion either by asking them to recollect a life situation that made them feel a certain emotion[17,33,34] or read a passage designed to influence one's emotions in a particular way[12]. The authenticity of emotions felt during different studies could deviate from natural emotion and may manifest differently in gait in unnatural ways. For example, we believe that the high classification accuracy of emotions in case of professional actors' gait could be attributed to unnaturally exaggerated movements. The students who were asked to feel a certain emotion could also have exaggerated body movements, considering their objective was to ensure the expression of certain emotions. Adhering to the theme, the experiment was designed to ensure authentic expression of emotion through gait, through visual stimulus only without providing any explicit instruction to feel a certain emotion. All changes in gait due to emotional expressions are natural. In other words, the subjects were not asked to act according to a certain emotion. In fact, there was no explicit instruction provided in terms of exhibiting a certain emotion.

### 6.1.1 Data Analysis

#### 6.1.1.1 Data Pre-processing

The three-dimensional trajectories of each of the 20 tracked joints were concatenated to form a vector representation of a static frame with a cardinality of 60, representing the location of the head, neck, shoulders, elbows, wrists, fingertips, mid-back, hips, knees, ankles and toes. Gait input to the model consisted of a sequence of vector representations of subsequent static frames, sampled at 24 frames per second. Joint trajectories were size normalised [127] and standardised with a zero mean and unit standard deviation.

In addition to the above dataset, the BM was also evaluated on synthetically generated gait data through the application of various spatiotemporal pre-processing steps on the veridical walker's dataset, similar to Experiment 3 in Chapter. 5. The positional data was smoothed using a 5-frame moving average filter before calculating the derivatives for the subsequent frames. The data underwent size normalisation and standardisation following the guidelines from the previous experiments before training and testing of the models. The final pre-processed dataset included, (1) position, (2) velocity and (3) acceleration and the corresponding temporal derivatives of the three-dimensional joint trajectories for the unit limb length skeleton as (4), (5) and (6) respectively. The additional dataset was generated based on the preference of HP in velocity of the movements, rather than position itself, when categorising emotion from biological motion [13,34].

#### 6.1.1.2 Biomimetic Model

The BM architecture for building the framework of the model, training and testing protocol is the same as the previous experiments, however, differences in the spatiotemporal pre-processing steps results in three different models. Considering the classification objective of three emotions of 'positive', 'negative' and 'neutral', the BM was modified to have three output nodes, as opposed to two output nodes all the previous experiments. The final affine layer was modified to

be fully connected to three output layer nodes, as shown in Fig. 24. The nomenclature of the models is established according to the pre-processing steps mentioned in Table 9.

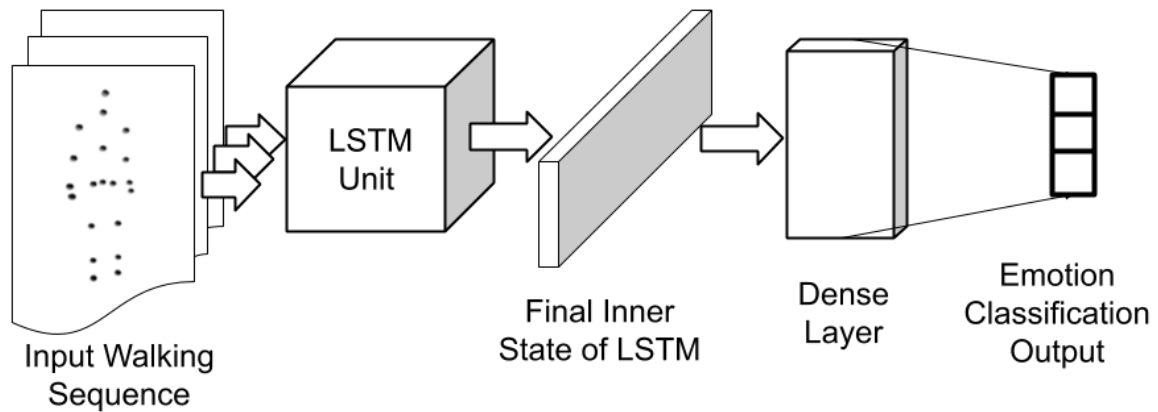


Figure 24: Model architecture of the BM for emotion classification of human gait into ‘neutral’, ‘positive’ and ‘negative’ emotions.

<b>SPATIAL\TEMPORAL</b>	<b>Position</b>	<b>Velocity</b>	<b>Acceleration</b>
<b>Standard Structure</b>	$BM_{pos}$	$BM_{vel}$	$BM_{acc}$
<b>Unit Limb Length Structure (gender-neutral)</b>	$BM_{pos,u1l}$	$BM_{vel,u1l}$	$BM_{acc,u1l}$

Table 12: Nomenclature of the various BM based on the spatiotemporal preprocessing steps used to for training and testing of the models

### 6.1.1.3 Model Training and Testing

Model training sessions included, initialization of the model weights, classification of the output probabilities based on the gait input, propagation of the classification error and updating the network weights. Model training was executed in batches of 50 and repeated for 100 epochs. Gait input sequence was provided for 2.5 seconds (60 frames) which completes two full gait cycles, to maximise the data availability to the BM to classify emotions from. During training, emotion was provided as one-hot encoded labels with neutral, positive and negative represented as [1,0,0], [0,1,0] and [0,0,1] respectively. A conservative cross-validation approach, leave-one-out policy was used for training and testing [172], to prevent lack of overestimation of performance accuracy. All six versions of the BM i.e.  $BM_{pos}$ ,  $BM_{vel}$ ,  $BM_{acc}$ ,  $BM_{pos,u1l}$ ,  $BM_{vel,u1l}$  and  $BM_{acc,u1l}$  were trained separately on position, velocity and acceleration datasets of veridical walkers and unit limb length normalised walkers datasets accordingly. Each version was initialized 10 times with different randomised weights and evaluated for statistical significance, thus providing 9 degrees of freedom for evaluating the performance accuracy. The models trained per session per duration were stored locally for future analyses.

## 6.1.2 Results

Table 10 details the results of the emotion classification task by the various BM,

<b>Biomimetic Model</b>	<b>Accuracy of Emotion Classification (%)</b>
$BM_{pos}$	43 (p<0.05)
$BM_{vel}$	41 (p<0.05)
$BM_{acc}$	44 (p<0.05)
$BM_{pos,ull}$	32
$BM_{vel,ull}$	41
$BM_{acc,ull}$	29

*Table 13: Correct emotion classification accuracy (in %) by various BM trained with a specific spatiotemporal modifications made to the original gait dataset*

The BM that were trained on the synthetically modified skeleton to obtain gender-neutral structure performed either at chance or at an accuracy that could not be distinguished with statistical significance.

Whereas, the models that were trained on the veridical walkers' skeletons performed at an accuracy significantly above chance (chance taken as 33.33% in this case). In terms of mean performance accuracy,  $BM_{acc}$  at 44% ( $t_9 = 2.36$ ,  $p < 0.05$ ) was closely followed by  $BM_{pos}$  (43%,  $t_9 = 2.51$ ,  $p < 0.05$ ) and  $BM_{vel}$  at 41% ( $t_9 = 2.41$ ,  $p < 0.05$ ). However, the difference in performance within the models was indistinguishable.

The result demonstrates the ability of the BM in being able to extract and correctly classify a temporarily induced extrinsic feature. Interestingly, this demonstrates the ability of the model in classification of emotions that are changed in quick succession, in about 5 minutes. Improving on this baseline, the next objective is to understand the preferences of the model in terms of data type for a particular emotion. While one could determine the preference from the current result, it was deemed better practice to run new experiments by isolating the emotions in pairs for a better understanding of individual emotions. Literature suggests that human walkers tend to stoop down (change in structure of skeleton) and reduce gait speed when induced with a negative emotion and move their arms more (change in motion of the skeleton) when induced with a positive emotion [13,34].

The inability of the BM models trained with gait of the gender-neutral skeleton i.e.  $BM_{pos,ull}$ ,  $BM_{vel,ull}$  and  $BM_{acc,ull}$  to classify emotion correctly, demonstrates a difference in expression of emotion between the genders. One might argue that this could be a result of a loss of information resulting from the gender-neutralisation modification. However, this argument can be disputed by the results in Chapter 4 and 5, where the performance accuracy increased after modifying the

veridical skeleton. This result corroborates with the claim that men and women express their emotions differently in terms of facial expressions [173]. However, this to our knowledge is a novel contribution towards the question, that is, if emotional expressions in skeletal motion are sex specific. Given a correlation between skeletal anatomy and gender, the results demonstrate the dependence of a particular way of emotional expression on sex, thus, on the skeletal anatomy. Thus, leading to a loss of relevant distinguishing features upon neutralisation of gender specific anatomical differences in the skeleton.

In order to understand the predominance of either structural or motion-based cues for different emotions, the next experiment focuses on a pair-wise training and testing of the two extremes of emotion. Specifically, the next experiment trains the BM models on neutral vs. positive and neutral vs. negative emotions to observe the preference of the models for either structure or motion of the skeleton for correct classification of induced emotion.

## 6.2 Experiment 2: Pairwise Comparison of Positive and Negative Emotion with Neutral, Using Biomimetic Models

### 6.2.1 Experiment Setup

#### 6.2.1.1 Data Input

The data collected for the previous experiment was separately collated to form two groups (1) neutral and positive and (2) neutral and negative. The dimensionality of the data was the same as the previous experiment and the pre-processed configurations of the dataset was the same as well i.e. temporal derivatives of the size normalised dataset to form three types of data, namely (1) position, (2) velocity and (3) acceleration of the three-dimensional joint trajectories of 20 tracked joints.

#### 6.2.1.2 Biomimetic Model

The BM model remains the same as the model in previous chapters. During training, a modification was made in the output layer to represent emotions as one-hot labels, representing both sets of emotion-pairs. In the first set, neutral and positive were represented as [1,0] and [0,1] respectively. In the second set, neutral and negative were represented as [1,0] and [0,1] respectively. Training and testing protocols were followed as per the sequence mentioned in Experiment. 3 in the previous chapter. 10 BM models were developed by initialising the weights randomly, conceptually representing 10 different perceptions, proving 9 degrees of freedom in statistical analysis of results. Three different variations of BM are trained based on three levels of temporal pre-processing steps, the nomenclature of the models provided in Table 11. As mentioned in the previous experiment, a conservative leave-one-out cross-validation policy was utilised for the training and testing protocols.



<b>Spatiotemporal Pre-processing</b>	<b>Position</b>	<b>Velocity</b>	<b>Acceleration</b>
<b>Biomimetic Model</b>	BM <sub>pos</sub>	BM <sub>vel</sub>	BM <sub>acc</sub>

Table 14: Nomenclature of the BM based on the temporal pre-processing of the dataset used for training and testing of the model

## 6.2.2 Results

### 6.2.2.1 Neutral versus Positive

The results of operating the BM models on neutral and positive emotions are detailed in Table 12.

<b>Biomimetic Model</b>	<b>Accuracy of Emotion Classification (%)</b>
BM <sub>pos</sub>	54 (p<0.5)
BM <sub>vel</sub>	66 (p<0.1)
BM <sub>acc</sub>	66 (p<0.05)

Table 15: Correct emotion classification accuracy of classification of 'neutral' and 'positive' emotion (in %) by various BM operated on temporally pre-processed human gait data

Temporal derivatives of gait data demonstrate a ready availability of emotion information for the BM to be able to learn from. Change performance, for the purpose of this experiment, is at 50%. Only the BM<sub>acc</sub> model was able to distinguish between neutral and positive emotion with a performance significantly above chance 66% ( $t_9 = 2.6$ ,  $p < 0.05$ ), closely followed by BM<sub>vel</sub>, which could classify emotions with the same average performance, but without a considerable statistical significance. However, BM<sub>pos</sub> was unable to distinguish between the emotions, with an average performance close to chance.

Notably, there was an increase of 12% in the mean classification accuracy between BM<sub>pos</sub> and BM<sub>acc</sub>. However, the difference in accuracy was not significant ( $F_{1,18} = 2.8$ ,  $p > 0.1$ ). However, a mean accuracy difference of 12% encouraged further exploration of HP literature and its application to improve BM classification. In Chapter 4, both HP and BM demonstrated a significant improvement upon increase in exposure to gait stimuli. Thus a variation of the experiment was conducted by increasing the exposure duration to 3.4 seconds (80 frames). Results of the increased exposure duration are detailed in Table 13 and Table 14. Results show a significant increase in the mean accuracy between BM<sub>pos</sub> and BM<sub>acc</sub>. Classification accuracy increased by 23%, from 43% to 65% ( $F_{1,18} = 7.4$ ,  $p < 0.05$ ). While there was a decrease in the mean accuracy from 54% to 43%, however the decrease was not statistically significant ( $F_{1,18} = 2.9$ ,  $p = 0.1$ ).

	<b>Accuracy</b>
BM <sub>pos</sub>	43%, p<0.25
BM <sub>vel</sub>	66%, p<0.1
BM <sub>acc</sub>	65%, p<0.05

*Table 16: Classification accuracy of the BM for positive-neutral emotions (in %) for stimuli exposure duration of 3.4 seconds. Chance accuracy at 50%*

	BM <sub>pos</sub>	BM <sub>vel</sub>	BM <sub>acc</sub>
BM <sub>pos</sub>	p=1.0	p<0.05	p<0.05
BM <sub>vel</sub>		p=1.0	p=0.93
BM <sub>acc</sub>			p=1.0

*Table 17: Difference in the mean accuracies (as p-values of one-way ANOVA) of the variations of BM. Null hypothesis assumes no significant difference in their means. Chance performance at 50%*

#### 6.2.2.2 Neutral versus Negative

The results of operating the BM model on neutral and positive emotions are detailed in Table 15.

<b>Biomimetic Model</b>	<b>Accuracy of Emotion Classification (%)</b>
BM <sub>pos</sub>	52 (p<1.0)
BM <sub>vel</sub>	59 (p<0.5)
BM <sub>acc</sub>	63 (p<0.5)

*Table 18: Correct emotion classification accuracy of classification of 'neutral' and 'negative' emotion by various BM operated on temporally preprocessed human gait data*

The results of the BM models do not demonstrate a performance significantly different from chance. Although the average classification increases with increasing levels of temporal derivatives, similar to the previous pair of emotions, there is no conclusive evidence of the BM model in being able to distinguish between neutral and negative emotion. The results show potential which can be investigated further by training additional models to strengthen the outcomes. In the current state of the experiment, taking a confidence level of 95%, sample size of 10 and a worst case classification accuracy percentage of 33.33%, provides a confidence interval of 29.22%.

Previous studies suggest a reduction of walking speed and downward tilting of the head (looking down) as two of the unique features of negative emotions [13,34], while no such claim was made for the positive emotional state. The treadmill forces the walker to maintain a constant walking speed throughout the negatively induced emotional state. One might argue that this artificially enforced treadmill walking could significantly mask the natural expressions of negative emotional state. Additionally, given the explicit instruction to the walkers to have a forward gaze to look at the sequence of photographs being displayed on the screen, it could artificially prevent the walkers from expressing the natural tendency to look down at the ground, thus, resulting in the inability of the BM in being to identify the negative emotion. For a deeper examination of this plausibility, a sensitivity analysis was performed, which revealed the reason for low accuracy being a very biased classification of gait towards the negative emotional state. Notably, 71% of gait provided to the model during testing was classified as 'negative', hinting at a possible dominance of certain negative emotional gait traits over neutral gait. This requires further study in terms of understanding the nuances in the numerical values in negative gait which could lead to higher values of the cost function which might lead to a harsher correction of classification in case of a wrong classification of the negative emotional state. While the same may not be true for the neutral emotional state, leading to a forced prejudice of the BM model in identifying the majority of gait samples as negative to prevent large penalties due to wrong judgement. Future work shall explore the BM biases and prejudices in judgement based on the values of the cost functions and corresponding corrective adjustments to the weights. An ideal backpropagation function shall take the same into account to ensure a normative range of errors in spite of the data sample provided.

### 6.2.3 Discussion

The series of experiments in this chapter evaluated the ability of the biomimetic BM model in being able to classify emotion, a transient, extrinsic feature that is externally and visually induced for a few minutes, from human gait. The mode of induction of emotion was chosen to ensure a more authentic and natural expression of emotion in the walkers' gait. Emotion classification was chosen as an extrinsic feature of gait, which can be argued to be a proxy for various such types of properties, such as clinical conditions, weight etc. The BM model was able to classify all three emotions with an overall accuracy significantly higher than chance. Thus demonstrating the ability of the model in being able to classify extrinsic features of gait, in addition to the intrinsic feature of gait from previous chapters (gender). Results from the binary-classification task for positive and neutral emotions further supported the bidirectional learning paradigm for improving the accuracy of the BM. Leveraging motion based cues over structural based cues results in an increase of 12% ( $F_{1,18} = 2.8, p > 0.1$ ). However the difference was not significant. Further, application of the shared aspect of increase in accuracy with increase in exposure duration between HP and BM (see Chapter 4) resulted in a significant improvement of 23%, from 43% to 65% ( $F_{1,18} = 7.4, p < 0.05$ ).

A deeper analysis of the model preferences revealed a similar preference to skeletal motion (as opposed to skeletal posture) as that of HP when distinguishing positive emotional affect from neutral emotion, corroborating with existing human psychological literature. Simultaneously, the

inability of the model in being able to distinguish between the negative and neutral emotion shows a deviation of the biomimicry from established human psychological literature. However, one could argue that the experimental conditions and constraints restricted the natural expression of the negative emotional affect in gait, leading to the inability of the BM in being able to distinguish the emotion. Analysing the plausible ways of explaining this anomaly in result lead to the discovery of a mode of analysis that could explain not only the biases in outputs of the BM, but also, lead to the design of a better cost function. This aspect of the model shall be explored as part of future work on the development of the model.

## Chapter 7 Discussion

This thesis aims to streamline the process of selecting and improving machine learning (ML) models, particularly biomimetic models (BM), for classification of a task which humans have already been shown to be adept at. Classification of human gait is taken as the exemplar classification task. The streamlined process aims to translate the insights gained from human perception (with regards to the specific task) onto the ML models, to make intuitive and human understandable changes to the models and improve their performance. Towards that objective, the thesis proposes and tests the validity of a novel bidirectional learning paradigm for improving the ML using established human perception preferences and limitations in the classification of gait. In addition to improving performance of the ML models, this approach also aims to increase understanding of human perception (HP). The validity of the proposed paradigm for improving the BM is established through the classification of two conceptually different classification objectives, namely an intrinsic feature, which tends to remain constant over one's lifetime and an extrinsic feature of human gait, which tends to change frequently, even in a matter of minutes. In this case, gender is taken as the exemplar intrinsic feature and human emotion is taken as the exemplar extrinsic feature. In theory, there are countless methods that can potentially improve the classification accuracy of the BM, requiring computational expertise, prior knowledge and intuition in the domain of application. In practice, however, given limited time and computing budget, the proposed approach could narrow the search space of such methods significantly.

In order to establish the validity of the proposed approach, numerous experiments were executed to, (1) select a biomimetic machine learning model (BM), (2) apply the approach by modifying the BM according to known human perception preferences and limitations and compare its performance with the unmodified BM, and (3) test the same approach on a conceptually different classification objective to test the validity again. The experiments in Chapter 4 help identify a viable BM through a broad comparison of numerous ML models to human performance, specifically in terms of gender classification. The resulting BM from Chapter 4 is then modified in Chapter 5 according to the proposed bidirectional approach with the aim of improving its performance. Results from the initial experiments in Chapter 5 discover a shared limitation in terms of gender classification between HP and BM, establishing an aspect of functional similarity between HP and BM to be improved upon in the subsequent experiments. Results of the latter experiments in Chapter 5 which modify the BM, not only show a significant improvement in terms of performance accuracy, but also a significant reduction in a shared limitation between BM and HP. The discovery of the shared limitation and subsequent alleviation of it in the BM provides a potential explanation for the presence of the limitation in HP which is discussed later in this chapter. After establishing the validity of the proposed approach in terms of gender (intrinsic feature) classification, Chapter 6 applies the same approach, but in terms of classifying human emotion (extrinsic feature) from gait. Results from the experiments in Chapter 6 further establish the validity of the approach. Despite the increase in classification accuracy, the significance of the improvement is modest, in comparison to the results from gender classification. The details of the results of all the experiments, the insights gained and their relation to the study's objectives (see Chapter 1) are discussed below.

The experiments were executed in three phases. Firstly, in Chapter 4, a broad comparison was performed under controlled conditions of non-biomimetic machine learning models (NBM), biomimetic machine learning models (BM) and human perception (HP), on the task of gender classification from gait. In the case of NBM, support vector machines [76] with multiple kernel functions [181] and decision forests [183-186] utilising Gini impurity [79] were considered as the exemplar NBM for the study. Whereas, in the case of BM, the Long-Short Term Memory cell [84] was considered as the exemplar BM. Results demonstrate shared aspects of classification between BM and HP. The shared aspects include an increase in accuracy with increase in stimulus exposure duration; higher accuracy in velocity-based cues compared to position-based cues, demonstrating a preference of motion-based compared with structural cues, and; a qualitative similarity in gender classification accuracy as a function of exposure duration. Notably, NBM did not share any of the above aspects with either BM or HP. In the case of NBM, classification accuracy did not change significantly with change in exposure duration. Upon provision of velocity-based cues, accuracy decreased significantly, demonstrating a deviation from the aspects shared between BM and HP. Thus, while intriguing parallels between human (HP) and biological models (BM) were apparent, a deeper investigation would be needed to account for differences in the present context. Notably, a different paradigm of neural networks based on spiking neurons [31] exists, that mimics low level biology rather than at a behavioural level as RNNs and CNNs do, which also has a promising potential for exploration.

In the second experiment phase, described in Chapter 5, the BM was evaluated for the presence of a well-studied effect in HP, which has also been reported to negatively affect H's gender classification accuracy; the inversion effect [8]. Results showed the emergence of a human-like inversion effect in BM, as a result of training. Applying findings from human perception to BM not only mitigated the magnitude of the effect but also improved its accuracy significantly in both upright and vertically inverted gait stimuli: mean gender classification accuracy improved by 6%, from 76% to 82% ( $F_{1,18} = 12, p < 0.05$ ) for upright stimuli, and by 45%, from 37% to 82% ( $F_{1,18} = 20, p < 0.05$ ), for inverted stimuli.

Lastly, the third experimental phase is described in Chapter 6, where the validity of the proposed approach was also evaluated on an extrinsic feature of human gait, through the classification of emotional state entirely from the pattern of gait. Existing literature on classification of affective emotional state using HP, insists on the role of motion based cues over structural cues. Thus this was taken into account for the bidirectional approach for improving the accuracy of the BM. This in combination with the shared aspect of increase in accuracy with increase in exposure duration between HP and BM (from Chapter 4) was taken into account. Application of this approach (providing explicit motion based cues of gait with long exposure duration) resulted in a significant improvement of 23%, from 43% to 65% ( $F_{1,18} = 7.4, p < 0.05$ ) in the binary classification task of positive and neutral emotion. Notably, the literature on affective state perception relies on either amateur or professional actors, acting under instructions from the experimenter [12–17,37,68]. By contrast, in the experiments reported here the emotional states were visually induced naturally by showing subjects standardised affective images and without giving them any explicit instructions. These results are therefore the first demonstration of classification of naturally induced (as opposed to performed) emotion from gait. The successful

automatic classification of gait changes in response to extrinsic features, in this case of emotional affect, may provide fruitful ground for exploring the effects of other extrinsic features in terms of both the capabilities of automatic BM classifiers and the physiological and psychological behaviours in people (HP).

From the above, one can argue for shared aspects of perception between HP and BM particularly in terms of gait, given the functional commonalities mentioned and biomimetic structure of the BM. However, despite the commonalities, the BM is not intended to fully mirror neuronal structures in HP, especially given the simplicity of its structure in comparison to HP. Several studies have formulated more complex models that mimic the neuronal pathways of visual perception in HP more closely [27, 70, 89, 90]. The studies have mimicked the underlying neural network for biological motion perception which have been claimed to be responsible for perception of gait, and the BM does not parallel this complexity and sophistication. However, the shared common functionalities between HP and BM, given the orders of magnitude of difference in the models in terms of scale and complexity, certainly enables the BM to be used as a practical automatic gait classifier. The applicability of the BM given a limited computing budget and availability of training data also supports its pragmatism. Additionally, given the computational nature of the model there are no prior assumptions of an anthropomorphic representation of human gait, allowing potentially for the useful exploitation of the model for other classification tasks in any multidimensional time series as a testable hypothesis than can be explored in further research. These results encourage the application of the BM to problems outside of gait classification that may perhaps require a functional approximation of biological neural substrates. Potential applications include intelligent robotics [174], smart prosthetic devices [56, 63], clinical diagnosis [25, 39, 102–104] and gait rehabilitation [175].

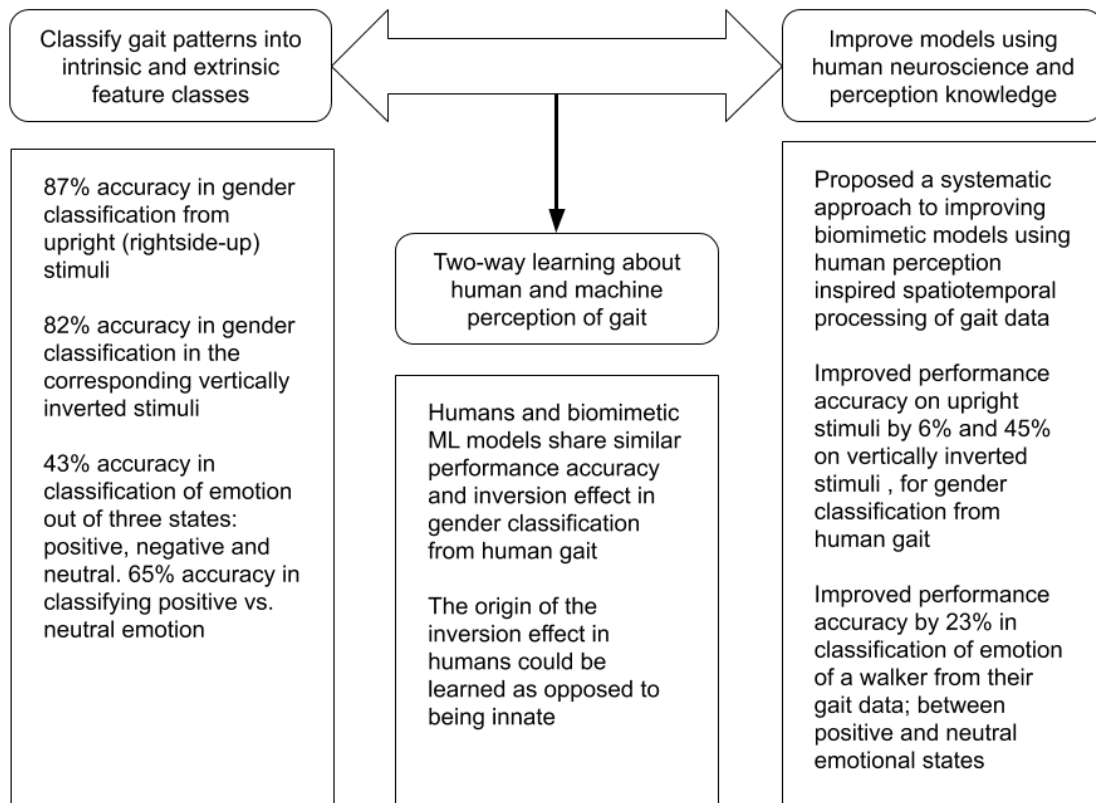
As proposed in the bidirectional learning paradigm, the BM could potentially be used as a tool for gaining insights into HP. As previously mentioned, the BM developed a human-like inversion effect, despite the absence of an anthropomorphic expectation of the skeleton in the gait information. The effect was an emergent behaviour in all the BM after training on veridical, right-side up skeletal gait stimuli, irrespective of their initial randomly generated configurations. The gender classification accuracy of the BM for vertically inverted stimuli decreased to below chance levels, to 37% ( $t_9 = -3.7$ ,  $p < 0.05$ ), similar to HP. This similarity could have hypothetically be brought about through by a significant disparity in the availability of veridical right-side-up walking datasets and upside-down walkers. Perhaps that could lead to the performance tilting below chance performance, because of the dependence of perception on hips and shoulder movements. Men are known to move their shoulders more, while women move their hips more [8]. In the vertically inverted stimuli, one could argue that the hips and shoulders approximately exchange positions anatomically, leading to a classification of gender to invert compared to what was inferred for the same subject in the right-side-up dataset, i.e. a male is reported as female, and vice versa. Studies have shown the indifference of new-borns to global configural characteristics of gait [169,176]. While the origin of inversion effect cannot be claimed with certainty from this research, the emergence of a similar behaviour in the BM and the similarity in gender classification outcomes suggests the possibility of the effect being a learned phenomenon rather than an innate feature of HP itself.

While the results of the experiments provide evidence for the applicability of the proposed bidirectional paradigm of improving BM, the impact of the proposed paradigm is probably only in its very beginnings. As computational resources become more affordable and the size of datasets increases, the model could be evolved readily to become significantly more sophisticated and better performing. Currently the scale and range of experimental results are limited by the computational budget, while the motivation for development is focused on the pragmatic application of the model as an automated gait classifier. A deviation of motivation towards biomimicry would encourage the exploration of other BM, including those that claim to mimic human perception more closely, making for interesting future avenues of exploration.

The objectives of the thesis have been achieved within the scope defined in the beginning: see Fig. 25 for a concise illustration of results against the objectives set in Chapter 1. Considerable effort was made to ensure unbiased evaluation across all the models including HP, BM and NBM. However, the nature of the models introduces bias, given their operations. For example, the inability of the NBM in processing temporal information forces the need for aggregating the time series into fixed length vectors, leading to information loss. These results suggested a markedly different approach to learning between HP, BM and NBM. One might argue that the static representation of gait as first order statistics may not be the best representation of gait for operation by the NBM. However, representation of human gait as a set of widely accepted clinical gait metrics was not able to provide the requisite information either.

Representation of a multi-dimensional time series data as a static numerical value, ensuring minimum loss of information is an ongoing researched topic. Future work shall also explore a multitude of static representations, such as the more recent. Recurrent Variational AutoEncoders (RVAE), which aims to develop a human interpretable static representation of time series data [177] that might also be used for human motion synthesis. Another variation of static representation could be based on PCA approach demonstrated by Troje et al. [127]. The use of Rough Path Signatures to generate a unique static representation of a time-series signal of fixed dimensionality is an additional promising avenue for exploration [125]. The architecture of the BM is only a functional emulation of our understanding of memory represented through perceptron-based neurons, deviating from the spiking biological neurons [31]. The various factors affecting the inference capabilities of the HP is not understood completely, introducing an unavoidable unfairness in the comparison. However, the conformation of BM to HP, shared aspects of perception and conduciveness to improvements based on perceptual findings, establishes the viability of the bidirectional paradigm of learning at a basic foundational level. While the artificial neural networks (ANNs) like the biomimetic model, demonstrate promise in analysis of gait and potentially other physiological phenomenon, they suffer from lack of interpretability. Being black-box models, understanding the criterion for a certain classification by the ANNs is a challenging task. In the case of humans, we get around this problem through verbal communication of the reasons for a particular classification. Interpretability of the decision by the ANNs is a current ongoing research topic, and would further strengthen the applicability of ANNs as an accurate and robust gait classifier.





*Figure 25: Concise representation of the main findings with reference to the stated research objectives (see Chapter 1). The horizontal arrow points to the improvements to machine perception of gait by leveraging human perception knowledge. The vertical arrow refers to the deeper understanding of human perception using insights gained from machine perception*

## Chapter 8 Conclusions

Humans are adept at extracting relevant features from gait for classifying the observed walker into an intrinsic or extrinsic feature class. Intrinsic features include features whose expression in gait remain comparatively unchanged in a person's lifetime, such as gender, while extrinsic features that can change frequently, possibly in a matter of minutes, such as emotional affect or clinical conditions. Some features can be extracted perceptually by humans from gait without any training, such as gender and emotional affect, while others may require extensive training and analysis, such as for clinical diagnosis and athletic performance evaluation. Learning to extract the latter accurately may require years and can be demanding on time and resources and prove difficult to transfer to another person. Automation of classification of the features would not only alleviate such expertise dependent problems in analysis of gait, but may also extend its viability of application in places that currently do not have access to such expertise. If successful, the automated gait analysis could be used in a myriad of fields from crowd analysis, clinical diagnosis, sports medicine to sentiment analysis.

Computational models that can be trained using machine learning (ML) techniques provide a viable approach to automating gait analysis. Fortunately, one can develop a ML model using various open source tools, making it an opportune time for exploration of such models. Towards this objective, in this thesis, psychophysical as well as computational experiments, spanning various machine learning models are executed with the aim of establishing a bidirectional learning approach between the two board fields of study; to improve our understanding of both human and machine perception concurrently. This approach leverages the strengths of either field to the study of the other, and inspires promising avenues for further research for both.

So called biomimetic or bio-inspired approaches to artificial intelligence (AI) continue to evolve. As knowledge of human perception and neurophysiology increases, the potential for powerful new models and applications will also increase. This thesis has demonstrated that development of biomimetic ML models and human psychology can be advanced not just independently but also concurrently, providing a synergistic leverage to both fields of research. An amalgamation of the robust versatility of humans with the finesse and attention to detail in computational models is shown to be a useful approach. As both fields continue to progress, perhaps an augmentation of biomimetic models into our biological neural networks can be imagined that is deeper and much more profound than brain-machine interface currently allow, aimed at overcoming disabilities while enhancing our natural and very human perceptual qualities.

## List of Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
BiLSTM	Bidirectional Long Short Term Memory
BM	Biomimetic Model
BP	Backpropagation
BPTT	Backpropagation Through Time
CART	Classification And Regression Tree
CCR	Correct Classification Rate
CNN	Convolutional Neural Network
CRPS	Complex Regional Pain Syndrome
DL	Deep Learning
DT	Decision Tree
EMG	Electromyogram
FAU	Florida Atlantic University
FFNN	Feed-Forward Neural Network
GA	Genetic Algorithms
GEI	Gait Energy Image
GEV	Gait Energy Volume
GRU	Gated Recurrent Unit
HP	Human Perception
ID3	Iterative Dichotomiser 3
ILP	Inductive Logic Programming
IMU	Inertial Measurement Unit
LSTM	Long Short Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron
NBM	Non-Biomimetic Model
PLD	Point Light Display
RBF	Radial Basis Function
RDF	Random Decision Forest
RGB	Red Green Blue
RGB-D	Red Green Blue Depth
RNN	Recurrent Neural Network
RVAE	Recurrent Variational Auto-Encoder
SVM	Support Vector Machine

## References

1. Davis JW, Gao H. Gender Recognition from Walking Movements using Adaptive Three-Mode PCA. 2004 Conference on Computer Vision and Pattern Recognition Workshop. doi:10.1109/cvpr.2004.354
2. Borràs R, Lapedriza À, Igual L. Depth Information in Human Gait Analysis: An Experimental Study on Gender Recognition. *Lecture Notes in Computer Science*. 2012. pp. 98–105. doi:10.1007/978-3-642-31298-4\_12
3. Yu S, Tan T, Huang K, Jia K, Wu X. A study on gait-based gender classification. *IEEE Trans Image Process*. 2009;18: 1905–1910.
4. Cao L, Dikmen M, Fu Y, Huang TS. Gender recognition from body. *Proceeding of the 16th ACM international conference on Multimedia - MM '08*. 2008. doi:10.1145/1459359.1459470
5. Huang G, Wang Y. Gender Classification Based on Fusion of Multi-view Gait Sequences. *Computer Vision – ACCV 2007*. pp. 462–471. doi:10.1007/978-3-540-76386-4\_43
6. Saunders DR, Williamson DK, Troje NF. Gaze patterns during perception of direction and gender from biological motion. *J Vis*. 2010;10: 9.
7. Mather, George, Murdoch L. Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. 1994. pp. 273–279. doi:10.1098/rspb.1994.0173
8. Barclay CD, Cutting JE, Kozlowski LT. Temporal and spatial factors in gait perception that influence gender recognition. *Percept Psychophys*. 1978;23: 145–152.
9. Pollick FE, Lestou V, Ryu J, Cho S-B. Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Res*. 2002;42: 2345–2355.
10. Pollick FE, Kay JW, Heim K, Stringer R. Gender recognition from point-light walkers. *J Exp Psychol Hum Percept Perform*. 2005;31: 1247–1265.
11. Yoo J-H, Hwang D, Nixon MS. Gender Classification in Human Gait Using Support Vector Machine. *Advanced Concepts for Intelligent Vision Systems*. 2005. pp. 138–145. doi:10.1007/11558484\_18
12. Schneider S, Christensen A, Häußinger FB, Fallgatter AJ, Giese MA, Ehlis A-C. Show me how you walk and I tell you how you feel - a functional near-infrared spectroscopy study on emotion perception based on human gait. *Neuroimage*. 2014;85 Pt 1: 380–390.
13. Venture G, Kadone H, Zhang T, Grèzes J, Berthoz A, Hicheur H. Recognizing Emotions Conveyed by Human Gait. *International Journal of Social Robotics*. 2014. pp. 621–632. doi:10.1007/s12369-014-0243-1
14. Roether CL, Omlor L, Christensen A, Giese MA. Critical features for the perception of emotion from gait. *J Vis*. 2009;9: 15.1–32.
15. Montepare JM, Goldstein SB, Clausen A. The identification of emotions from gait information. *Journal of Nonverbal Behavior*. 1987. pp. 33–42. doi:10.1007/bf00999605
16. Atkinson A, Dittrich WH, Gemmell AJ. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*. 2004. Available: journals.sagepub.com

17. Janssen D, Schöllhorn WI, Lubienetzki J, Fölling K, Kokenge H, Davids K. Recognition of Emotions in Gait Patterns by Means of Artificial Neural Nets. *J Nonverbal Behav.* 2008;32: 79–92.
18. Cutting JE, Kozlowski LT. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society.* 1977. pp. 353–356. doi:10.3758/bf03337021
19. Beardsworth T, Buckner T. The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society.* 1981. pp. 19–22. doi:10.3758/bf03333558
20. Hill H, Johnston A. Categorizing sex and identity from the biological motion of faces. *Curr Biol.* 2001;11: 880–885.
21. Cesari M, Kritchevsky SB, Penninx BWHJ, Nicklas BJ, Simonsick EM, Newman AB, et al. Prognostic value of usual gait speed in well-functioning older people--results from the Health, Aging and Body Composition Study. *J Am Geriatr Soc.* 2005;53: 1675–1680.
22. Morris ME, Ianssek R, Matyas TA, Summers JJ. The pathogenesis of gait hypokinesia in Parkinson's disease. *Brain.* 1994;117: 1169–1181.
23. Statistical tools for clinical gait analysis. *Gait Posture.* 2004;20: 204–212.
24. McGibbon CA, Krebs DE. Compensatory gait mechanics in patients with unilateral knee arthritis. *J Rheumatol.* 2002;29: 2410–2419.
25. Sarangi V, Algahtani E, Kazakov D, Pelah A, "Explainable AI for Clinical Gait Analysis: Developing Diagnostics with Inductive Logic Programming." 29<sup>th</sup> ILP Conference, Bulgaria, Sept. 2019
26. Cortes C, Vapnik V. Support-vector networks. *Machine Learning.* 1995. pp. 273–297. doi:10.1007/bf00994018
27. Mikolov T, Zweig G. Context dependent recurrent neural network language model. 2012 IEEE Spoken Language Technology Workshop (SLT). 2012. doi:10.1109/slt.2012.6424228
28. An improved ID3 decision tree algorithm - IEEE Conference Publication. [cited 14 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/5228509/>
29. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM.* 2017. pp. 84–90. doi:10.1145/3065386
30. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. *Computer Vision – ECCV 2014.* 2014. pp. 818–833. doi:10.1007/978-3-319-10590-1\_53
31. Izhikevich EM. Simple model of spiking neurons. *IEEE Transactions on Neural Networks.* 2003. pp. 1569–1572. doi:10.1109/tnn.2003.820440
32. Barclay CD, Cutting JE, Kozlowski LT. Temporal and spatial factors in gait perception that influence gender recognition. *Percept Psychophys.* 1978;23: 145–152.
33. Roether CL, Omlor L, Christensen A, Giese MA. Critical features for the perception of emotion from gait. *J Vis.* 2009;9: 15.1–32.
34. Montepare JM, Goldstein SB, Clausen A. The identification of emotions from gait information. *Journal of Nonverbal Behavior.* 1987. pp. 33–42. doi:10.1007/bf00999605
35. McGlothlin B, Jiacoletti D, Yandell L. The Inversion Effect: Biological Motion and Gender

- Recognition. *Psi Chi Journal of Psychological Research*. 2012. pp. 68–72. doi:10.24839/2164-8204.jn17.2.68
36. Troje NF, Westhoff C. The Inversion Effect in Biological Motion Perception: Evidence for a “Life Detector”? *Current Biology*. 2006. pp. 821–824. doi:10.1016/j.cub.2006.03.022
  37. Atkinson AP, Tunstall ML, Dittrich WH. Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures. *Cognition*. 2007;104: 59–72.
  38. Zegers J, Van hamme H. CNN-LSTM Models for Multi-Speaker Source Separation Using Bayesian Hyper Parameter Optimization. *Interspeech 2019*. 2019. doi:10.21437/interspeech.2019-2423
  39. Abu-Faraj ZO, Harris GF, Smith PA. Human gait and Clinical Movement Analysis. In: Webster JG, editor. *Wiley Encyclopedia of Electrical and Electronics Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1999. pp. 1–34.
  40. Clinical gait analysis: A review. *Hum Mov Sci*. 1996;15: 369–387.
  41. Mündermann L, Corazza S, Andriacchi TP. The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *J Neuroeng Rehabil*. 2006;3: 1–11.
  42. Pfister A, West AM, Bronner S, Noah JA. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *J Med Eng Technol*. 2014;38: 274–280.
  43. Ceseracciu E, Sawacha Z, Cobelli C. Comparison of Markerless and Marker-Based Motion Capture Technologies through Simultaneous Data Collection during Gait: Proof of Concept. *PLoS ONE*. 2014. p. e87640. doi:10.1371/journal.pone.0087640
  44. Carnegie Mellon University, School of computer science. Marker placement guide. Available: <http://mocap.cs.cmu.edu/markerPlacementGuide.pdf>
  45. Nagymáté G, Kiss RM. Application of OptiTrack motion capture systems in human movement analysis A systematic literature review. *Recent Innovations in Mechatronics*. 2018. doi:10.17667/riim.2018.1/13
  46. Clarke S, Institute for Marine Dynamics (Canada), Memorial University of Newfoundland. Faculty of Engineering and Applied Science. *QUALISYS Optical Tracking System*. 1992.
  47. A new paradigm of human gait analysis with Kinect - IEEE Conference Publication. [cited 14 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/7346722/>
  48. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Comput Vis Image Underst*. 2015;139: 1–20.
  49. Performance measurements for the Microsoft Kinect skeleton - IEEE Conference Publication. [cited 14 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/6180911/>
  50. Documentation about Microsoft Kinect v2s Joint type enumeration. Available: <https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx>
  51. Woolford K. Defining accuracy in the use of Kinect v2 for exercise monitoring. *Proceedings of the 2nd International Workshop on Movement and Computing - MOCO '15*. 2015. doi:10.1145/2790994.2791002
  52. Wong CK, Mentis HM, Kuber R. The bit doesn't fit: Evaluation of a commercial activity-tracker at slower walking speeds. *Gait Posture*. 2018;59: 177–181.

53. Lu CX, Du B, Wen H, Wang S, Markham A, Martinovic I, et al. Snoopy: Sniffing Your Smartwatch Passwords via Deep Sequence Learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2018;1: 29.
54. Loeb GE, Loeb G, Gans C. *Electromyography for Experimentalists*. University of Chicago Press; 1986.
55. Dubo HI, Peat M, Winter DA, Quanbury AO, Hobson DA, Steinke T, et al. Electromyographic temporal analysis of gait: normal human locomotion. *Arch Phys Med Rehabil*. 1976;57: 415–420.
56. York G, Chakrabarty S. A survey on foot drop and functional electrical stimulation. *International Journal of Intelligent Robotics and Applications*. 2019. pp. 4–10. doi:10.1007/s41315-019-00088-1
57. Shiavi R, Frigo C, Pedotti A. Electromyographic signals during gait: criteria for envelope filtering and number of strides. *Med Biol Eng Comput*. 1998;36: 171–178.
58. Sutherland DH. The evolution of clinical gait analysis part I: kinesiological EMG. *Gait Posture*. 2001;14: 61–70.
59. Meng M, Luo Z, She Q, Ma Y. Automatic recognition of gait mode from EMG signals of lower limb. 2010 The 2nd International Conference on Industrial Mechatronics and Automation. 2010. doi:10.1109/icindma.2010.5538164
60. Ounpuu S, Gage JR, Davis RB. Three-dimensional lower extremity joint kinetics in normal pediatric gait. *J Pediatr Orthop*. 1991;11: 341–349.
61. Nymark JR, Balmer SJ, Melis EH, Lemaire ED, Millar S. Electromyographic and kinematic nondisabled gait differences at extremely slow overground and treadmill walking speeds. *J Rehabil Res Dev*. 2005;42: 523–534.
62. Wentink EC, Schut VGH, Prinsen EC, Rietman JS, Veltink PH. Detection of the onset of gait initiation using kinematic sensors and EMG in transfemoral amputees. *Gait Posture*. 2014;39: 391–396.
63. Russell C, Roche AD, Chakrabarty S. Peripheral nerve bionic interface: a review of electrodes. *International Journal of Intelligent Robotics and Applications*. 2019. pp. 11–18. doi:10.1007/s41315-019-00086-3
64. Jin D, Yang J, Zhang R, Wang R, Zhang J. Terrain identification for prosthetic knees based on electromyographic signal features. *Tsinghua Science and Technology*. 2006. pp. 74–79. doi:10.1016/s1007-0214(06)70157-2
65. Aeyels B, Van Petegem W, Sloten JV, Van Der Perre G, Peeraer L. An EMG-based finite state approach for a microcomputer-controlled above-knee prosthesis. *Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society*. doi:10.1109/iembs.1995.579704
66. Au SK, Bonato P, Herr H. An EMG-Position Controlled System for an Active Ankle-Foot Prosthesis: An Initial Experimental Study. 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005. doi:10.1109/icorr.2005.1501123
67. Johansson G. Visual perception of biological motion and a model for its analysis. *Percept Psychophys*. 1973;14: 201–211.
68. Dittrich WH, Troscianko T, Lea SE, Morgan D. Perception of emotion from dynamic point-light displays represented in dance. *Perception*. 1996;25: 727–738.
69. Chang DHF, Troje NF. Characterizing global and local mechanisms in biological motion perception.

J Vis. 2009;9: 8.1–10.

70. Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. *J Microbiol Methods*. 2000;43: 3–31.
71. Smith J. *Machine Learning With Matlab: Supervised Learning and Regression*. Createspace Independent Publishing Platform; 2017.
72. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995. pp. 273–297. doi:10.1007/bf00994018
73. Guo G, Li SZ, Chan K. Face recognition by support vector machines. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. doi:10.1109/afgr.2000.840634
74. Smith, Smith, Gales. Using SVMs and discriminative models for speech recognition. *IEEE International Conference on Acoustics Speech and Signal Processing*. 2002. doi:10.1109/icassp.2002.1005680
75. Joachims T. Text categorization with Support Vector Machines: Learning with many relevant features. *Machine Learning: ECML-98*. 1998. pp. 137–142. doi:10.1007/bfb0026683
76. Shilton A. Design and training of support vector machines. PhD, The University of Melbourne. 2006.
77. Suykens JAK, Vandewalle J. Least Squares Support Vector Machine Classifiers. *Neural Process Letters*. 1999;9: 293–300.
78. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, et al. Real-time human pose recognition in parts from single depth images. *CVPR 2011*. 2011. doi:10.1109/cvpr.2011.5995316
79. Raileanu LE, Stoffel K. Theoretical Comparison between the Gini Index and Information Gain Criteria. *Annals of Mathematics and Artificial Intelligence*. 2004. pp. 77–93. doi:10.1023/b:amai.0000018580.96245.c6
80. Deep learning in neural networks: An overview. *Neural Netw*. 2015;61: 85–117.
81. Slotnick SD. Visual Memory and Visual Perception Recruit Common Neural Substrates. *Behavioral and Cognitive Neuroscience Reviews*. 2004. pp. 207–221. doi:10.1177/1534582304274070
82. Tang D, Qin B, Liu T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2015. doi:10.18653/v1/d15-1167
83. S. Angra and S. Ahuja, "Machine learning and its applications: A review," 2017 *International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, 2017, pp. 57-60, doi: 10.1109/ICBDACI.2017.8070809.
84. Graves A. Long Short-Term Memory. *Studies in Computational Intelligence*. 2012. pp. 37–45. doi:10.1007/978-3-642-24797-2\_4
85. Kim M, Cao B, Mau T, Wang J. Speaker-Independent Silent Speech Recognition from Flesh-Point Articulatory Movements Using an LSTM Neural Network. *IEEE/ACM Trans Audio Speech Lang Process*. 2017;25: 2323–2336.
86. Nelson DMQ, Pereira ACM, de Oliveira RA. Stock market's price movement prediction with LSTM neural networks. *2017 International Joint Conference on Neural Networks (IJCNN)*. 2017.



doi:10.1109/ijcnn.2017.7966019

87. Wong KH, Tang CP, Chui KL, Yu YK, Zeng Z. Music genre classification using a hierarchical long short term memory (LSTM) model. Third International Workshop on Pattern Recognition. 2018. doi:10.1117/12.2501763
88. Weiss RJ, Chorowski J, Jaitly N, Wu Y, Chen Z. Sequence-to-Sequence Models Can Directly Translate Foreign Speech. Interspeech 2017. 2017. doi:10.21437/interspeech.2017-503
89. Giese MA, Poggio T. Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*. 2003. pp. 179–192. doi:10.1038/nrn1057
90. Lange J. A Model of Biological Motion Perception from Configural Form Cues. *Journal of Neuroscience*. 2006. pp. 2894–2906. doi:10.1523/jneurosci.4915-05.2006
91. Comparing support vector machines with Gaussian kernels to radial basis function classifiers - IEEE Journals & Magazine. [cited 14 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/650102/>
92. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. In: Google Books [Internet]. [cited 14 Jan 2020]. Available: [https://books.google.com/books/about/An\\_Introduction\\_to\\_Support\\_Vector\\_Machin.html?id=\\_PXJn\\_cxv0AC](https://books.google.com/books/about/An_Introduction_to_Support_Vector_Machin.html?id=_PXJn_cxv0AC)
93. Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 1958. pp. 386–408. doi:10.1037/h0042519
94. Poggio T, Mhaskar H, Rosasco L, Miranda B, Liao Q. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. *Int J Autom Comput*. 2017;14: 503–519.
95. Delalleau O, Bengio Y. Shallow vs. Deep Sum-Product Networks. *Advances in Neural Information Processing Systems*. 2011. pp. 666–674.
96. Xin D, Hua-hua C, Wei-kang G. Neural network and genetic algorithm based global path planning in a static environment. *Journal of Zhejiang University-SCIENCE A*. 2005. pp. 549–554. doi:10.1631/jzus.2005.a0549
97. Gupta JND, Sexton RS. Comparing backpropagation with a genetic algorithm for neural network training. *Omega*. 1999. pp. 679–684. doi:10.1016/s0305-0483(99)00027-4
98. Sexton RS, Dorsey RE. Reliable classification using neural networks: a genetic algorithm and backpropagation comparison. *Decision Support Systems*. 2000. pp. 11–22. doi:10.1016/s0167-9236(00)00086-5
99. Rajasekaran S, Vijayalakshmi Pai GA. NEURAL NETWORKS, FUZZY LOGIC AND GENETIC ALGORITHM: SYNTHESIS AND APPLICATIONS (WITH CD). PHI Learning Pvt. Ltd.; 2003.
100. Örkücü HH, Hasan Örkücü H, Bal H. Comparing performances of backpropagation and genetic algorithms in the data classification. *Expert Systems with Applications*. 2011. pp. 3703–3709. doi:10.1016/j.eswa.2010.09.028
101. Gao C, Smith SL, Lones M, Jamieson S, Alty J, Cosgrove J, et al. Objective assessment of bradykinesia in Parkinson's disease using evolutionary algorithms: clinical validation. *Translational Neurodegeneration*. 2018. doi:10.1186/s40035-018-0124-x
102. Lones MA, Alty JE, Cosgrove J, Duggan-Carter P, Jamieson S, Naylor RF, et al. A New Evolutionary Algorithm-Based Home Monitoring Device for Parkinson's Dyskinesia. *Journal of*

Medical Systems. 2017. doi:10.1007/s10916-017-0811-7

103. Smith SL, Lones MA, Bedder M, Alty JE, Cosgrove J, Maguire RJ, et al. Computational approaches for understanding the diagnosis and treatment of Parkinson's disease. *IET Syst Biol.* 12.2015;9: 226–233.
104. Lones MA, Alty JE, Duggan-Carter P, Turner AJ, Stuart Jamieson DR, Smith SL. Classification and characterisation of movement patterns during levodopa therapy for parkinson's disease. *Proceedings of the 2014 conference companion on Genetic and evolutionary computation companion - GECCO Comp '14.* 2014. doi:10.1145/2598394.2609852
105. Kruschke JK. Human Category Learning: Implications for Backpropagation Models. *Connection Science.* 1993. pp. 3–36. doi:10.1080/09540099308915683
106. Cohen JD, Dunbar K, McClelland JL. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychol Rev.* 1990;97: 332–361.
107. Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. *Psychological Review.* 1989. pp. 523–568. doi:10.1037/0033-295x.96.4.523
108. Gluck MA. Stimulus Generalization and Representation in Adaptive Network Models of Category Learning. *Psychological Science.* 1991. pp. 50–55. doi:10.1111/j.1467-9280.1991.tb00096.x
109. McCloskey M, Cohen NJ. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. *Psychology of Learning and Motivation.* 1989. pp. 109–165. doi:10.1016/s0079-7421(08)60536-8
110. Makioka S. A Connectionist Model of Phonological Working Memory. *New Developments in Psychometrics.* 2003. pp. 577–584. doi:10.1007/978-4-431-66996-8\_66
111. Werbos PJ. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE.* 1990. pp. 1550–1560. doi:10.1109/5.58337
112. VISUAL MOTION PERCEPTION on JSTOR. [cited 14 Jan 2020]. Available: <https://www.jstor.org/stable/24949822>
113. Johansson G. Spatio-temporal differentiation and integration in visual motion perception. *Psychol Res.* 1976;38: 379–393.
114. Sumi S. Upside-down presentation of the Johansson moving light-spot pattern. *Perception.* 1984;13: 283–286.
115. Kozlowski LT, Cutting JE. Recognizing the sex of a walker from a dynamic point-light display. *Percept Psychophys.* 1977;21: 575–580.
116. Pollick FE, Paterson HM, Bruderlin A, Sanford AJ. Perceiving affect from arm movement. *Cognition.* 2001. pp. B51–B61. doi:10.1016/s0010-0277(01)00147-0
117. Gait analysis for recognition and classification - IEEE Conference Publication. [cited 14 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/1004148/>
118. Han J, Bhanu B. Individual recognition using gait energy image. *IEEE Trans Pattern Anal Mach Intell.* 2006;28: 316–322.
119. Sivapalan S, Chen D, Denman S, Sridharan S, Fookes C. Gait energy volumes and frontal gait recognition using depth images. *2011 International Joint Conference on Biometrics (IJCB).* 2011. doi:10.1109/ijcb.2011.6117504

120. Lee L, Grimson WEL. Gait analysis for recognition and classification. Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition. doi:10.1109/afgr.2002.1004148
121. Li X, Wang L, Sung E. AdaBoost with SVM-based component classifiers. Engineering Applications of Artificial Intelligence. 2008. pp. 785–795. doi:10.1016/j.engappai.2007.07.001
122. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002; 2: 18–22.
123. The Principles of Psychology. By William James, Professor of Psychology in Harvard University. Two Vols. American Science Series. New York: Henry Holt and Co. 1890. London: Macmillan and Co. 1890. Journal of Mental Science. 1891. pp. 428–434. doi:10.1192/bjp.37.158.428
124. Wood MM. Book Reviews : KINESICS AND CONTEXT: ESSAYS ON BODY MOTION COMMUNICATION. Ray L. Birdwhistell. Philadelphia, University of Pennsylvania Press, 1970. The ABCA Journal of Business Communication. 1972. pp. 68–69. doi:10.1177/002194367200900311
125. Henley N. Body politics: power, sex, and nonverbal communication. Prentice Hall; 1977.
126. Gelder B de, de Gelder B, Vroomen J. The perception of emotions by ear and by eye. Cognition & Emotion. 2000. pp. 289–311. doi:10.1080/026999300378824
127. Troje NF. Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. J Vis. 2002;2: 371–387.
128. Atkinson A, Dittrich WH, Gemmell AJ. Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception. 2004. Available: journals.sagepub.com
129. Wallbott HG. Bodily expression of emotion. European Journal of Social Psychology. 1998. pp. 879–896. doi:10.1002/(sici)1099-0992(199811)28:6<879::aid-ejsp901>3.0.co;2-w
130. Wallbott HG, Scherer KR. Cues and channels in emotion recognition. Journal of Personality and Social Psychology. 1986. pp. 690–699. doi:10.1037/0022-3514.51.4.690
131. Sawada M, Suda K, Ishii M. Expression of Emotions in Dance: Relation between Arm Movement Characteristics and Emotion. Perceptual and Motor Skills. 2003. pp. 697–708. doi:10.2466/pms.2003.97.3.697
132. Yang L, Zhang L, Dong H, Alelaiwi A, El Saddik A. Evaluating and Improving the Depth Accuracy of Kinect for Windows v2. IEEE Sensors Journal. 2015. pp. 4275–4285. doi:10.1109/jsen.2015.2416651
133. Setup tips for your Kinect sensor and play space. Available: <https://beta.support.xbox.com/help/hardware-network/kinect/kinect-sensor-setup-tips>
134. Body Class. In: Microsoft Docs [Internet]. Available: [https://docs.microsoft.com/en-us/previous-versions/windows/kinect/dn772824\(v%3Dweb.10\)](https://docs.microsoft.com/en-us/previous-versions/windows/kinect/dn772824(v%3Dweb.10))
135. Clark RA, Pua Y-H, Fortin K, Ritchie C, Webster KE, Denehy L, et al. Validity of the Microsoft Kinect for assessment of postural control. Gait Posture. 2012;36: 372–377.
136. Clark RA, Bower KJ, Mentiplay BF, Paterson K, Pua Y-H. Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. J Biomech. 2013;46: 2722–2725.
137. van Diest M, Stegenga J, Wörtche HJ, Postema K, Verkerke GJ, Lamothe CJC. Suitability of Kinect for measuring whole body movement patterns during exergaming. J Biomech. 2014;47: 2925–2932.

138. Shi Y, Ma X, Ma Z, Wang J, Yao N, Gu Q, et al. Using a Kinect sensor to acquire biological motion: Toolbox and evaluation. *Behav Res Methods*. 2018;50: 518–529.
139. The Center for the Study of Emotion and Attention. [cited 14 Jan 2020]. Available: <https://csea.phhp.ufl.edu/Media.html#topmedia>
140. Siciliano A. *MATLAB: Data Analysis and Visualization*. World Scientific; 2008.
141. Anaconda | The World's Most Popular Data Science Platform. In: Anaconda | The World's Most Popular Data Science Platform [Internet]. Available: <https://www.anaconda.com/>
142. Mishra RK. Introduction to Python and NumPy. *PySpark Recipes*. 2018. pp. 45–83. doi:10.1007/978-1-4842-3141-8\_3
143. Blanco-Silva FJ. *Learning SciPy for Numerical and Scientific Computing*. Packt Publishing Ltd; 2013.
144. Trappenberg TP. Machine learning with sklearn. *Fundamentals of Machine Learning*. 2019. pp. 38–65. doi:10.1093/oso/9780198828044.003.0003
145. Ketkar N. Introduction to Tensorflow. *Deep Learning with Python*. 2017. pp. 159–194. doi:10.1007/978-1-4842-2766-4\_11
146. Cython: C-Extensions for Python. [cited 15 Jan 2020]. Available: <https://cython.org/>
147. LIBSVM -- A Library for Support Vector Machines. [cited 15 Jan 2020]. Available: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
148. Shangareeva GR, Mustafina SA. Parallelization of the conjugate gradient method using the technology NVidia Cuda. *Scientific Bulletin*. 2014. pp. 155–162. doi:10.17117/nv.2014.02.155
149. ONEWAY Analysis of Variance. *Quantitative Research in Communication*. pp. 51–70. doi:10.4135/9781452274881.n4
150. Saunders DR, Williamson DK, Troje NF. Gaze patterns during perception of direction and gender from biological motion. *J Vis*. 2010;10: 9.
151. Draganski B, Gaser C, Busch V, Schuierer G, Bogdahn U, May A. Changes in grey matter induced by training. *Nature*. 2004. pp. 311–312. doi:10.1038/427311a
152. Graves A, Fernández S, Gomez F, Schmidhuber J. Connectionist temporal classification. *Proceedings of the 23rd international conference on Machine learning - ICML '06*. 2006. doi:10.1145/1143844.1143891
153. Efficient Learning of Sparse Representations with an Energy-Based Model. *Advances in Neural Information Processing Systems 19*. 2007. doi:10.7551/mitpress/7503.003.0147
154. Hinton GE, Osindero S, Teh Y-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*. 2006. pp. 1527–1554. doi:10.1162/neco.2006.18.7.1527
155. Osindero S, Welling M, Hinton GE. Topographic Product Models Applied to Natural Scene Statistics. *Neural Computation*. 2006. pp. 381–414. doi:10.1162/089976606775093936
156. Ito M. Representation of Angles Embedded within Contour Stimuli in Area V2 of Macaque Monkeys. *Journal of Neuroscience*. 2004. pp. 3313–3324. doi:10.1523/jneurosci.4364-03.2004
157. Lee H, Ekanadham C, Ng AY. Sparse deep belief net model for visual area V2. 2008. pp. 873–880.

158. Bengio Y, Frasconi P, Simard P. The problem of learning long-term dependencies in recurrent networks. *IEEE International Conference on Neural Networks*. doi:10.1109/icnn.1993.298725
159. Chen T, Xu R, He Y, Wang X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications*. 2017. pp. 221–230. doi:10.1016/j.eswa.2016.10.065
160. Maas AL, Hannun A, Ng A. Rectifier nonlinearities improve neural network acoustic models. *robotics.stanford.edu*. 2013. Available: [http://robotics.stanford.edu/~amaas/papers/relu\\_hybrid\\_icml2013\\_final.pdf](http://robotics.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf)
161. Brébisson D, Alexandre, Vincent P. An exploration of softmax alternatives belonging to the spherical loss family. 2015. Available: <https://arxiv.org/abs/1511.05042>
162. Kingma DP, Ba J. Adam: A method for stochastic optimization. Available: <https://arxiv.org/abs/1412.6980>
163. McKone E, Yovel G. Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. *Psychonomic Bulletin & Review*. 2009. pp. 778–797. doi:10.3758/pbr.16.5.778
164. Pavlova M, Sokolov A. Orientation specificity in biological motion perception. *Percept Psychophys*. 2000;62: 889–899.
165. Yin RK. Looking at upside-down faces. *Journal of Experimental Psychology*. 1969. pp. 141–145. doi:10.1037/h0027474
166. Bertenthal BI, Pinto J. Global Processing of Biological Motions. *Psychological Science*. 1994. pp. 221–225. doi:10.1111/j.1467-9280.1994.tb00504.x
167. Saunders DR, Suchan J, Troje NF. Off on the wrong foot: local features in biological motion. *Perception*. 2009;38: 522–532.
168. Chang DHF, Troje NF. Acceleration carries the local inversion effect in biological motion perception. *J Vis*. 2009;9: 19.1–17.
169. Simion F, Regolin L, Bulf H. A predisposition for biological motion in the newborn baby. *Proc Natl Acad Sci U S A*. 2008;105: 809–813.
170. Civile C, McLaren RP, McLaren IPL. The Face Inversion Effect—Parts and Wholes: Individual Features and Their Configuration. *Quarterly Journal of Experimental Psychology*. 2014. pp. 728–746. doi:10.1080/17470218.2013.828315
171. VISUAL MOTION PERCEPTION on JSTOR. [cited 14 Jan 2020]. Available: <https://www.jstor.org/stable/24949822>
172. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*. 2017. pp. 1413–1432. doi:10.1007/s11222-016-9696-4
173. Buck R, Miller RE, Caul WF. Sex, personality, and physiological variables in the communication of affect via facial expression. *J Pers Soc Psychol*. 1974;30: 587–596.
174. Lewis MA, Fagg AH, Solidum A. Genetic programming approach to the construction of a neural network for control of a walking robot. *Proceedings 1992 IEEE International Conference on Robotics and Automation*. doi:10.1109/robot.1992.220047

175. Martin JH, Chakrabarty S, Friel KM. Harnessing activity-dependent plasticity to repair the damaged corticospinal tract in an animal model of cerebral palsy. *Developmental Medicine & Child Neurology*. 2011. pp. 9–13. doi:10.1111/j.1469-8749.2011.04055.x
176. Chakrabarty S, Martin JH. Postnatal Development of the Motor Representation in Primary Motor Cortex. *Journal of Neurophysiology*. 2000. pp. 2582–2594. doi:10.1152/jn.2000.84.5.2582
177. Pereira J, Silveira M. Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). 2018. doi:10.1109/icmla.2018.00207
178. Sarangi V, Pelah A, Hahn W, Barenholtz E. Biological and biomimetic perception: A comparative study of gender recognition from human gait. *Journal of Perceptual Imaging*. 1.
179. Sarangi V, Pelah A, Hahn W, Barenholtz E. Biological and biomimetic perception: A comparative study through gender recognition from human gait. *Human Vision and Electronic Imaging 2020*. Society for Imaging Science and Technology;
180. An Introduction to Statistical Learning | SpringerLink. [cited 14 Jan 2020]. Available: <https://link.springer.com/content/pdf/10.1007/978-1-4614-7138-7.pdf>
181. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. In: Google Books [Internet]. [cited 14 Jan 2020]. Available: [https://books.google.com/books/about/An\\_Introduction\\_to\\_Support\\_Vector\\_Machin.html?id=\\_PXJn\\_cxv0AC](https://books.google.com/books/about/An_Introduction_to_Support_Vector_Machin.html?id=_PXJn_cxv0AC)
182. Breiman L, Friedman JH, Olshen RA, Stone CJ. Regression Trees. *Classification And Regression Trees*. 2017. pp. 216–265. doi:10.1201/9781315139470-8
183. An improved ID3 decision tree algorithm - IEEE Conference Publication. [cited 14 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/5228509/>
184. Hssina B, Merbouha A, Ezzikouri H, Erritali M. A comparative study of decision tree ID3 and C4.5. *International Journal of Advanced Computer Science and Applications*. 2014. doi:10.14569/specialissue.2014.040203
185. Lakshminarayanan B. Decision trees and forests: a probabilistic perspective. PhD, University College London. 2016.
186. Criminisi A, Shotton J, Konukoglu E. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *CGV*. 2012;7: 81–227.
187. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature*. 1986. pp. 533–536. doi:10.1038/323533a0
188. Jain AK, Farrokhnia F. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*. 1991. pp. 1167–1186. doi:10.1016/0031-3203(91)90143-s
189. ImageNet Large Scale Visual Recognition Competition (ILSVRC). [cited 15 Jan 2020]. Available: <http://www.image-net.org/challenges/LSVRC/>
190. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998. pp. 2278–2324. doi:10.1109/5.726791
191. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.

doi:10.1109/cvpr.2015.7298594

192. Visual Geometry Group - University of Oxford. [cited 14 Jan 2020]. Available: [http://www.robots.ox.ac.uk/~vgg/research/very\\_deep/](http://www.robots.ox.ac.uk/~vgg/research/very_deep/)
193. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. doi:10.1109/cvpr.2016.90
194. Yuan X, Feng Z, Norton M, Li X. Generalized Batch Normalization: Towards Accelerating Deep Neural Networks. Proceedings of the AAAI Conference on Artificial Intelligence. 2019. pp. 1682–1689. doi:10.1609/aaai.v33i01.33011682
195. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2000. doi:10.1002/0471722146
196. Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation. 1997. pp. 67–82. doi:10.1109/4235.585893
197. Begg R, Kamruzzaman J. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. Journal of Biomechanics. 2005. pp. 401–408. doi:10.1016/j.jbiomech.2004.05.002
198. Assessment of gait patterns using neural networks. J Biomech. 1993;26: 645–651.
199. Cunado D, Nash JM, Nixon MS, Carter JN. Gait Extraction and Description by Evidence-Gathering. 1999. pp. 43–48.
200. Liu L-F, Jia W, Zhu Y-H. Survey of Gait Recognition. Emerging Intelligent Computing Technology and Applications With Aspects of Artificial Intelligence. Springer, Berlin, Heidelberg; 2009. pp. 652–659.
201. Model-based human gait recognition using leg and arm movements. Eng Appl Artif Intell. 2010;23: 1237–1246.
202. A Review of Vision-Based Gait Recognition Methods for Human Identification - IEEE Conference Publication. [cited 15 Jan 2020]. Available: <https://ieeexplore.ieee.org/abstract/document/5692583/>
203. Zheng H, Yang M, Wang H, McClean S. Machine Learning and Statistical Approaches to Support the Discrimination of Neuro-degenerative Diseases Based on Gait Analysis. Studies in Computational Intelligence. pp. 57–70. doi:10.1007/978-3-642-00179-6\_4
204. Maki BE. Gait Changes in Older Adults: Predictors of Falls or Indicators of Fear? Journal of the American Geriatrics Society. 1997. pp. 313–320. doi:10.1111/j.1532-5415.1997.tb00946.x
205. O'Malley MJ, Abel MF, Damiano DL, Vaughan CL. Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. IEEE Transactions on Rehabilitation Engineering. 1997. pp. 300–309. doi:10.1109/86.650282
206. Barton JG, Lees A. An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. Gait & Posture. 1997. pp. 28–33. doi:10.1016/s0966-6362(96)01070-3
207. Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. IEEE Trans Neural Netw. 1999;10: 1055–1064.
208. Ding CH, Dubchak I. Multi-class protein fold recognition using support vector machines and neural networks. Bioinformatics. 2001;17: 349–358.

209. Hinton GE. Reducing the Dimensionality of Data with Neural Networks. *Science*. 2006. pp. 504–507. doi:10.1126/science.1127647
210. Sarangi V, Pelah A, Hahn W, Barenholtz E., “Gender Perception from Gait: A comparison between biological, biomimetic and non-biomimetic learning paradigms” *Frontiers in Human Neuroscience - Motor Neuroscience*



# Appendix

## A Support Vector Machines

### A.I Hyperplane

In a  $p$ -dimensional space, a hyperplane is a flat affine subspace (a subspace need not pass through the origin) of dimension  $p-1$ . It is defined by equation

$$\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_px_p = 0 \quad (1)$$

for the parameter  $\beta_1 \dots \beta_p$ . In this case, the point  $X=(x_1, x_2, x_3, \dots, x_p)$  lies on the hyperplane. If  $X$  lies on the hyperplane then it satisfied equation 1. But if it doesn't then it the equation can take two forms, essentially signifying the point lying on either side of the hyperplane. The hyperplane divides the infinite hyperspace into two parts, each denoting a certain class.

$$\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_px_p < 0 \quad (2)$$

$$\beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_px_p > 0 \quad (3)$$

### A.II Classification using a Separating Hyperplane

Given an  $n \times p$  data matrix that consists of  $n$  observations in a  $p$  dimensional space

$$x_1 = (x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16} \dots x_{1p}) \dots x_n = (x_{n1}, x_{n2}, x_{n3}, x_{n4}, x_{n5}, x_{n6} \dots x_{np})$$

with each point  $x_i$  belonging to a class  $y_i \in \{-1, 1\}$ , where  $i = 1, 2, 3, 4, \dots, n$  i.e.  $x_1$  belongs to  $y_1$  and so on, where  $y_i$  is either +1 or -1 depending on which class it belongs to. The objective with SVMs is to develop a classifier that can classify a new given point  $X$  in  $p$ -dimensional space based on the pattern of classification of the points  $x_1, x_2 \dots x_n$ . The separating hyperplane is based on the concept developed by [180]. Given any point in hyper-dimensional space,  $x_i$

$$\beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \dots + \beta_px_{ip} < 0 \text{ if } y_i = -1 \quad (4)$$

$$\beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \dots + \beta_px_{ip} > 0 \text{ if } y_i = 1 \quad (5)$$

Combining equations 4 and 5, we get  $y_i(\beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \dots + \beta_px_{ip}) > 0 \quad (6)$

Thus, in practice SVMs classify the new point  $X$  based on which side of the hyperplane it lies in the hyperspace. When substituted for  $x$  in the equation

$$f(x) = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \dots + \beta_px_{ip} \quad (7)$$

a negative value classifies it into class -1 and a positive value places it in class 1 as shown in Fig. A- 1.

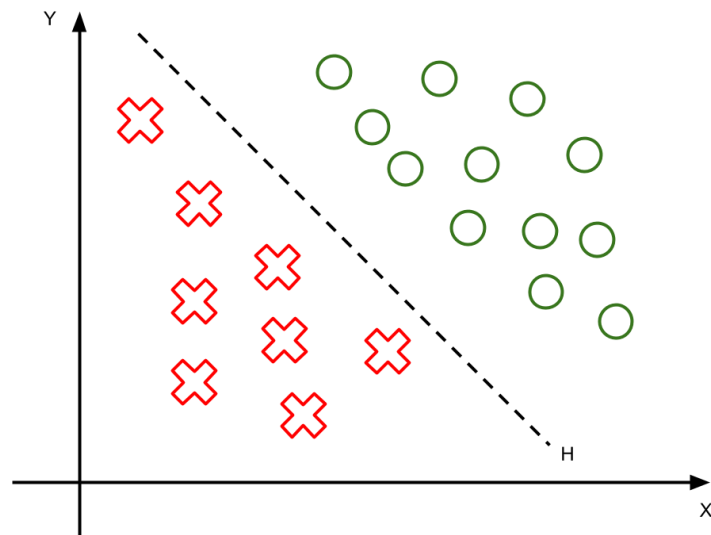


Figure A- 1 Separating hyperplane denoted by HP in two-dimensional space

### A.III Maximal Margin Classifier

Assuming a perfectly linearly separable dataset, there exists an infinite number of separating hyperplanes that could classify the data points. The objective of the SVM is to select the hyperplane that can separate the data points belonging to the different classes with the maximum margin of separation to ensure high confidence in classification. The maximal margin classifier is also known as the optimal separating hyperplane. The minimum distance of the training data points from the separating hyperplane is called the ‘margin’ [180]. The optimal hyperplane attempts to maximise the margin of separation,  $M$ , denoted by Fig. A- 2 and the equations,

$$\max M \text{ given } \beta_0, \beta_1, \beta_2, \dots, \beta_p \quad (8)$$

$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1, \text{ as the hyperplane vector is of unit length} \quad (9)$$

$$\text{and } y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip}) \geq M \quad \text{for all } i = 1, 2, \dots, n \quad (10)$$

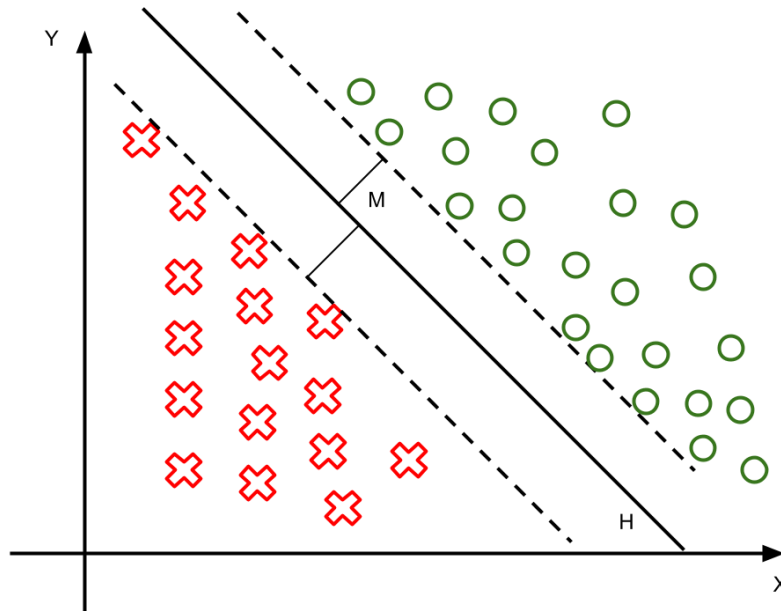


Figure A- 2 Maximal margin hyperplane, HP and the margin  $M$  in two-dimensional space

Equation 6 ensures a positive (or zero) value for the separation from the hyperplane. The left side of Equation 10 provides the distance of the point  $x_i$  from the separating hyperplane  $H$  as a positive value  $M$ . Thus maximising the value of  $M$  would provide the maximal margin hyperplane that results in the value.

#### A.IV Support Vector Classifier

The maximal margin classifier has the ability to perfectly separate two linearly separable sets of data points. However, the strict margins of separations is not robust to small perturbations in data points. Mislabelling of a few data points penalises the performance heavily and results in a non-optimal solution, as shown Fig. A- 3.

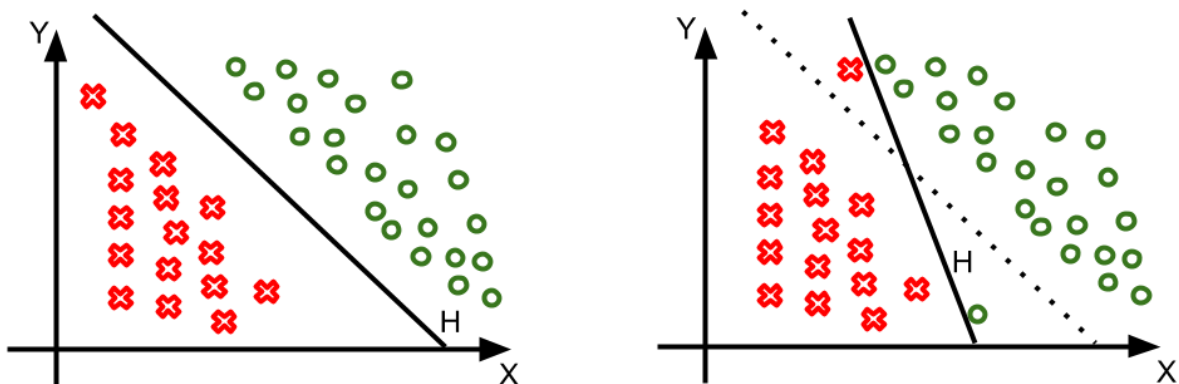


Figure A- 3 Hyperplane changes affected by a small number of mislabelled data points

This is generally termed as ‘overfitting’ in the training phase of machine learning algorithms, where the algorithm tries to fit all of the training data into the model, resulting in a less than ideal solution. Thus the need for introducing a margin of error allowed during training, also known as

soft margin classifier or support vector classifier. Unlike the maximal margin classifier, the soft margin classifier has some leeway in allowing some data points to fall on the wrong side of the dividing hyperplane with an error value  $\varepsilon$  as shown in Fig. A- 4.

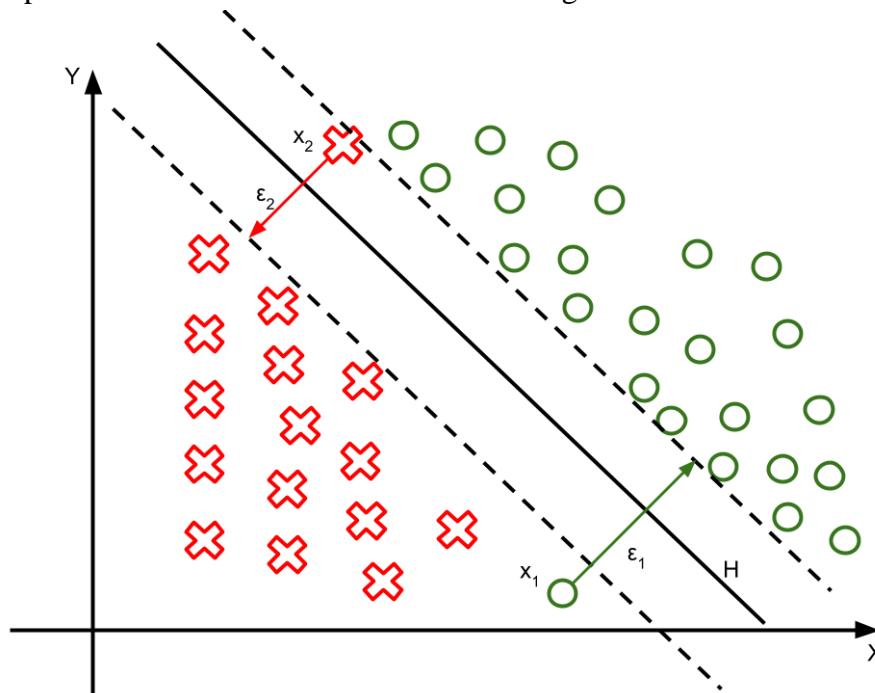


Figure A- 4 Soft Margin Classifier

As shown in the figure, the points  $x_1$  and  $x_2$  fall on the wrong side of the hyperplane and have a distance of  $\varepsilon_1$  and  $\varepsilon_2$  from their corresponding *support vectors*. Similar to the maximal margin classifier, the soft margin classifier is the solution to the optimisation problem

$$\max M \text{ given } \beta_0, \beta_1, \beta_2, \dots, \beta_p \quad (11)$$

$$\text{subject to } \sum_{j=1}^p \beta_j^2 = 1 \quad (12)$$

$$\text{and } y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip}) \geq M(1 - \varepsilon_i) \quad \text{for all } i = 1, 2, \dots, n \quad (13)$$

$$\varepsilon_i \geq 0 \text{ and } \sum_{i=1}^n \varepsilon_i \leq C \quad (14)$$

where  $C$  is the nonnegative tuning parameter. As shown in Fig. A- 4,  $M$  is the allowed ideal margin width which is intended to be as large as possible. The values  $\varepsilon_i$  for the data points that were erroneously classified are called '*slack variables*' that relax the margin requirements for the soft margin classifier. From the equations, it is clear that when  $0 \leq \varepsilon_i \leq 1$ , the data point lies on the wrong side of the margin or hyperplane (as shown by the point  $x_2$  with the slack variable  $\varepsilon_2$  in the Fig. A- 4). However when  $\varepsilon_i > 1$ , the data point lies on the wrong side of both hyperplanes (as shown by the point  $x_1$  with the slack variable  $\varepsilon_1$  in the Fig. A- 4)

## A.V Solving for the Support Vector Classifier

Combining equations 11 and 12

$$\max M \text{ given } \beta, \beta_0, \|\beta\|=1 \quad (13)$$

$$y_i(x_i^T \beta + \beta_0) \geq M(1 - \varepsilon_i), \text{ for all } i = 1, 2, \dots, n \quad (14)$$

$$\varepsilon_i \geq 0 \text{ and } \sum_{i=1}^n \varepsilon_i \leq C \quad (15)$$

where  $x_i \in R^p$  with unit vector  $\|\beta\| = 1$ . If we ignore the norm constraint on  $\beta$ , where  $M$  is defined as  $M = \frac{1}{\|\beta\|}$ , equations 13 to 15 can be represented as

$$\min \|\beta\| \quad (16)$$

$$\text{subject to } y_i(x_i^T \beta + \beta_0) \geq 1 - \varepsilon_i, \text{ for all } i = 1, 2, \dots, n \quad (17)$$

$$\varepsilon_i \geq 0 \text{ and } \sum_{i=1}^n \varepsilon_i \leq C \quad (18)$$

The problem 16-18 is a quadratic equation with linear inequality constraints. Hence the solution for the same can be obtained using Lagrangian multipliers [181]. The problem can be rephrased as

$$\min \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \varepsilon_i \quad (19)$$

$$\text{subject to } y_i(x_i^T \beta + \beta_0) \geq 1 - \varepsilon_i, \text{ for all } i = 1, 2, \dots, n \quad (20)$$

$$\varepsilon_i \geq 0 \quad (21)$$

Equation 19 is also known as the loss function that needs to be optimised. The first term represents the regularisation term while the second term represents the loss function penalty term. The lagrangian function is

$$L = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n \varepsilon_i - \sum_{i=1}^N \alpha_i [y_i(x_i^T \beta + \beta_0) - (1 - \varepsilon_i)] - \sum_{i=1}^N \mu_i \varepsilon_i \quad (22)$$

Taking the partial derivative to zero, we get

$$\beta = \sum_{i=1}^N \alpha_i y_i x_i, \text{ taking the partial derivative w.r.t. } \beta \quad (23)$$

$$\sum_{i=1}^N \alpha_i y_i = 0, \text{ taking the partial derivative w.r.t. } x_i \quad (24)$$

$$\alpha_i = C - \mu_i \text{ for } i = 1, 2, \dots, N, \text{ taking the partial derivative w.r.t } \varepsilon_i \quad (25)$$

where constraints  $\alpha_i, \mu_i, \varepsilon_i \geq 0$ . By substituting 23-25 into 22, we get

$$\max L = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (26)$$

$$\text{subject to } 0 \leq \alpha_i \leq C \quad (27)$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad (28)$$

In addition, the Karush-Kuhn-Tucker conditions are satisfied by the solution as following

$$\alpha_i [y_i (x_i^T \beta + \beta_0) - (1 - \varepsilon_i)] = 0 \quad (29)$$

$$\mu_i \varepsilon_i = 0 \quad (30)$$

$$[y_i (x_i^T \beta + \beta_0) - (1 - \varepsilon_i)] \geq 0 \quad (31)$$

For  $i = 1, 2, \dots, N$

The equations from 23 to 29 show characteristics of solutions to the primal and dual problems as described in the kernel functions [181]. Meanwhile Equation. 23 can provide us with the solution for  $\beta$  which is in the form of  $\beta = \sum_{i=1}^N \alpha_i y_i x_i$  and the observations with constraints in Equation. 31 are notably known as ‘*support vectors*’ which refer to the data points that lie on the soft margin hyperplane. Some of the support vectors meet the constraints referred to in Equation. 29 and lie on the edge of the margin and can be utilised in solving for  $\beta_0$

## A.VI Kernel functions: Classification with Nonlinear Decision Boundaries

The margin based classifiers, either soft or maximal margin classifiers are linear boundary classifiers. However, data points are not always linearly separable. Thus developing the need to develop a nonlinear decision boundary for proper classification. The feature space is required to be enlarged using functions of predictors, like quadratic or cubic projectors. An example of non-linearly separable data points is shown in Fig. A- 5.

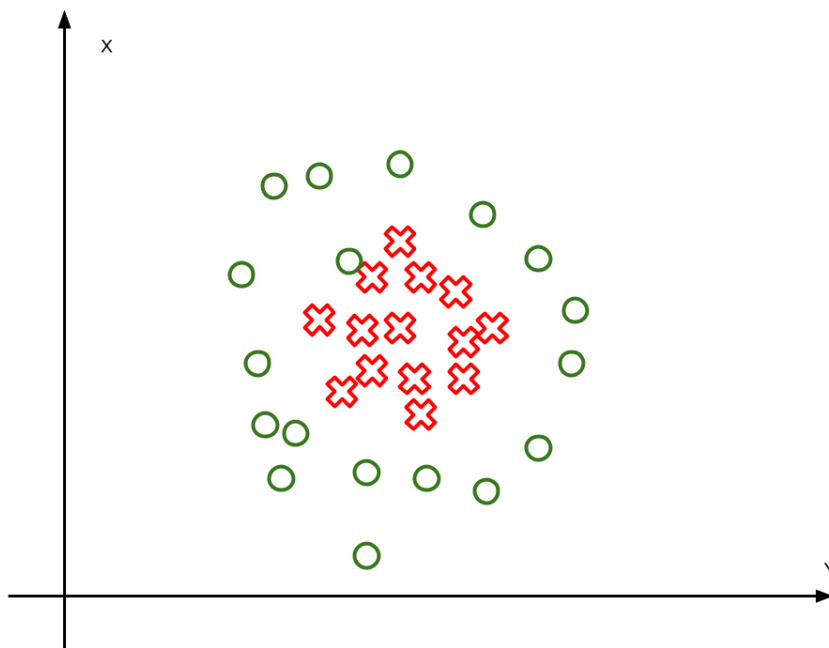


Figure A- 5 Example of data points distribution where linear separating boundary performs poorly

From the figure, it is visually evident that a linear hyperplane would be unable to separate the

two sets of data points. Intuitively a circular nonlinear separation boundary would be best able to separate the data points into their respective classes and thus can be used as the decision boundary for test data points, as shown in Fig. A- 6.

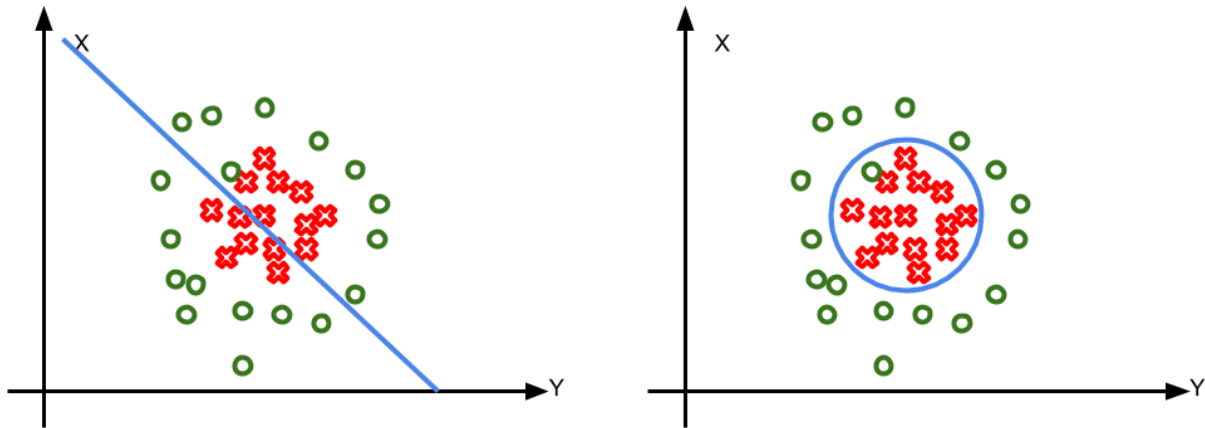


Figure A- 6 Better separation of data points using a nonlinear decision boundary compared to linear (separation boundary shown in blue)

Mercer's theorem states that if a function  $K(a, b)$  satisfies all the mercer's constraints [181] then there exists a function that can map  $a$  and  $b$  into a higher dimension.  $K(a, b) = \Phi(a)^T \cdot \Phi(b)$ , where  $\Phi()$  is known as the *kernel function*. It leverages the method of using the linear classifier to classify nonlinear data points. Mathematically, the above mentioned kernel function maps the input data points into a higher dimension making them more conducive to being separated by a linear separating hyperplane. Intuitively, if we use the kernel function that maps the points in Fig. A- 6 into a higher dimension  $z(x, y) = x^2 + y^2$  then the resulting hyperspace can be divided into binary classes by using a linear hyperplane as shown in Fig. A- 7.

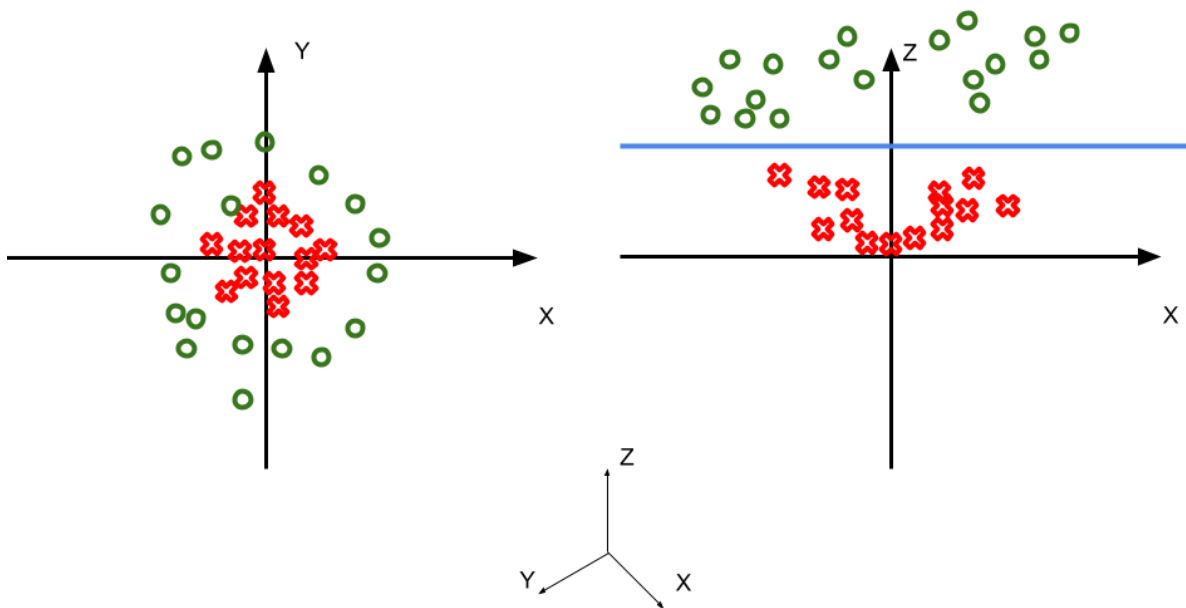


Figure A- 7 Projection into higher dimension based on quadratic function enables linear separation

Based on the above example, mathematically, with  $p$  observations  $x_1, x_2, \dots, x_p$  a vector classifier along the quadratic form of  $x_1, x_1^2, \dots, x_p, x_p^2$  can be fit. Thus, the optimisation problem mentioned in 9 and 12 becomes

$$\max M \text{ given } \beta_0, \beta_{11}, \dots, \beta_{p1}, \beta_{12}, \dots, \beta_{p2} \quad (32)$$

$$\text{subject to } \sum_{j=1}^p \sum_k^2 \beta_{jk}^2 = 1 \quad (33)$$

$$y_i(\beta_0 + \sum_{j=1}^p \beta_{j1}x_{ij} + \sum_{j=1}^p \beta_{j2}x_{ij}^2) \geq M(1 - \varepsilon_i), \text{ for all } i = 1, 2, \dots, n \quad (34)$$

$$\varepsilon_i \geq 0 \text{ and } \sum_{i=1}^n \varepsilon_i \leq C \quad (35)$$

The solution to the above equations will lead to the determination of the nonlinear decision boundary which would also be known as the support vector machine. In other cases the kernel can also be linear, radial basis function etc. In its most general form the support vector classifier can be represented as

$$f(x) = \beta_0 + \sum_{j=1}^p \alpha_j K(x_i, x_j) \quad (36)$$

where  $K(x_i, x_j)$  represents the generic kernel function.

The most popular kernel functions used in applications are,

1. Gaussian radial basis function, where  $K(x_i, x_j) = \exp(-\gamma||x_i - z_j||^2)$
2. Gaussian function, where  $K(x_i, x_j) = \exp(-||x_i - z_j||^2/2\sigma^2)$
3. Polynomial kernel function, where,  $K(x_i, x_j) = (x_i \cdot x_j + a)^b$
4. Linear kernel function, where  $K(x_i, x_j) = x_i \cdot x_j$

As a general empirical rule of thumb, in practical applications

- Computational requirement for SVM learning: linear < polynomial < radial basis function
- Performance in fitting any data: linear < polynomial < radial basis function
- Risk of overfitting the training data: linear < polynomial < radial basis function
- Risk of underfitting the training data: radial basis function < polynomial < linear
- Tunable hyperparameters: linear (0) < radial basis function (1) < polynomial (3)

## B ID3 Decision Trees

### B.I Information Entropy

Information entropy or simply entropy is a quantifiable measure of the amount of resources/capacity required to describe all the information in a sample. A homogenous sample with all similar elements results in an entropy of 0, while a purely random sample possesses the maximum entropy of 1. Mathematically, entropy is defined as

$$\text{Entropy } (E) = -\sum_{i=1}^n p_i * \log(p_i) \quad (37)$$

where  $p_i$  is defined as the probability of the sample belonging to class/category  $i$ , while  $i$  represents the different classes from 1 to  $n$ .

### B.II Gini Index/Gini Impurity

Gini index is a quantifiable measure of the inequality in the sample. Similar to entropy, a gini index of 0 signifies a homogeneous sample set with all samples belonging to the same class, while a value of 1 signifies maximal inequality among the elements of the sample set. Mathematically, it is the sum of probabilities of each class, illustrated as



$$Gini\ index\ (G) = 1 - \sum_{i=1}^n p_i^2 \quad (39)$$

where  $i$  represents the classes 1 to  $n$  of the sample set. Notably,  $0 \leq G \leq 1$ .

### B.III Decision Tree Algorithm

The algorithm itself can be visually represented as a tree, where each node in the tree represents the features (or attributes), each branch represents the decision (or the rule) and each leaf node represents the outcomes (classification, regression, whether discrete or continuous) as shown in Fig. B- 1.

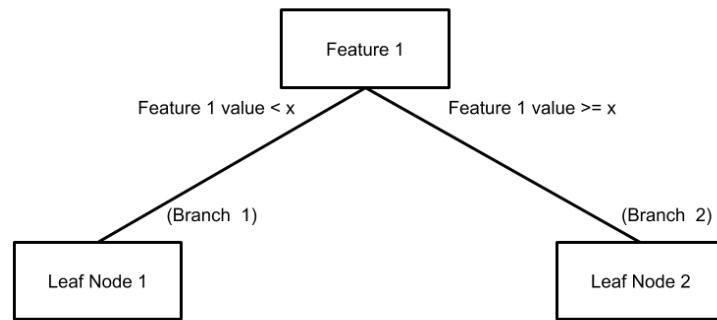


Figure B- 1 Schematic visual representation of a decision tree algorithm

DTs are implemented using a variety of algorithms. The most widely used being,

1. CART - Classification And Regression tree [182]
2. ID3 [183]
3. ID4.5 [184]

The two most widely used being CART and ID3, which differ in the metric used for quantifying impurity as well their their area of application.

### B.IV CART - Classification and Regression Tree

This generic DT learning algorithm can be used for both classification and regression. The training phase comprises minimising the Gini index as a measure of cost function to evaluate the quality of the split in feature selection in case of a classification tree. The direct application of CART is in binary classification. In case of regression, it uses least squares as a metric to select features. The case of CART algorithm shall be explained with a commonly used *weather* dataset.

The dataset describes the decision of playing (either Yes or No) given the outlook, temperature, humidity and wind conditions, which shall be treated as features/attributes.

Day	Outlook	Temperature	Humidity	Wind	Play Decision
1	Sunny	Hot	High	Weak	No

2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rainfall	Mild	High	Weak	Yes
5	Rainfall	Cool	Normal	Weak	Yes
6	Rainfall	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rainfall	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rainfall	Mild	High	Strong	No

Outlook is a nominal feature, which can take three values i.e. sunny, overcast or rain, summarised as follows

Outlook	Yes	No	Number of instances
Sunny	2	3	5
Overcast	4	0	4
Rainfall	3	2	5

The Gini index of *Outlook* is defined as follows

Outlook	Gini Index
Sunny	$1 - (\frac{2}{5})^2 - (\frac{3}{5})^2 = 1 - 0.16 - 0.36 = 0.48$
Overcast	$1 - (\frac{4}{4})^2 - (\frac{0}{4})^2 = 1 - 1 - 0 = 0$
Rainfall	$1 - (\frac{3}{5})^2 - (\frac{2}{5})^2 = 1 - 0.36 - 0.16 = 0.48$
<b>Weighted sum of Gini index</b>	<b><math>(\frac{5}{14}) * 0.48 + (\frac{4}{14}) * 0 + (\frac{5}{14}) * 0.48 = 0.342</math></b>

Similarly for temperature

Temperature	Yes	No	Number of instances
Hot	2	2	4
Cool	3	1	4
Mild	4	2	6

Temperature	Gini Index
Hot	$1 - (2/4)^2 - (2/4)^2 = 0.5$
Cool	$1 - (3/4)^2 - (1/4)^2 = 0.375$
Mild	$1 - (4/6)^2 - (2/6)^2 = 0.445$
<b>Weighted sum of Gini index</b>	<b><math>(4/14)*0.5 + (4/14)*0.375 + (6/14)*0.445 = 0.439</math></b>

Humidity	Yes	No	Number of instances
High	3	4	7
Normal	6	1	7

Humidity	Gini Index
High	$1 - (3/7)^2 - (4/7)^2 = 0.489$
Normal	$1 - (6/7)^2 - (1/7)^2 = 0.244$
<b>Weighted sum of Gini index</b>	<b><math>(7/14)*0.489 + (7/14)*0.244 = 0.367</math></b>

Wind	Yes	No	Number of instances
Weak	6	2	8

Strong	3	3	6
--------	---	---	---

Wind	Gini Index
Weak	$1 - (6/8)^2 - (2/8)^2 = 0.375$
Strong	$1 - (3/6)^2 - (3/6)^2 = 0.5$
<b>Weighted sum of Gini index</b>	<b><math>(8/14)*0.375 + (6/14)*0.5 = 0.428</math></b>

The final decision for the root node among all the features,

Features/Attributes	Gini Index
Outlook	0.342
Temperature	0.439
Humidity	0.367
Wind	0.428

From the table, the lowest Gini index of the *Outlook* feature is considered as the root node of the tree, as shown in Fig. B- 2.

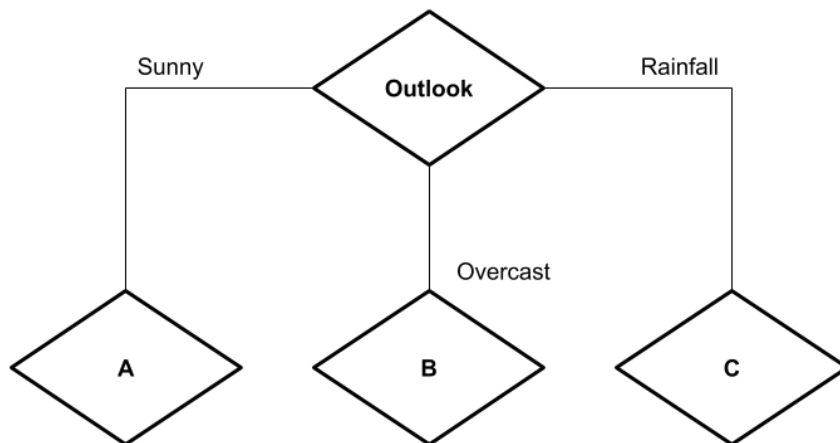


Figure B- 2 Decision tree with root node chosen based on Gini impurity

Next steps involve the corresponding weighted Gini impurity index for the subsets of data for the particular *Outlook* feature. For example for the Sunny instances, the subset is represented as

Day	Outlook	Temperature	Humidity	Wind	Play Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

Following the steps for the *Sunny Outlook* as before the final Gini index are

Features	Gini Index
Temperature	0.2
Humidity	0
Wind	0.466

Thus, the *Humidity* attribute results in pure splits of the dataset as can be seen from the perfect correlation between *Humidity* and *Play Decision*. Thus the resulting decision tree is represented in Fig. B- 3

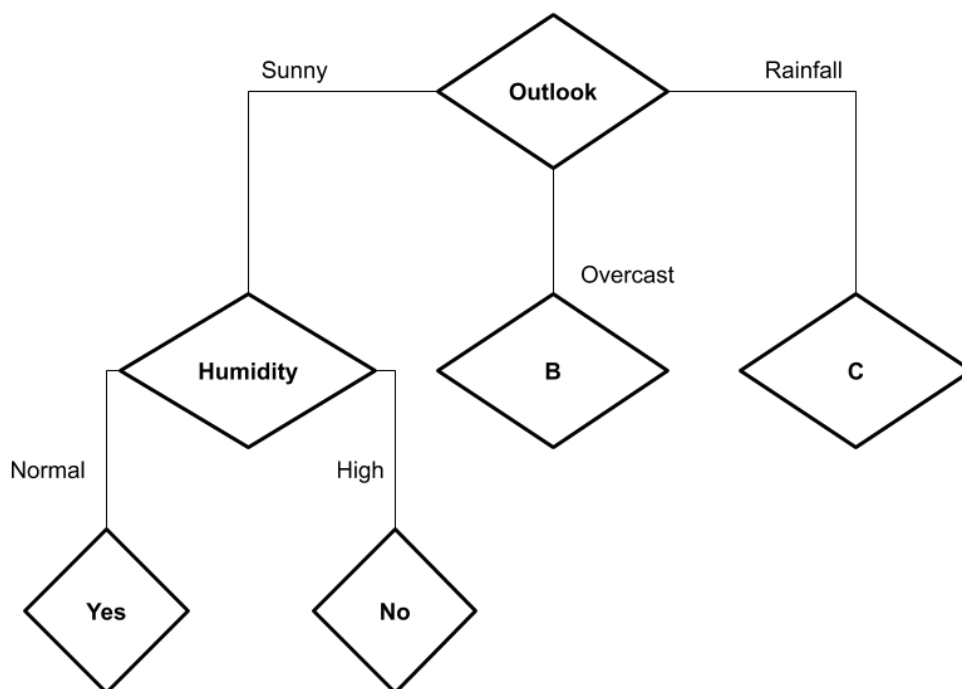


Figure B- 3 Decision tree after analysis of the Sunny Outlook subset

Following the same steps for all the features/attributes, the final decision trees can be represented as in Fig. B- 4. Detailed calculation of all the steps is omitted due to redundancy in conceptual development.

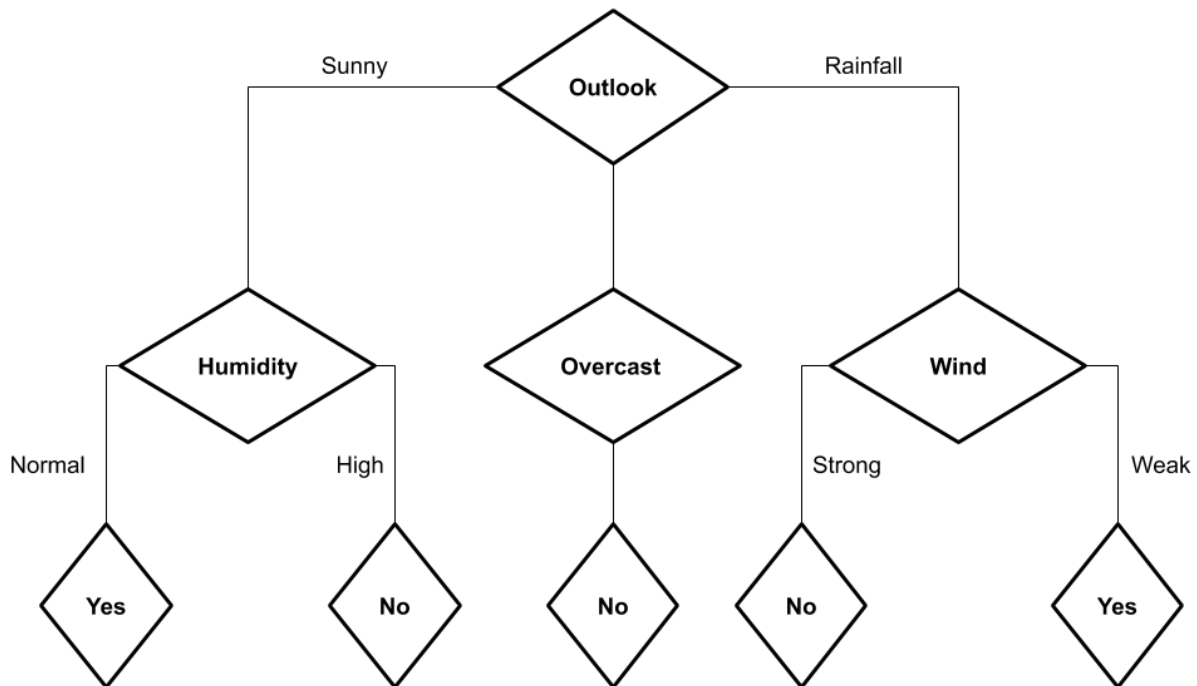


Figure B- 4 The final decision tree finishing Gini impurity analysis of all the features

### B.V ID3 - Iterative Dichotomiser 3

The ID3 algorithm was invented to generate a decision tree using *Entropy* as the metric for dataset impurity (as opposed to Gini index used by the CART algorithm). However, unlike the CART technique, ID3 is used for classification problems only. Feature splits are decided based on the amount of information gain obtained on splitting on a particular feature

$$Entropy (E) = - \sum_{i=1}^n p_i * \log(p_i)$$

$$Information\ Gain (S, A) = Entropy(S) - \sum P(S|A) * Entropy(S|A) \quad (40)$$

The feature chosen to split upon results in the highest information gain. As an example, the ID3 algorithm shall be applied to the previously mentioned dataset to build the concept.

From the dataset,

- Number of observations = 14
- Number of observations resulting in a ‘Yes’ decision = 9
- Probability of a ‘Yes’ decision  $p(Yes) = 9/14$
- Number of observations resulting in a ‘No’ decision = 5
- Probability of a ‘No’ decision  $p(No) = 5/14$

$$\text{Entropy}(\text{decision}) = -p(\text{Yes}) * \log_2 p(\text{Yes}) - p(\text{No}) * \log_2 p(\text{No})$$

Substituting values:  $\text{Entropy}(\text{decision}) = -(9/14) * \log_2(9/14) - (5/14) * \log_2(5/14) = 0.94$

Given the four attributes of *Outlook*, *Temperature*, *Humidity* and *Wind*, the Information Gain that would be obtained on splitting based on any of the attributes would provide us with the best attribute to split on. As an example, the steps involved in calculating the Gain for *Outlook - Sunny* is detailed here. The subset of data for *Outlook - Sunny* would look like the following (as presented earlier)

Day	Outlook	Temperature	Humidity	Wind	Play Decision
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

1. Number of instances with *Sunny outlook* = 5
2. Play Decision = 'Yes' given that *Outlook* is *Sunny*  $p(\text{decision} = 'Yes' | \text{Outlook} = 'Sunny') = 2/5$
3. Play Decision = 'No' given a *Sunny outlook*  $p(\text{decision} = 'No' | \text{Outlook} = 'Sunny') = 3/5$

$$\text{Entropy}(\text{decision} | \text{outlook} = \text{sunny}) = -(2/5) * \log_2(2/5) - (3/5) * \log_2(3/5) = 0.97$$

Similarly

$$\text{Entropy}(\text{decision} | \text{outlook} = \text{overcast}) = -(4/4) * \log_2(4/4) = 0$$

$$\text{Entropy}(\text{decision} | \text{outlook} = \text{rainfall}) = -(3/5) * \log_2(3/5) - (2/5) * \log_2(2/5) = 0.97$$

Thus

$$\begin{aligned} \text{Information Gain}(\text{decision}, \text{outlook}) &= \text{Entropy}(\text{decision}) - \\ & p(\text{decision} | \text{outlook} = \text{sunny}) * \log_2 p(\text{decision} | \text{outlook} = \text{sunny}) - \\ & p(\text{decision} | \text{outlook} = \text{overcast}) * \log_2 p(\text{decision} | \text{outlook} = \text{overcast}) - \\ & p(\text{decision} | \text{outlook} = \text{rainfall}) * \log_2 p(\text{decision} | \text{outlook} = \text{rainfall}) \end{aligned}$$

$$\text{Gain}(\text{decision}, \text{outlook}) = 0.94 - (5/14) * 0.97 - (4/4) * 0 - (5/14) * 0.97 = 0.247$$

The information gain obtained on splitting on either attributes,

- $\text{Gain}(\text{decision, outlook}) = 0.247$
- $\text{Gain}(\text{decision, wind}) = 0.048$
- $\text{Gain}(\text{decision, temperature}) = 0.029$
- $\text{Gain}(\text{decision, humidity}) = 0.151$

Thus the first node (root node) shall be based on *Outlook* given the highest gain in information. The subsequent steps are similar to the ones mentioned in the previous section, leading to the final decision tree structure of Fig. B- 4. Thus in this particular example case, leading to the same tree structure irrespective of the impurity metric considered.

## C Random Forests

As the name suggests *random forests* are a collection of *randomly developed decision trees*. The development of randomised trees is based on three popular strategies, (1) training the individual trees on bootstrapped versions of the original dataset, (2) randomly sampling a subset of the original features before optimising for split dimension and split location, and (3) randomly sampling candidate pairs of split dimensions and split locations and restricting the search to just these pairs [185]. Although originally proposed for supervised learning, random forests are extremely flexible and can be employed for other purposes such as density estimation, manifold learning and semi-supervised learning [186]. Unlike decision trees, random forests are less prone to overfitting on the training dataset.

The random forest model used in the thesis combines the bootstrapped method along with feature bagging to develop highly robust and accurate random decision trees. Given the availability of sufficient training data, the choice does not suffer from lack of enough training samples. Each decision tree in the random forest was generated to select, at each candidate split in the learning process, a random subset of the features also known as ‘feature bagging’. The ‘bootstrapping’ method selects a set of samples at random (with replacement) for training each decision tree. The generalisation accuracy was estimated on the left-out samples. The number of randomly chosen features for ‘feature-bagging’ was maintained to be the square root of the total number of features available, as suggested by the original inventors of the algorithm. The minimum number of samples required for proceeding with the split was kept at 2 to ensure pure or homogeneous sets at the leaf nodes. The features are randomly permuted at each split to maintain the same criterion across the splits. Gini impurity [79] is used as a quantitative measure of impurity of the samples. In the classification phase, the predicted class of an input sample is a vote by the trees in the forest, weighted by their probability estimates. That is, the predicted class is the one with the highest mean probability estimate across the trees.

## D Artificial Neural Networks

### D.I From Biological Neurons to Mathematical Model Neurons

In a biological neuron, the membrane of the neuron maintains the concentration differences of various ions between the inside and outside of the cell through active ion pumps and controllable ion channels. At rest, the channels are closed resulting in a net negative



potential of -70 mV on the inside of the cell compared to the outside fluid. When a sufficiently strong electrical excitation is provided, it results in a temporary less negative potential, triggering an opening of specific ion channels and leading to a chain reaction of other channels opening and/or closing. As a net result, an electrical peak of height around +40mV is generated for about 1 msec and propagates along the membrane at a speed of about 5 m/sec. This peak is also known as the action potential also known as the *spike* or an *impulse*. The spike is followed by a refractory period, when no excitation can occur. This action potential serves as an electric communication signal propagating and spreading along the output channel of the neuron, the axon and other connected neurons. The junction between an output channel of one neuron and the input channel of another neuron is known as the synapse. The propagating spike impulse maintains its shape and strength between neurons, due to the trigger of release of a chemical, a *neurotransmitter*, when the spike reaches a synapse. Upon reaching a synapse, the neurotransmitter selectively opens ion channels in the membrane of the receiving neuron. If the channel being opened is the  $Na^+$  (Sodium ion) channel, it results in an increase in the probability of the receiving neuron to start firing a spike impulse itself, also known as *excitatory synapse*. However if the channel being opened is the  $Cl^-$  (Chlorine ion) channel, it decreases the probability, also known as an *inhibitory synapse*. If the postsynaptic potential exceeds the neuron-specific threshold (of about -30 mV), it results in an action potential getting fired. The key to learning, adaptive and self-programming properties of the neuron is that the synapses and firing thresholds are not fixed and are updated on the fly.

Mathematically, the biological neuron is represented as a perceptron model. A perceptron takes some input values, called ‘features’ represented by  $x_1, x_2, x_3, x_4, x_5 \dots x_n$  weighted (through multiplication) with ‘weights’ represented as  $w_1, w_2, w_3, w_4 \dots w_n$  and combined linearly (through summation) over all the neurons to provide the output  $o = f(w_1x_1 + w_2x_2 + w_3x_3 + \dots w_ix_i + \dots w_nx_n)$ . The function  $f()$  is determined based on the type of perceptron neuron. The equation is represented visually in Fig. D- 1.

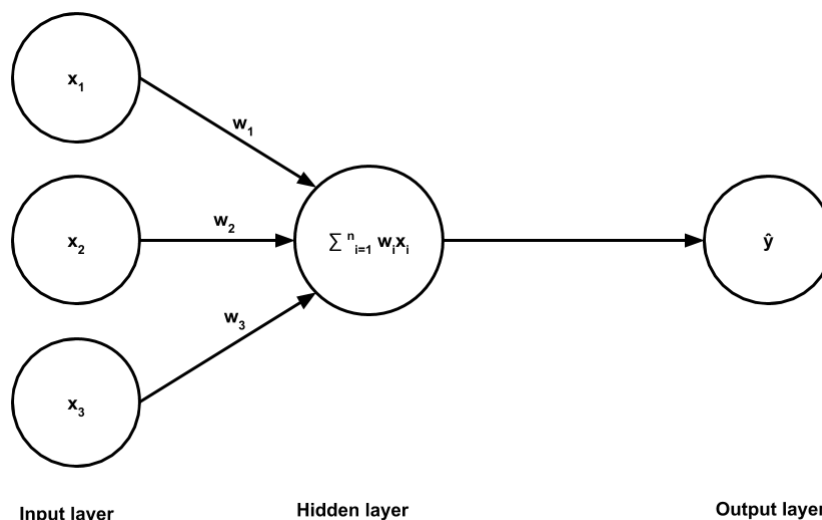


Figure D- 1 Mathematical model of a single perceptron-based neuron

Generalising the summation equation with the subscript  $i$ , the equation can be written as

$$\sum_{i=1}^n w_i x_i \quad (41)$$

As,  $x_i$  and  $w_i$  are of a fixed number of elements i.e. from 1 to  $n$ , they can be visualised as a vector of  $n$ -dimensions and the summation as a dot product of the two vectors.

$$x = [x_1, x_2, x_3, x_4, \dots, x_n] \quad (42)$$

$$w = [w_1, w_2, w_3, w_4, \dots, w_n] \quad (43)$$

$$w \cdot x \quad (44)$$

Extending this logic, in case of multiple neurons in the hidden layer, and additional subscript is used to describe the weights for each connection between the input-hidden, hidden-hidden and hidden-output layers, as shown in Fig. D- 2. This architecture is known as the Multilayer Perceptron architecture (MLP).

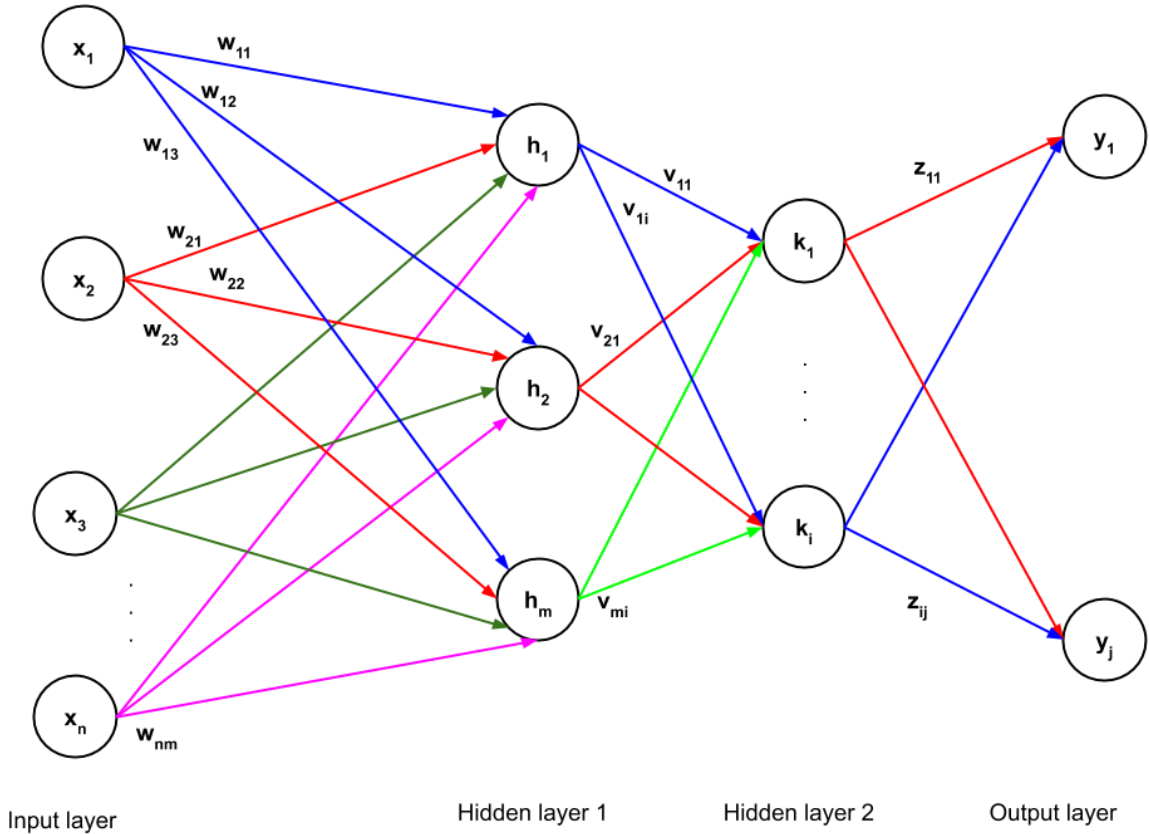


Figure D- 2 Multi layer perceptron (MLP) based artificial neural network architecture

As a generalised extension of the previous equations, for MLP,

$$h = f_1(w \cdot x), k = f_2(v \cdot h), y = f_3(z \cdot k)$$

where  $f_1, f_2$  and  $f_3$  are the nonlinear activation functions between the input-hidden, hidden-hidden and hidden-output layers.  $y$  represents the output vector from the network.

In addition to the *weights* components, each node also possesses a *bias* component to provide an offset to the input to that node. Similar to weights, the biases are also tunable parameters that can be learned, changing the equations to,

$$h = f_1(w \cdot x + b_1), k = f_2(v \cdot h + b_2), y = f_3(z \cdot k + b_3)$$

where  $b_1, b_2$  and  $b_3$  are the bias vectors for *Hidden layer 1*, *Hidden layer 2* and the *Output layer*. The nonlinear activation functions are also known as the *transfer functions* and are essential to the introduction of nonlinearity in the functional mappings between the inputs and outputs (also

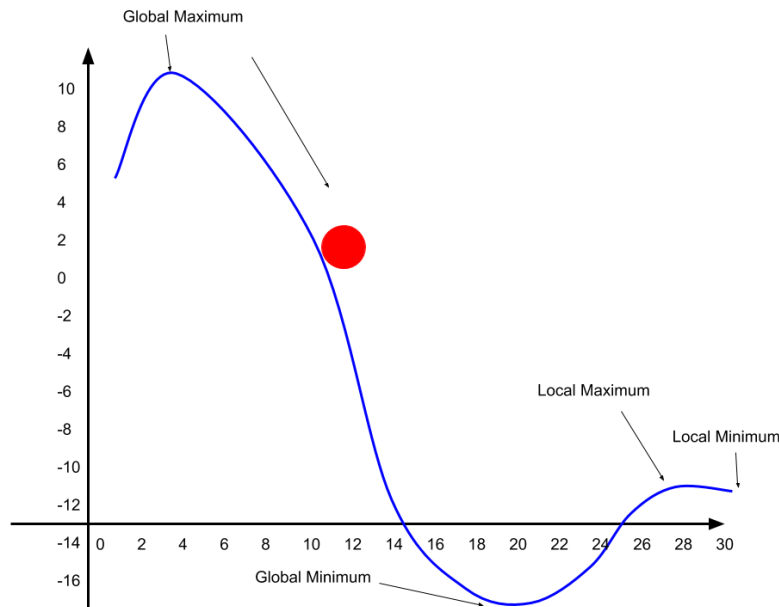
known as labels or response variables). In the absence of the activation function, the mapping would be linear, limiting the complexity of mapping that can be learned. This feature of ANNs has credited them with the term *Universal Function Approximators* given their ability to learn any nonlinear function given enough complexity and data. A requirement of the activation function is its differentiability. The differentiated errors are used for back propagation of the computed gradients of error (loss) values with respect to the weights for optimisation of the weight values using either Gradient Descent or any other optimisation technique for error reduction. The most commonly used activation functions are, (1) sigmoid or logistic, (2) tanh or hyperbolic tangent and (3) ReLU or Rectified Linear Units. Mathematically the functions have the following definition,

$$\begin{aligned} \text{sigmoid}(x) &= 1 / (1 + e^{-x}) \\ \text{tanh}(x) &= (1 - e^{-2x}) / (1 + e^{-2x}) \\ \text{ReLU}(x) &= \max(0, x) \end{aligned}$$

The output of the final activation function is provided as an input to a *softmax* [161] function to obtain the probabilities of belonging to a class, for classification problems. In case of a regression problem, a simple linear function is used.

## D.II Learning in Artificial Neural Networks

As mentioned in the previous section, each neuron sums all the inputs from the previous layer, nonlinearly activates the output and passes directly to all neurons in the next layer. As values are sent from one layer to the next, a weight is assigned to each interconnecting line and is multiplied by the values. Values of the interconnecting weights predetermine the neural network's computation reaction to any arbitrary input pattern. As information is passed forward from the inputs towards the outputs, interconnecting weights are adjusted by a *back-propagation* algorithm during the learning/training phase so that known outputs will best match predicted outputs (also known as labels). The back-propagation is the most common algorithm used to train ANNs due to its ability to generalise well on a wide variety of problems. Training of an ANN is the procedure by which the values for the individual weights (and biases) are determined such that the relationship between the inputs, predicted outputs and labels of the network are modelled accurately. By varying the weights and determining the errors of classification (based on the modified weights) an error surface is developed. The dimensionality of the error surface is determined by the number of tuneable weights in the network. A simplified two-dimensional error surface is shown in Fig. D- 3 to illustrate the concept.



*Figure D- 3 A schematic representation of an error surface*

The objective of training a neural network is to find the combination of weights which will result in the smallest error. In practice, it is not possible to plot such a surface due to the multitude of weights and hence it is more relevant to find the minimum point of the error. One possible technique is to use a procedure known as gradient descent. The backpropagation training algorithm uses this procedure to attempt to locate the global minimum of the error surface. The backpropagation algorithm is the most computationally straightforward algorithm for training the multilayer perceptron. Back-propagation has been shown to perform adequately in many applications. A great number of applications discussed in this paper used back-propagation to train ANNs. Conceptually, the back-propagation algorithm is an iterative process involving a series of steps that are repeated sequentially until the desired error function value is obtained. The steps include, (1) network initialisation (or preparation), (2) generation of the input vector from training data, (3) propagation of the input vector through the network, (4) calculation of error signal value, (5) back propagation of the error through the network as delta changes to the weight to minimise the error, (6) adjustment of the weight values for reduction of the error, and (7) iteration through the same procedures for further reduction in the error.

The above mentioned backpropagation algorithm is also known as online training whereby the network weights are adapted after each pattern has been presented. The alternative is known as batch training, where the summed error for all patterns is used to update the weights. In practice, many thousands of training iterations will be required before the network error reaches a satisfactory level as determined by the problem being addressed. Training should be stopped when the performance of the ANN on the independent test data reaches a maximum.

### D.III Error Function

Classification error function used for calculating the deviation of the predicted classes from the actual classes (labels) have a few options within them. The error (also known as cost

function) function to use has been empirically tried and tested and the most popular list of functions collated.

For regression, the most common function is the *Mean Squared Error (MSE)* approach that is used to find the deviation from actual numerical value,

$$MSE = 1/n * \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where  $n$  represents the number of nodes in the output layer,  $y_i$  represents the actual regressed value that should be output, and  $\hat{y}_i$  represents the predicted regressed value.

For classification, the most effective error function has been empirically determined to be the *cross-entropy* function, also referred to as the *logarithmic loss* function. In binary classification, if the number of classes equals 2, then cross-entropy can be calculated as,

$$-(y \cdot \log(p) + (1 - y)\log(1 - p))$$

and if the number of classes,  $M$  is greater than 2 (i.e. multiclass classification), then the error (or loss) is calculated for each class label per observation and the result is summed,

$$-\sum_{c=1}^M y_{o,c} \log(p_{o,c})$$

where  $M$  represents the number of classes,  $\log$  is the natural logarithm,  $y$  is the binary indicator (0 or 1), if class label  $c$  is the correct classification for observation  $o$ ,  $p$  is the predicted probability observation  $o$  of class  $c$ . In addition to the above functions, other loss functions include, (1) huber loss - typically used for regression. It's less sensitive to outliers than the *MSE* is, as it treats the error as the square only inside a given interval, (2) hinge loss - typically used for classification, and (3) mean absolute error (L1) represented as  $\sum_{i=1}^n |\hat{y}_i - y_i|$ , where  $\hat{y}_i$  represents the actual class labels,  $y_i$  represents the predicted class labels and  $n$  represents the number of class nodes

#### D.IV Backpropagation for updating the weights

Backpropagation was invented in the 1970s as a general optimisation method for performing automatic differentiation of complex nested functions. However, it wasn't until 1986, with the publishing of a paper by Rumelhart, Hinton, and Williams, titled "Learning Representations by Back-Propagating Errors," that the importance of the algorithm was appreciated by the machine learning community at large [187]. Backpropagation was one of the first methods able to demonstrate that artificial neural networks could learn good internal representations, i.e. their hidden layers learned nontrivial features. Experts examining multilayer feedforward networks trained using backpropagation actually found that many nodes learned features similar to those designed by human experts and those found by neuroscientists investigating biological neural networks in mammalian brains (e.g. certain nodes learned to detect edges, while others computed Gabor filters [188]). Even more importantly, because of the efficiency of the algorithm and the fact that domain experts were no longer required to discover appropriate features (this diverting away from the expert system based artificially intelligent systems), backpropagation allowed artificial neural networks to be applied to a much wider field of problems that were previously off-limits due to time and cost constraints.

Backpropagation requires three things:

- Dataset  $(x_i, y_i)$ , where  $x_i$  is the input and  $y_i$  is the desired output of the network on the input. The set of input-output pairs of size  $n$  is denoted as  $X = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$
- An ANN whose parameters are collectively denoted by  $\theta$ . In backpropagation, the parameters of importance are  $w_{ij}^k$ , the weights between the node  $j$  in layer  $k$  and node  $i$  in layer  $k - 1$ , and  $b_i^k$ , the biases for node  $i$  in layer  $k$ .
- Error function represented as  $E(X, \theta)$ , which defines the error between the desired output  $y_i$  and the predicted/calculated output  $\hat{y}_i$  of the neural network on the input  $x_i$  for a set of input-output pairs and a particular value of the parameters.

Training a neural network with gradient descent requires the calculation of the *error deltas*. The deltas define the contribution of each of the tuneable parameters such as the weights  $w_{ij}^k$  and the biases  $b_i^k$  gradient to the value of the error function. Subsequently, the deltas are multiplied with a learning rate  $\alpha$  for updating the weights and biases in each iteration, according to,

$$\theta^{t+1} = \theta^t - \alpha \frac{\partial E(X, \theta^t)}{\partial \theta}$$

where  $\theta^t$  collectively denotes the tunable parameters of the neural network at iteration  $t$  in gradient descent.

The following terms are utilised in a general formulation of a neural network,

- $w_{ij}^k$ : weight for node  $j$  for the layer  $k$  for incoming node  $i$
- $b_i^k$ : bias for node  $i$  in layer  $k$
- $a_i^k$ : product sum plus bias (activation) for node  $i$  in layer  $k$
- $o_i^k$ : output for node  $i$  in layer  $k$
- $r_k$ : number of nodes in layer  $k$
- $g$ : activation function for hidden layer nodes
- $g_o$ : activation function for the output layer nodes

The general backpropagation is dependent on the following five equations,

- For the partial derivatives,  $\frac{\partial E_d}{\partial w_{ij}^k} = \delta_j^k o_i^{k-1}$
- For the final layer's error term,  $\delta_1^m = g'_o(a_1^m)(\hat{y}_d - y_d)$
- For the hidden layer's error terms,  $\delta_j^k = g'(a_j^k) \sum_{l=1}^{r^{k+1}} w_{jl}^{k+1} \delta_l^{k+1}$
- For combining the partial derivatives for each input-output pair,  $\frac{\partial E(X, \theta)}{\partial w_{ij}^k} = \frac{1}{n} \sum_{d=1}^n \frac{\partial}{\partial w_{ij}^k} \left[ \frac{1}{2} (\hat{y}_d - y_d)^2 \right] = \frac{1}{n} \sum_{d=1}^n \frac{\partial E_d}{\partial w_{ij}^k}$
- For updating the weights,  $\Delta w_{ij}^k = -\alpha \frac{\partial E(X, \theta)}{\partial w_{ij}^k}$

The general backpropagation algorithm proceeds in the following way,

- Calculation of the forward phase for each input-output pair  $(x_i, y_i)$  and storage of the results

- Calculation of the backward phase for each input-output pair and store the results for each weight connecting the nodes (within the layers)
  - Evaluation of the error for the final layer  $\delta_1^m$  by using the second equation mentioned above
  - Backpropagation of the error terms for the hidden layers, working backwards from the final hidden layer, by repeatedly using their equation
  - Evaluation of the partial derivatives of the individual errors with respect to the weights by using the first equation
- Combination of the individual gradients for each input-output pair to get the total gradient for the entire set of input-output pairs by using the fourth equation
- Updating the weights according to the learning rate  $\alpha$  and a total gradient by using the fifth equation (moving in the direction of the negative gradient)

The backpropagation algorithm is repeated for multiple iterations for multiple batches of data, till stable minimum error point is reached. The accepted error values then leads to a trained neural network. The exact nature of the equations are highly dependent on the nature of the activation function, the number of layers in the neural network and the number of nodes per layer.

## E Artificial Neural Network Architectures

In addition to the multi-layer perceptron (MLP) model, the two major biologically inspired paradigms of architectures are (1) convolutional neural networks, and (2) recurrent neural networks.

### E.I Convolutional Neural Networks (CNN)

Convolutional neural networks (CNN) [30] are very similar to ordinary neural networks with the addition of multiple layers convolution kernel filters, which are learned. The presence of learnable kernel functions optimises the network for image analysis. In addition, the presence of an enormous amount of layers and weights (and biases) per layer makes it a ‘deep learning’ neural network. The architecture of a CNN is analogous to that of the connectivity pattern of neurons in the human brain and was inspired by the organisation of the *visual cortex*. Individual neurons respond to stimuli only in a restricted region of the visual field known as the *receptive field*. A collection of such fields overlap to cover the entire visual area. Similar to the biological neural connections in the visual cortex, the convolutional neural network is designed to capture the spatial correlations across the regions in the image for a particular objective such as object identification or pattern recognition.

The mathematics of the CNN is very similar to the MLP, however the arrangement of the neurons is what makes it unique. The first layer of the CNN is a convolutional layer. Intuitively, it can be visualised as a flashlight shining across an image and sliding across it. In ML terms, the flashlight is called the *filter* (also known as the kernel) and the region that it is shining over is called the *receptive field*. The filter (kernel) is a set of numerical values represented in the form of an array. The depth of array is usually the same as the depth of the image its analysing. For example, in the case of a normal RGB image, the depth of both the image array and the kernel

array is 3. As the filter slides over the image, it convolves with the underlying receptive field, though matrix multiplication (element wise multiplication) to result in a single numerical value, which in turn forms a part of the filtered image. This process is repeated multiple times, similar to the presence of multiple layers of neurons in the human brain’s visual cortex (V1 through V5) to learn increasingly complicated features with increasing filtration (through kernels) of the original image. The resulting array of numbers after the final filtration is called the activation map or the feature map, which is then connected to the output layer. The results of each convolutional layer is usually activated through a nonlinear function (such as ReLU). It is common to periodically insert a *Pooling layer* in-between successive convolutional layers in a CNN. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control overfitting. The Pooling Layer operates independently on every depth slice of the input and resizes it spatially, using the MAX operation (choosing the maximum value). The process is visually represented in Fig. E- 1.

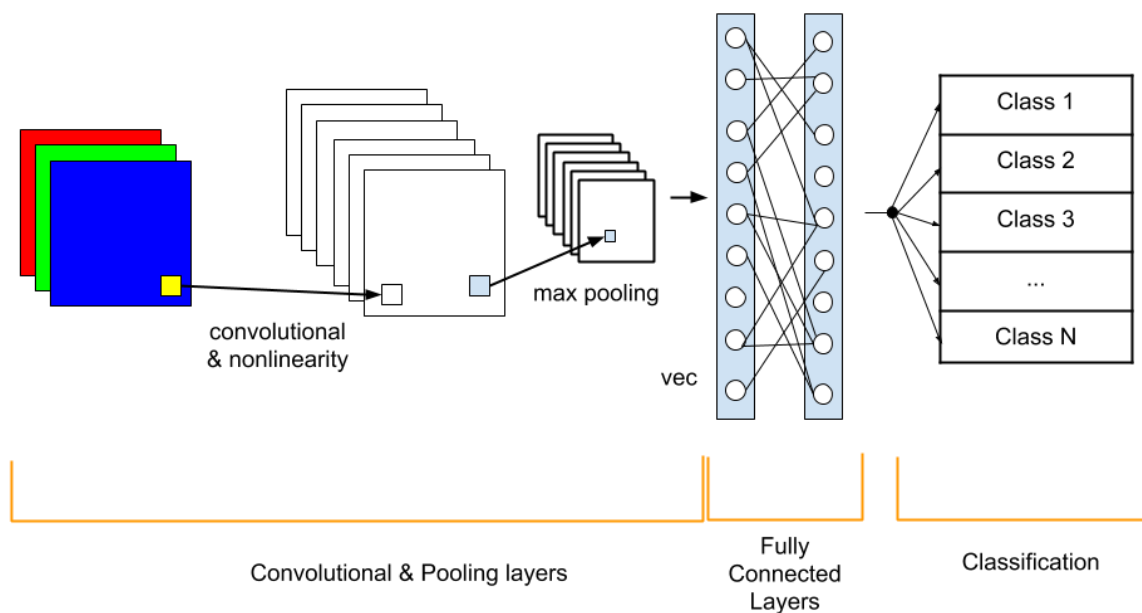


Figure E- 1 Schematic representation of a convolutional neural network

The most popular case studies for CNNs are the ones that have performed exceptionally well in competitions such as Image Recognition in ILSVRC [189] such as,

- **LeNet.** One of the first successful applications of Convolutional Networks were developed by Yann LeCun in 90’s. One of the most well-known of them being LeNet [190] architecture that was used to read zip codes, digits, etc.
- **AlexNet.** The first network that shot CNN to fame in the Computer Vision community was the AlexNet [29] , developed by Alex Krizhevsky, Ilya Sutskever and Geoff Hinton. The AlexNet was a submission in the ImageNet ILSVRC challenge in 2012 and significantly outperformed the second runner-up (top 5 error of 16% compared to runner-up with 26% error). The Network had a very similar architecture to LeNet, but was deeper, has more learnable parameters, and featured convolutional layers stacked on top of each other immediately followed by a pooling layer.



- **ZF Net.** The ILSVRC 2013 winner was a Convolutional Network from Matthew Zeiler and Rob Fergus. It became known as the ZFNet (short for Zeiler & Fergus Net). It was an improvement on AlexNet by adjusting several architectural hyperparameters, especially by expanding the size of the middle convolutional layers and making the stride and filter size on the first layer smaller.
- **GoogLeNet.** The ILSVRC 2014 winner was an official CNN from Google developed by Szegedy et al. [191]. It initiated the development of an Inception Module that dramatically reduced the number of parameters in the network (4M, compared to AlexNet with 60M). Additionally, this paper uses Average Pooling instead of Fully Connected layers at the top of the CNN, eliminating a large amount of redundant parameters. GoogLeNet was followed up by several versions, most notably Inception-v4.
- **VGGNet.** VGGNet. [192], from Karen Simonyan and Andrew Zisserman is the runner-up in ILSVRC 2014. It emphasised that network depth was an important factor in performance. Their final best network contains 16 CONV/FC layers and, appealingly, features an extremely homogeneous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end. A downside of the VGGNet is that it is more expensive to evaluate and uses a lot more memory and parameters (140M). Most of these parameters are in the first fully connected layer, and it was since found that these FC layers can be removed with no performance downgrade, significantly reducing the number of necessary parameters.
- **ResNet.** Residual Network [193] developed by Kaiming He et al. was the winner of ILSVRC 2015. It features special skip connections and an impressive use of batch normalisation [194]. The architecture ignores the presence of a fully connected network at the end of the CNN. ResNets are currently by far the state of the art CNN and are the default choice for using CNNs in practice (as of May 10, 2016).

## F Canonical Problems in Machine Learning

Table F- 1 succinctly presents the canonical problems encountered in machine learning as well as the typical models used to address the problems.

Canonical Problem	Machine Learning Models	Example Application
Classification	Support vector machines Decision trees Random forest Neural networks Boosted trees Nearest neighbour	<i>Medical diagnosis:</i> Does this tissue show signs of disease? What can we say about the person from his/her walking style? <i>Banking:</i> Is this transaction fraudulent? <i>Computer vision:</i> what type of object is in this picture? Is it a person? Is it a building?
Regression	Simple linear regression Polynomial regression Support vector regression	<i>Finance:</i> what is the value of this stock going to be tomorrow? <i>Housing:</i> what would the price of this

	Decision tree regression Random forest regression	house be if it were sold today? <i>Food quality</i> : how many days before this strawberry is ripe? <i>Image processing</i> : how old is the person in this photo? <i>Rehabilitation</i> : Given the speed and state of recovery, how long might it take for the patient to get back to normal life?
Clustering	K-means Mean-shift DBSCAN EM clustering using GMM Agglomerative Hierarchical clustering	<i>E-commerce</i> : which customers are exhibiting similar behaviour to each other, how do they group together? <i>Video Streaming</i> : what are the different types of video genres in our catalogue, and which videos are in the same genre?
Dimensionality Reduction	Principal component analysis Factor analysis Linear Discriminant Analysis Multidimensional scaling Isometric feature mapping Locally linear embedding Hessian Eigen mapping Autoencoders	<i>E-commerce</i> : what combinations of features allow us to summarise the behaviour of our customers? <i>Molecular biology</i> : how can scientists summarise the behaviour of all 20,000 human genes in a particular diseased tissue? <i>Gait Analysis</i> : What combinations of joint motions allows us to diagnose someone with the least amount of computation and which motions are therefore redundant?
Semi-supervised Learning	Generative models Low-density separation Graph-based methods Heuristic approaches	<i>Computer vision</i> : how can object detection be developed, with only a small training data set? <i>Drug discovery</i> : which of the millions of possible drugs could be effective against a disease, given we have so far only tested a few?
Reinforcement Learning	Monte Carlo Q-learning Deep Q network Deep reinforcement learning Inverse reinforcement learning Apprenticeship learning	<i>Robots</i> : how can a robot move through its environment? <i>Games</i> : which moves were important in helping the computer win a particular game? <i>Gait Rehabilitation</i> : Having tested rehabilitation techniques on people with gait-impairments, which rehabilitation techniques should be used for optimal rehabilitation, given this patient's gait?

*Table F- 1 Canonical problems encountered in machine learning*

## G Classification Machine Learning Algorithms

### G.I (Regularised) Logistic Regression

Logistic regression [195] is the classification counterpart to linear regression. Classifications are mapped to be between 0 and 1 through the logistic function, enabling classifications to be interpreted as class probabilities.

The models themselves are still ‘linear’, thus they work at their best when the classes are linearly separable (i.e. they can be separated by a single decision surface). Logistic regression can also be regularised by penalising coefficients with a tuneable penalty strength.

- **Strengths:** Outputs have an easily interpretable probabilistic distribution, and the algorithm can be regularised to avoid overfitting. Logistic models can be updated easily with new data using stochastic gradient descent.
- **Weaknesses:** Logistic regression tends to underperform when there are multiple or non-linear decision boundaries. They are not flexible enough to capture more complex relationships naturally.

### G.II Classification Trees (Ensembles)

Classification trees are the classification counterparts to regression trees. They are both commonly referred to as "decision trees" or by the umbrella term "classification and regression trees (CART) [122]"

- **Strengths:** They are robust to outliers, scalable, and able to naturally model non-linear decision boundaries owing to their hierarchical structure.
- **Weaknesses:** Unconstrained, individual trees are prone to overfitting, but can be alleviated by ensemble methods.

### G.III Deep Learning (Artificial Neural Networks)

Deep learning is can also be easily adapted to classification problems. In practice, deep learning techniques are used more commonly for classification tasks compared to regression.

- **Strengths:** Deep learning performs very well when classifying for audio, text, and image data. It is inherently capable of handling both static and dynamic data.
- **Weaknesses:** As with regression, deep neural networks require very large amounts of data to train, thus, is not treated as a general-purpose algorithm.

### G.IV Support Vector Machines (SVMs)

Support vector machines (SVM) use kernels [181], to calculate the distance between two observations. The SVM algorithm maximises the distance between the closest members of separate classes, to place the classification decision boundary.

For example, an SVM with a linear kernel is similar to logistic regression, where the SVM places a linear hyperplane such that the distance of each of the observations from the hyperplane is maximised. It then uses the linear hyperplane to decide the classes of the test dataset. Thus, in practice, the benefit of SVMs typically come from using non-linear kernels to model non-linear decision boundaries.

- **Strengths:** SVMs can model non-linear decision boundaries, and have a plethora of nonlinear kernel options to choose from. The hyperplanes are fairly robust against overfitting, especially in high-dimensional space.
- **Weaknesses:** SVMs are memory intensive, harder to tune due to the importance of choice of the kernel, and don't scale well to larger datasets. Typically random forests or artificial neural networks are usually preferred over SVMs. The latter having the advantage of using multiple nonlinear activations to map the nonlinearities between the dataset features.

In machine learning, no one algorithm works best for every problem, and it's especially relevant for supervised learning (i.e. predictive modelling). This is also known as the "No Free Lunch" theorem [196]. For example, one can't say that neural networks are always better than decision trees or vice-versa. There are many factors at play, such as the size and structure of the dataset and amount of computation available. As a result, many different algorithms should be tried for a particular problem, while using a holdout "test set" of data to evaluate performance and select the final solution.

## H Use of Machine Learning in Clinical Gait Analysis

Machine learning techniques have been used for gait classification, person recognition from movement patterns using basic, kinetic and kinematic gait data, and for gait event detection. The most common ML technique used in gait analysis by far is Support Vector Machines (SVMs). The results have been very promising in terms of person recognition, diagnosis of neurodegenerative diseases, especially in ageing population and studying the cause of gait degeneration with increasing age. It is well established that ageing, diseases and disorders influence gait patterns and considerable research has documented changes during unobstructed and obstructed walking that suggest age-related declines in lower limb control [197]. The major aim has been to identify key variables of gait degeneration in elderly individuals that might be predictors of falling behaviour due to loss of balance or use of gait as a measurement for efficacy of treatment of a certain disease. Research has shown that significant changes in gait can occur with age and disease in temporal and distance measures such as gait velocity, stride length, and stance and swing phase times [21]. In addition, foot-ground reaction force data during braking and propulsive phases [198] and joint angular motion data such as the ankle, knee and hip joint angles have shown effects of aging. To date, however, the relative influence of these measures in differentiating the age groups has not been demonstrated.

Systems facilitating robust, automatic identification of persons have gained increasing acceptance during recent years. Systems for automatic identification play a decisive role in

surveillance scenarios (e.g., monitoring high security areas like banks or airports). Biometric techniques use characteristic physiological and behavioural specifics of different persons for identification. Examples for such techniques are the recognition of iris, face, fingerprint, gait, or the handwriting. Gait classification is a relatively new biometric technique. Using gait as a biometric gained increasing attention during the past years, since it offers many advantages compared to other biometrics [199–202]. Allowing for marker-free person identification is one of the key advantages of using gait as a biometric [201] In contrast to systems using for example the iris or the fingerprint of a person as biometrics, identification by gait does neither require the cooperation nor the attention of the subject. To this end Kinect enables skeleton-detection and tracking of people in real time by an integrated depth camera. In comparison with traditional person recognition techniques, gait based person recognition has the advantage of providing reliable results even if the face is occluded from the camera view or the natural fingerprint altered in any way. Because of the high affordability of RGBD cameras (like the Kinect sensor), it could easily be used for surveillance, implementing multiple person recognition techniques.

Techniques like SVMs have been employed in identifying neurodegenerative diseases like multiple sclerosis, Huntington’s disease, Parkinson’s disease, stroke and osteoarthritis [203]. But the accuracy of detection is usually limited by the representation of gait as the few hand crafted feature vectors used for classification. Techniques like SVMs, Random Forests and Bagged and Truncated Decision Trees rely on the feature vector directly, assuming complete representation of the principal components of the sample space by the feature vector. Thus, unless a dimensionality reduction step is implemented, such techniques do not consider the interdependencies of the features.

Automated classification of gait pattern changes by a machine classifier from their respective measures is expected to offer many potential advantages. For example, Maki [204] using spatial-temporal measures of gait has shown significant changes in gait characteristics in the elderly fallers when compared to gait characteristics of elderly non-fallers. This research has particularly shown that some foot placement gait measures (e.g., step width and stride variability) displayed greater associations with falls classification. Therefore, early identification of gait changes due to falling behaviour by a machine classifier might trigger initiation of necessary measures to prevent injurious falls such as an exercise intervention program. Similar benefits could also be obtained in a clinical context via classification of abnormality in gait patterns and also by evaluating the effectiveness of treatment outcomes. In order to facilitate automated classification of gait patterns, neural networks and fuzzy clustering techniques have been applied for classification of normal and pathological gait [198,205], and also to differentiate gait simulations, such as leg length discrepancy from joint-angle measures [206]. However, it is well known that there are several limitations of neural network-based modelling, including: (i) dependency on a large number of parameters, e.g., network size, learning parameters and selection of initial weights, (ii) the possibility of being trapped into local minima, and (iii) overfitting on training data resulting in poor generalisation. Support vector machines (SVM) have emerged as a powerful technique for general purpose pattern classification. It has been applied to classification and regression problems with exceptionally good performance on a

range of binary classification tasks [11,73,207,208]. The primary advantage of SVM is its ability to minimise both structural and empirical risk leading to better generalisation for new data classification.

Although the use of advanced technology has led to increased use of gait in biometrics, the applications of said technology as a diagnostic and rehabilitation tool is still lacking; metrology of gait being a huge bottleneck to its advancement. The lack of established quantitative reporting practices combined with the lack of normative data have hindered the exploration of its deterministic nature in diagnosis of neurodegenerative diseases and rehabilitation monitoring measure. Because of a lack of publicly available normative data and gait-analysis results based diagnostics or rehabilitation objectives, the interdependence and contribution of the gait metrics towards the final objectives is either unknown or highly qualitative, hindering its candidature for automation as well.

One of the research objectives is to advance the metrics that are derived from gait, beyond the standard measures such as gait velocity, stride length and swing stance ratio, starting with narrow focus on stroke detection, mobility and rehabilitation and generalising towards other gait-impairing conditions. The achievement of this objective will not only result in a more detailed gait analysis report which would give deeper insights into the state of a patient to the clinician, but also, act as a highly reliable feature space for use in various classification tasks. Given the highly deterministic nature of diagnosis of diseases through gait analysis, the advanced feature space would enable highly accurate automated diagnosis of diseases, even at a very early symptom stage. In addition, keeping track of the metrics would provide useful insights into the recovery of the patient (thus, the efficacy of the rehabilitation measures) with time. Fortunately, the advent of the deep learning space combined with advanced sensors and the availability of high computation capabilities of machines, makes the topic a prime target for research at this time.

The recent emergence of deep learning architectures have refocused the attention on artificial neural networks again. Artificial Neural Networks have the added advantage of finding multiple non-linear correlations between the features and reducing the dimensionality of the feature space, by preserving the principal components [209]. Moreover, with the increase of computation capabilities of systems of late and advancements in GPU and increased implementation of GPU based parallel processing programming, deep learning neural networks have been applied and tested thoroughly on various datasets and have proved their efficacy over almost all of the classical machine learning techniques, including perceptron based feedforward neural networks. The exploration of applications of deep learning networks in gait analysis is still in its infancy and is a major part of this thesis.

# I Ethics Application, Protocol and Consent Forms

PSEC Application Form V4



## Application Form for Physical Sciences Ethics Committee Approval

### *Advice for applicants on completing the form*

*Please ensure that the information provided is:*

- *Accurate and concise*
- *Clear and simple and easily understood by a lay person*
- *Free of jargon, technical terms and abbreviations*

*Further advice and information can be obtained from your departmental representative on the PSEC and at: <http://www.york.ac.uk/admin/aso/ethics/ctee.htm>*

*Please return completed (typed) form to your departmental representative via email to:*

elec-ethics@york.ac.uk

### ***Title of project:***

Extended sensing for treadmill-based motion capture and gait analysis

### **SECTION 1 DETAILS OF APPLICANTS**

#### **Details of principal investigator (name, appointment and qualifications)**

Lewis Bellwood, Student, 3<sup>rd</sup> Year Electronic Engineering (BEng)

#### **Names, appointments and qualifications of additional investigators (student applicants should include their project supervisor(s) here)**

Oliver Tutt, Student, 3<sup>rd</sup> Year Electronic Engineering (BEng)  
Joshua Danks-Smith, 4<sup>th</sup> Year Electronic Engineering Student (MEng)  
Thomas Talbot, 4<sup>th</sup> Year Electronic Engineering Student (MEng)  
James Winters, 4<sup>th</sup> Year Electronic Engineering Student (MEng)  
Viswadeep Sarangi, PhD student  
Dr Adar Pelah, Supervisor

May 2015

**Location(s) of project**

Laboratory for Intelligent Virtual Environments (LIVE), Room S008, Physics-Electronics Building

**SECTION 2 FUNDERS**

**What is the funding source(s) for the project?**

Projects allowance from the Department of Electronic Engineering

**Please answer the following:**

- (i) Does the express and direct aim of the research or other activity raise ethical issues?  

YES	<input type="checkbox"/>
-----	--------------------------

NO	<input checked="" type="checkbox"/>
----	-------------------------------------
- (ii) Is there any obvious or inevitable adaptation of research findings to ethically questionable aims?  

YES	<input type="checkbox"/>
-----	--------------------------

NO	<input checked="" type="checkbox"/>
----	-------------------------------------
- (iii) Is the work being funded by organisations tainted by ethically questionable activities?  

YES	<input type="checkbox"/>
-----	--------------------------

NO	<input checked="" type="checkbox"/>
----	-------------------------------------
- (iv) Are there any restrictions on academic freedoms – notably, to adapt and withdraw from ongoing research, and to publish findings?  

YES	<input type="checkbox"/>
-----	--------------------------

NO	<input checked="" type="checkbox"/>
----	-------------------------------------

If you answered **Yes** to any of the above, please give details below:



**SECTION 3 DETAILS OF PROJECT OR OTHER ACTIVITY****Aims (100 words max)**

This application related to extending a treadmill-based motion capture suite that uses depth sensors with additional optical and wearable sensors. The research will involve subjects walking on the treadmill during which their gait is measured and the data stored.

**Background (250 words max)**

StroMoHab is a gait capture and analysis system developed in the Electronic Engineering Department, University of York, and currently in use in Dept. It was created in consultation with clinicians to introduce an inexpensive yet accurate way to analyse the movements of patients while also engaging them with virtual reality-based activities intended to help correct gait and alleviate other conditions, such complex regional pain syndrome (CRPS). The system has undergone review, risk assessment and received ethics approval in the NHS and is currently in use in research and clinical studies at Cambridge University Hospitals. The treadmill is fitted with handlebars, an automatic stop button and a body harness suspended from the ceiling. The system uses depth-based motion capture sensors (currently the marker-less MS Kinect™ V2) to track a user walking on a treadmill. The proposed extension involves the addition of other sensors, as described in this document.

7<sup>th</sup> May 2015

**Brief outline of project/activity (250 words max)**

The purpose of this application is to update the existing motion capture suite to allow the use and data analysis of additional methods of motion capture, including marker-based optical capture and wearables such as inertial measurement units (IMUs). These would require placement of reflective markers or small wireless sensors, attached to the walker's body using stickers or Velcro. Data collected from the walker specifies the x, y, z coordinates, and can include specific sensor-dependent data such as 3-DOF acceleration, EMG, gyroscopes, or magnetometry. Information collected from the subjects will not include their names or other identifiable features,

**Study design** (*if relevant – e.g. randomised control trial; laboratory-based*)

This application pertains to the extension of an existing motion capture suite rather than to a specific study design. Testing will involve participants walking or standing on a treadmill whose movement is recorded using the sensing methods described above.

**If the study involves participants, how many will be recruited?** About 25 participants will be recruited for initial testing in final year MEng and BEng student projects.

**If applicable, what is the statistical power of the study, i.e. what is the justification for the number of participants needed?** Standard parametric statistics extracted from the mocap software will sufficiently assess accuracy using the above numbers of subjects. Other measures may include qualitative responses by subjects to questions.

**SECTION 4 RECRUITMENT OF PARTICIPANTS**

**How will the participants be recruited?**

Participants will be recruited through personal contacts or via email, texts or social media.

**What are the inclusion/exclusion criteria?**

Participants with any impairment or other self-reported condition that could affect their safety while walking on a treadmill would be excluded from participation.

**Will participants be paid reimbursement of expenses?**

YES

NO

**Will participants be paid?**

YES

NO

*If yes, please obtain signed agreement*

**Will any of the participants be students?**

YES

NO

7<sup>th</sup> May 2015

**SECTION 5 DATA STORAGE AND TRANSMISSION**

**If the research will involve storing personal data, including sensitive data, on any of the following please indicate so and provide further details (answers only required if *personal* data is to be stored).**

<b>Manual files</b>	
<b>University computers</b>	Alongside the described sensor data, anonymised individual information may be recorded, such as age in years, height and weight.
<b>Home or other personal computers</b>	
<b>Laptop computers, tablets</b>	
<b>Website</b>	

**Please explain the measures in place to ensure data confidentiality, including whether encryption or other methods of anonymisation will be used.**

Collected data is stored according to established rules of compliance. Confidentiality is ensured by the collection of limited datasets that is sufficient for purposes of the study and anonymised such that participants cannot be identified from it.

**Please detail who will have access to the data generated by the study.**

Only the listed investigators will have access to the data

**Please detail who will have control of and act as custodian for, data generated by the study.**

Dr Adar Pelah

**Please explain where, and by whom, data will be analysed.**

Collected data will be analysed and used for establishing the accuracy of the extended sensing in the laboratory and for completion of final year project student reports.

**Please give details of data storage arrangements, including where data will be stored, how long for, and in what form.**

Data is recorded locally as anonymised datasets then sent to a secure AWS EC2 instance

7<sup>th</sup> May 2015

which in turn encrypts and secures data within the running instance. The AWS server is accessible only by Two Factor Authentication with permissions granted only by Dr. Adar Pelah. For development purposes, anonymised data sets can be accessed locally from the laboratory computers. The access is only to the consented data and is secured by network privileges from data controllers. This data is password protected at multiple levels and access requires authentication from the custodian of the data.

7<sup>th</sup> May 2015

**SECTION 6 CONSENT**

**Is written consent to be obtained?**

YES

NO

*If yes, please attach a copy of the information for participants*

*If no, please justify*

**Will any of the participants be from one of the following vulnerable groups?**

Children under 18	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
People with learning difficulties	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
People who are unconscious or severely ill	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
People with mental illness	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
NHS patients	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
Other vulnerable groups (if 'yes', please give details)	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>

**If so, what special arrangements have been made for getting consent?**

N/A

**SECTION 7 DETAILS OF INTERVENTIONS**

**Indicate whether the study involves procedures which:**

Involve taking bodily samples	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
Are physically invasive	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>
Are designed to be challenging/disturbing (physically or psychologically)	YES	<input type="checkbox"/>	NO	<input checked="" type="checkbox"/>

**o, please list those procedures to which participants will be exposed:**

N/A

**List any potential hazards:**

User falls while walking on treadmill

**List any discomfort or distress:**

N/A

**What steps will be taken to safeguard**

- (i) the confidentiality of information

Confidentially is protected by the measures taken to anonymise the data and the unidentifiable information collected for each participant. As well as these safeguards, as stated above, the data is held in secured AWS servers that meet GDPR compliance regulations and access to this is protected by Two Factor Authentication under controlled access to AWS servers only by Dr Adar Pelah. This data is password protected at multiple levels.

- (ii) the specimens themselves?

N/A

**What particular ethical problems or considerations are raised by the proposed study?**

Potential injury from falls while walking on a treadmill. These are mitigated by standard safety features as used for treadmill walking, including handlebars, a body harness and automatic treadmill stop buttons.

**What do you anticipate will be the output from the study? *Tick those that apply:***

Peer-reviewed publications	<input checked="" type="checkbox"/>
Non-peer-reviewed publications	<input checked="" type="checkbox"/>
Reports for sponsor	<input type="checkbox"/>
Confidential reports	<input type="checkbox"/>
Presentation at meetings	<input checked="" type="checkbox"/>
Press releases	<input type="checkbox"/>
Student project	<input checked="" type="checkbox"/>

**Is there a secrecy clause to the research?**

YES

NO

*If yes, please give details below*

**SECTION 8 SIGNATURES**

The information in this form is accurate to best of my knowledge and belief and I take full responsibility for it.

I agree to advise of any adverse or unexpected events that may occur during this project, to seek approval for any significant protocol amendments and to provide interim and final reports. I also agree to advise the Ethics Committee if the study is withdrawn or not completed.

Signature of Investigator(s): .....

.....

Date: .....

***Responsibilities of the Principal Researcher following approval***

- If changes to procedures are proposed, please notify the Ethics Committee
- Report promptly any adverse events involving risk to participants



## ADULT CONSENT FORM

---

**1) Title of Research Study:** Interactions Between Vision and Locomotion in Simulated Environments

**2) Investigators:** Dr Adar Pelah, Dr Howard Hock, Dr Elan Barenholtz

**3) Purpose:** The purpose of this study is to further our understanding of the relationship between vision and locomotion (walking). It is anticipated that the results of this research will provide insights that could lead to new treatments of impairments in mobility due to stroke and other brain injuries.

**4) Procedures:** As a participant you will be asked to stand or walk on a treadmill while viewing an image on a large screen in front of you. Sometimes you may see motion images on the screen, for example of a figure that appears to copy your movements as you make them. You will be asked to walk or stand for periods of between 5-10 minutes at a time. Each session will last about 1 hour and will include not more than 30min of walking. You will be offered the opportunity to return for up to 5 additional sessions, but you will be under no obligation to do so. The treadmill belt will be set at a slight incline of 0-10% and at a maximum speed of 3 mph for patients or 6 mph for others. You will be able to stop and rest, or sit down, at any time. You will also be able to stop the session and no longer participate in the experiment at any time that you wish. You may be asked questions by your experimenter about what you are seeing on the large screen or about your walking, for example whether you experience changes in your speed of walking. And you may also be asked to press a button to indicate your answer to questions that your experimenter will ask you. At the end of the session, and before you go home, we recommend that you sit and rest for a few minutes and until you feel comfortable.

**5) Risks:** There are some risks that you should be aware of if you choose to participate in this experiment. As you will be walking on a treadmill, there is a small chance that you may lose your footing and fall. Because of this, we will ask you to wear a harness that is anchored securely to the ceiling. That way, should you happen to fall the harness will hold you securely and you will not hurt yourself. You may also experience slight eye strain or similar visual discomfort, but if so this will not last for a long time. Remember, that if you feel any kind of discomfort you can rest at any time. You may also stop the experiment at any time, and we will not be upset or penalize you for this in any way. As with other forms of physical exertion, there is also a risk of having a heart attack or similar cardiovascular event. In case of such an emergency, at least one member of the experimental staff with CPR training will be available during the experiments. The standard operating procedures, including information on contacting EMT in case of emergency, are posted in the laboratory.

Important notice regarding insurance coverage in case of research injury: *"If you are injured or get sick as a result of your participation in this study, you should obtain medical treatment and notify the FAU researchers. Payment for this medical treatment is not routinely available from the researchers or FAU. You, or any available health insurance you have, will be billed for this treatment. Your health insurance company may not pay for treatment of injuries as a result of your participation in this study. Also, no funds are available to pay for any wages you may lose if you are harmed by this study. You do not give up your legal rights by signing this consent form."*

**6) Benefits:**

As well as possibly contributing to medical research, the potential benefits to you will be to become familiar with new research techniques, and in particular, with the novel use of computers and specially designed visual displays. The main benefits of your participation is to society to contribute ultimately to better treatments for people who have had a brain injury and similar conditions.

**7) Data Collection & Storage:**

Any information collected about you will be kept confidential and secure and only the people working with the study will see your data, unless required by law. The data will be kept securely for at most 5 years in a locked cabinet or password-protected computer in the principal investigator's office. After 5 years, paper copies will be destroyed by shredding and electronic data will be deleted. We may publish what we learn from this study. If we do, we will not let anyone know your name/identity unless you give us permission.

**8) Contact Information:**

- If you have questions about the study, you should call or email the principal investigator(s), Dr Adar Pelah at (561) 827-9171 and [apelah@fau.edu](mailto:apelah@fau.edu), Dr Howard Hock at (561) 843-1930 and [hockhs@fau.edu](mailto:hockhs@fau.edu), and Dr Elan Barenholtz at (561) 297-3343 and [elan.barenholtz@fau.edu](mailto:elan.barenholtz@fau.edu).
- If you have questions or concerns about your rights as a research participant, contact the Florida Atlantic University Division of Research at (561) 297-0777 or send an email to [fau.research@fau.edu](mailto:fau.research@fau.edu).

Initials \_\_\_\_\_

**9) Consent Statement:**

I have read or had read to me the preceding information describing this study. All my questions have been answered to my satisfaction. I am 18 years of age or older and freely consent to participate. I understand that I am free to withdraw from the study at any time without penalty. I have received a copy of this consent form.

I agree \_\_\_\_\_ I do not agree \_\_\_\_\_ be audiotaped/videotaped

Signature of Participant: \_\_\_\_\_ Date: \_\_\_\_\_

Printed Name of Participant: First Name \_\_\_\_\_ Last Name \_\_\_\_\_


Signature of Investigator: \_\_\_\_\_ Date: \_\_\_\_\_

**1. Project Summary:** This study will use simulated environments to investigate interactions between emotion, cognition, and motor movement during locomotion. It is expected that emotion will interact with locomotion, that both of these will interact with cognition, and that differences between these interactions will be reflected in behavioral responses to the emotion-inducing stimuli, behavioral responses to the cognitive stimuli, locomotion parameters during exposure to both types of stimuli, and hemodynamic activity during exposure to both types of stimuli. There are several novel features in this research, some involving the novelty of each individual component (e.g., locomotion in a simulated environment, a probabilistic spatial learning task for probing cognition, functional near-infrared spectroscopy [fNIRS] for the measurement of hemodynamic activity) and some involving the combination of these techniques (e.g., the effect of emotion on locomotion and vice versa).

**2. Objectives:** The primary objective is to deepen our understanding of the effects of different types of emotional load, cognitive load, and their interaction on locomotion and vice versa. A secondary objective is to identify unique patterns of hemodynamic response associated with the different constellations of emotion, cognition, and locomotion interaction.

**3. Background & Rationale:** There is evidence that different forms of probabilistic learning can result in dependence on one (or both) of two independent memory systems in the brain: one, a “cognitive, spatial”, hippocampal system and the other a “habit, stimulus-bound”, dorsal-striatal system. Different versions of the probabilistic learning task used in this research were designed to differentially access these different systems. In addition, these two systems are differentially affected by the presence of emotional arousal. Different types of emotion induction used in this research were designed to differentially affect access to these different systems (Packard & Goodman, 2012). Much of this research was done with rats in which the rats spatially navigate their path through a maze, although some more recent work has been done with stationary human participants responding to a static environment (e.g., Wirz, Bogdanov, & Schwabe, 2018). The use of locomotion (via the use of a treadmill) in a simulated environment (via the use of a Kinect-based app) will allow for our research with human participants to more closely approximate natural spatial navigation through a three-dimensional environment using a probabilistic learning task that, at least in some of its versions, demands spatial learning. Locomotion itself can influence the manner in which we perceive our progress through an environment, as when people perceive themselves as accelerating through an environment after walking on a treadmill (e.g., Durgin, Pelah, Fox, Lewis, Kane, & Walley, 2005). Finally, other work implicates structures in human inferior frontal cortex in cognitive “model-based” forms of learning that are engaged by some versions of our probabilistic learning task (e.g., Lee, Shimojo, & O’Doherty, 2014). We will use fNIRS to determine whether distinct patterns of cortical activation are associated with the different versions of our probabilistic learning task and the interactions of that task with both emotion and locomotion.

Durgin, F. H., Pelah, A., Fox, L. F., Lewis, J., Kane, R., & Walley, K. A. (2005). Self-motion perception during locomotor recalibration: more than meets the eye. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3), 398.

	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81(3), 687-699.

Packard, M. G., & Goodman, J. (2012). Emotional arousal and multiple memory systems in the mammalian brain. *Frontiers in Behavioral Neuroscience*, 6, 14.


Wirz, L., Bogdanov, M., & Schwabe, L. (2018). Habits under stress: mechanistic insights across different types of learning. *Current opinion in behavioral sciences*, 20, 9-16.

#### 4. Research Plan:

**(a) Study design.** The study design is a 3 x 3 x 2 design where probabilistic learning scenario (i.e., participants will either be in the color-driven, corner-driven, or location-driven version of the task; see below), emotional arousal (i.e., participants will either be in a positive, neutral, or negative emotion arousal condition; see below), and gender (male vs. female) are between-subject's variables. Data collection will begin spring of 2020 and continue through fall of 2020 and spring of 2021 if necessary in order to test approximately 12 participants in each cell of the design (3x3x2=18; 18x12=216; additional participants may be needed to replace participants with outlying performances).

**(b) Research Methods.** The experimental setup includes a treadmill, a movement recording sensor, a visual display, and an fNIRS machine along with its optical fibers and headgear. The display will provide visual information to the participant as they either walk or stand on the treadmill. The movement recording sensor will non-invasively measure a participant's walking parameters, such as their step length or the symmetry of their body movements. The participant is not required to wear or attach sensors or markers in order for body movement measurements to take place. A participant's perception of the visual stimulus will be recorded from either their verbal responses (to the emotion stimuli; see below) or by walking onto the tiles (in the probabilistic learning task; see below). Cortical activation during both phase three (see below) and phase four (see below) will be assessed using functional near-infrared spectroscopy (fNIRS), which is a non-invasive, safe optical imaging technique commonly used to measure changes in blood flow in different areas of the brain. These blood flow changes can in turn be used to draw inferences about the association of activity in these brain areas with the mental processes in which participants were engaged during the experimental tasks. In fNIRS, near-infrared light at 690 nm (which is sensitive to deoxy-hemoglobin or HbR) and 830 nm (which is sensitive to oxy-hemoglobin or HbO) is delivered to the scalp via light-emitting diodes called sources. The amount of light refracted is measured via light-detecting optodes called detectors. A change in the amount of light refracted (i.e., the relative change in HbO and HbR) indicates changes in blood volume and blood oxygenation and is used as an indicator of cortical activation. The sources and detectors are embedded in a soft, comfortable headgear that is placed on the head. The configuration of the sources and detectors in the headgear will allow measurement of hemodynamic responses in the orbital and lateral prefrontal cortex in both hemispheres.

In the first phase of these experiments, participants will complete both the state version and the trait version of the State-Trait Anxiety Inventory for Adults (STAI; Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A., 1983). Both versions have 20 questions. A short

 FAU Institutional Review Board	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable

version of the state form (i.e., six questions) will be used after the emotion induction (see below).


In the second phase of these experiments, the headgear will be placed on the participant's head. The participant will then acclimate to the treadmill while wearing the headgear (e.g., Herold, Wiegel, Scholkmann, & Muller, 2018).

In the third phase of these experiments, an emotion induction will take place. Pictures to induce positive (e.g., happy), negative (e.g., anger, fear, sadness, etc.), and/or neutral emotions will be displayed. The pictures are part of the International Affective Picture System (IAPS). The IAPS is a set of normed pictures that have been assessed for their emotional valence (i.e., positive, negative, or neutral) and emotional arousal across different age groups and genders (Lang, Bradley, & Cuthbert, 1999). These pictures have been used in dozens of studies conducted across the world. In the third phase of these particular experiments, participants will be exposed to 11 triplets of pictures, all of similar emotional valence, and will be asked to rate the quality of their emotional arousal (ranging from happy to sad) and the quantity of their emotional arousal (ranging from bored to tense/excited) after each triplet.

In the fourth phase, participants will be exposed to one of three probabilistic learning scenarios. In all three of the scenarios, each trial consists of two or three different color tiles located on a brown square on the floor in the middle of a virtual room. The tiles will be presented in two or three of the four locations on the brown square. Each wall of the room consists of different geometric figures (e.g., stars on one wall, triangles on another wall, etc.). As the participant walks on the treadmill, their avatar walks into the room from a particular direction. They select a tile by walking onto it. On each trial, only one tile is the winning tile. The participant receives positive feedback upon selection of the winning tile, but negative feedback upon selection of a losing tile. Then on the next trial, the participant enters the room from a different direction, different tiles appear in different locations on the brown square, and the participant chooses again. In the color-driven version of the task, the blue tile (for example) is the winning card 70% of the time, regardless of its relation to a particular corner of the room or its particular location on the brown square. In the corner-driven version, the star-diamond corner (for example) has the winning tile 70% of the time, regardless of its particular location on the brown square or its color. Finally, in the location-driven version, the upper left location (for example) has the winning card 70% of the time, regardless of its color or its relation to a particular corner of the room. In probabilistic learning tasks of this nature, participants improve their performance (i.e., how often they pick the winning card) over the course of several trials, often times without being able to describe why they are getting better at the task. Participants will begin with 75 points (displayed on a gauge on the display) and will gain 5 additional points if they select the winning tile, but will lose 5 points if they select a losing tile.

In the fifth phase of these experiments, participants will be given a learning awareness questionnaire. This questionnaire is internally generated in my lab. It assesses the degree to which participants understood the nature of the board game task, their use of different strategies in the board game, and the degree to which they were aware of their strategy for picking a winning card.

**(c) Study Population.** Participants will be part of the Psychology Subject Pool. Approximately 250 participants will be tested, yielding approximately 12 participants per cell in the 3x2x2

 FAU Institutional Review Board	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable



between-subjects design, which is a standard sample size in cognitive research with between-subjects designs. There will be no criteria for inclusion and/or exclusion from study participation.

**(d) Recruitment Plan.** Participants will be recruited through the Psychology Subject Pool. Participants will be tested in BS-12, Room 518, on the FAU, Boca Raton campus.

**(e) Analysis Plan.** The behavioral results (i.e., participants' verbal ratings of their emotional response to the pictures; degree of learning in the probabilistic learning task) and the hemodynamic results will be analyzed using factorial ANOVA. Additionally, responses to the State and Trait anxiety questionnaires and the learning awareness questionnaire will be correlated with performance in the emotion and learning tasks, as well as with the hemodynamic response. The results will be disseminated through publication in a peer-reviewed journal.


**(f) Compensation to Participants.** Participants get first-hand experience with psychological research, have opportunities to ask questions of the experimenters (who are typically undergraduate psychology majors), and receive a debriefing that builds on their classroom knowledge of psychology. Other than that, the only "compensation" participants will receive will be credit toward the research requirement component of their General Psychology course.

**5. Benefits:** All of us experience cognitive and emotional loads while performing motor tasks. This research explores the nature of those interactions and potential consequences of those interactions for our cognitive, emotional, and motor functioning. Understanding the interactions of cognitive, emotion, and motor processes could have wide-ranging consequences not only for normal functioning, but possibly for functioning in aging populations, for functioning in stroke patients, and/or for functioning in populations with disorders that have both a physical and mental manifestation (e.g., Parkinson's disease).

**6. Risks:** The effects of emotion on learning has long been a topic of interest to researchers and one effective method of investigating this topic has been to experimentally induce (i.e., manipulate) emotional state during the experiment, rather than simply measuring the affective state of the participant upon arrival, hoping for the emotional state of interest to be present. The experimental approach not only provides control of the emotional state uncorrelated with other life circumstances but is also much more efficient, saving many work hours that would be otherwise spent testing participants that would not be relevant to the central hypotheses.

With this being said, participants in the negative, emotionally arousing condition may experience brief discomfort commensurate with the emotional state that is being induced. The experimenter will observe the participant during this time. If a participant comments/complains about this discomfort, the experimenter will respond in the following manner: "I know these pictures are difficult to view. Please do the best you can." If a participant verbally indicates that they wish to discontinue viewing the pictures, the experimenter will allow the participant to turn away from the screen during the IAPS picture presentation, but then continue the rest of the experiment as normal. Of course, if the participant wishes to discontinue the experiment, they can do so at any time and still receive their research requirement credit.

With regard to breach of confidentiality or invasion of privacy, participants will sign and date a consent form. These signed consent forms will be kept in a locked file cabinet in BS-12, Room 518. Other than the consent form, all participants are identified in the data files and on the questionnaires they complete by a code that is relevant only to the condition of the experiment

	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable

they are in and has no connection to any of the participant's identifying information. Thus, the data is anonymized. In addition, only group results will be disseminated.


**7. Informed Consent Process:** After greeting the participant, the experimenter will give a brief and general oral description of what the participant will be expected to do, how long the experiment will last, and the number of research requirement credits the participant will earn. Then the participant will be given a consent form, be instructed to read the consent form as carefully as they wish, be given an opportunity to ask any questions, and finally, if the participant consents to participate, be asked to sign the consent form.

At the conclusion of the experiment, the participant will be debriefed as follows:

“Thank you for participating. One goal of this experiment was to see whether you got better at picking the winning tile as the room game went on. The winning strategy could have been picking a particular tile color, picking a tile near a particular intersection of two walls, or picking a tile on a particular location on the board. We think that those different types of strategies involve different types of learning, like learning the habit of picking the winning blue tile vs. thinking carefully about which intersection of two walls is the winning tile nearest to. In addition, you rated some pictures displayed on the computer monitor at the beginning of the experiment. Some participants see emotionally positive pictures, some see regular, everyday pictures, while other participants see disturbing pictures of violence and disease. We think that the emotional arousal that occurs when participants see disturbing pictures can influence how well a person picks up on the winning strategy in the room game, depending upon which strategy – card color, wall intersection, or location – is in play. Finally, another goal of this experiment was to study locomotor activity and brain activity during the different kinds of tasks you were asked to do. For example, we think that both locomotor activity and brain activity might differ, depending on which emotion condition you were in and which learning strategy was the winning strategy. Discovering the ways in which different forms of learning are affected by emotion and the effects that these kinds of interactions have on both brain activity and locomotor activity may help us to develop ways of assessing peoples' capacity for operating efficiently when experiencing different emotional states – as when first responders face a stressful situation – or may help us to develop methods for assessing the mental, emotional, and physical impairments in people suffering from medical disorders, such as stroke.

**8. Informed consent document:** See attached. Signed consent forms will be kept in a locked file cabinet on the FAU Boca Raton campus in BS-12, Room 518. They will be destroyed by shredding after 5 years.

**9. Research Materials, Records, and Privacy:** Four paper and pencil questionnaires will be collected from each participant. These will be kept in a locked file cabinet on the FAU Boca Raton campus in BS-12, Room 518. Participants' responses on these questionnaires will be transcribed into electronic files. The verbal responses made by participants to the emotion-inducing stimuli will be recorded by an experimenter into an electronic file as they are being produced by the participant. Participants' selections in the room game will be collected electronically. All electronic data will be kept on a password-protected computer in BS-12, Room 518. This room is locked except when being used for the collection of data with human

	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable

participants. No demographic information will be collected from the participants except for sex, handedness, and age. Experimenters will be instructed to keep the names of individual participants confidential. Even if the identity of an individual participant was to become known, it would be very difficult to connect that individual to their data. In addition, there is nothing about participation in this experiment that, in and of itself, would reveal anything about the participant beyond what would be known in the course of a normal day. The data will be reported as group data in a peer-reviewed journal. The paper and pencil questionnaires will be destroyed by shredding after 5 years.

**10. Resources:** The research will be conducted on the FAU Boca Raton campus in BS-12, Room 518. The room is approximately 12'x15'. One participant at-a-time will be tested. Participants will be tested by undergraduate students enrolled in directed independent research (DIR). These students will complete the Social and Behavioral Investigators course at the CITI website. These students will be individually trained in the standardized procedure that we will use to this study and, in addition, they will have a detailed script on how to interact with their participants. There will no need for any additional services beyond those in normal everyday life. No other sites or resources are needed for the conduct of this study.

FAU Institutional Review Board	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable



ADULT CONSENT FORM

Version 1.0 -- January 14, 2020

1) **Title of Research Study:** Locomotion, emotion, and cognition in simulated environments

2) **Investigator:** Dr. Terrence M. Barnhardt; Dr. Adar Pelah

3) **Purpose:** The purpose of the experiment is to explore the interaction of locomotion, emotion, and learning.

4) **Procedures:**

- You will be tested individually with two experimenters present. In one task, you'll walk on a treadmill while viewing pictures, which may be emotional in nature, and you'll rate your emotional reaction to those pictures by giving an oral response. In another task, you'll walk on a treadmill while navigating your avatar to a destination in a virtual room. The treadmill will be set at speeds ranging between 0 and 3.6 miles/hour, a typical casual walking speed. Each task will last between 5 and 10 minutes. Prior to the tasks, a headband that contains optic fibers will be placed on your head. During the tasks, the amount of light refracted from your head will be used as an indicator of changes in blood flow in the brain. This technique is safe and frequently used with infants, children, and adults.
- Your participation will begin and end today and will last no longer than 90 minutes.
- This research will be conducted in BS-12, Room 518, typically during normal school hours.
- By participating in this experiment, you will earn 1.5 credits towards satisfying the research participation requirement of Psychology 1012, General Psychology.
- If you do not wish to participate in any experiments, you can write research reports as specified in the research requirement handout associated with your General Psychology course.

5) **Risks:** While walking on a treadmill, there is a small chance that you may lose your footing and fall. Because of this, you will be provided with an automatic-stop pull-cord and magnet tethered between you and the treadmill. You can use this feature to immediately halt the treadmill if you feel it is necessary. If walking is too fatiguing, you may request rest between tasks. In addition, there is some risk for experiencing some negative, emotional arousal in response to some of the pictures.

6) **Benefits:** It is hoped that participation in this experiment will increase your familiarity with the nature of psychological research and increase your general knowledge about psychology. You can ask questions at any time, although the experimenter may need to postpone a definitive answer until the end of the experiment. At the end of the experiment, the experimenter will fully explain the nature of the research and will answer any questions you may have.


7) **Data Collection & Storage:** Any information collected about you will be kept confidential and secure. Only people working with or overseeing the study will see your data, unless required by law. The paper-pencil questionnaires will be kept in a locked file cabinet and the data generated by your oral responses to the pictures or your choices in the game will be stored on a password protected computer. All data will be stored in on the FAU Boca Raton campus in BS-12, Room 518, and, after 5 years, the paper-pencil questionnaires will be shredded and the computer files will be deleted. Sometimes researchers need to share information that may identify you and your research records with people that work for the University, the Institutional Review Board (IRB), Research Integrity staff, regulators or the study sponsor. These people are responsible for making sure the research is done safely and properly. If this does happen, we will take precautions to protect the information you have provided. We may publish what we learn from this study. If we do, we will not let anyone know your name/identity unless you give us permission.

8) **Contact Information:** If you have questions about the study, you should call or email the investigator Terrence M. Barnhardt, PhD at (561)297-0650 or tbarnhardt@fau.edu. If you have questions or concerns about your rights as a research participant, contact the Florida Atlantic University Division of Research, Research Integrity Office at (561) 297-1383 or send an email to researchintegrity@fau.edu.

9) **Consent Statement:** I have read or had read to me the information describing this study. All my questions have been answered to my satisfaction. I am 18 years of age or older and freely consent to participate. I understand that I am free to withdraw from the study at any time without penalty. I have received a copy of this consent form.

Printed Name of Participant: \_\_\_\_\_

Signature of Participant: \_\_\_\_\_ Date: \_\_\_\_\_

	1550546-1	
	Approved On:	February 3, 2020
	Expires On:	Not Applicable