The Classification of Multiple Interacting Gait Abnormalities Using Insole Sensors and Machine Learning

1st Alexander Turner
Computer Science
University of Nottingham
Nottingham, UK

2nd David Scott
Department of Sport, Health and Exercise Science
University of Hull
Hull, UK

3rd Stephen Hayes
Engineering
Nottingham Trent University
Nottingham, UK

Abstract—In this work we investigate the effectiveness of a wireless in-shoe pressure sensing system used in combination with a type of machine learning referred to as long term short term memory networks (LSTMs) to classify multiple interacting gait perturbations. Artificially induced gait perturbations consisted of restricted knee extension and altered under foot centre of pressure (COP). The primary aim was to assess the capacity to diagnose gait abnormalities without the need to attend a gait laboratory or visit a clinical healthcare professional, through the use of technology. Ultimately, such a system could be used to autonomously generate therapeutic guidance and provide healthcare professionals with accurate up to date information about a patients gait. The results show that LSTMs are capable of classifying complex interacting gait perturbations using in-shoe pressure data. When testing, 11 of 12 perturbation conditions were correctly classified overall and 58.8% of all data instances were correctly classified (8.3% is random classification). This work illustrates that an automated low cost, non-invasive gait diagnosis system with minimal sensors can be used to identify interacting gait abnormalities in individuals and has further potential to be used in a healthcare setting.

Index Terms—Gait abnormalities, insole technology, long term short term memory networks, high performance computing

I. INTRODUCTION

Gait is the coordinated movement of an individuals limbs enabling the forward propulsion of the center of mass (COM) [1]. Whilst this is a simple routine activity for most able-bodied individuals, various gait abnormalities affect numerous people for multiple reasons. Such abnormalities are commonly associated with neurodegenerative diseases, but are also synonymous with brain injuries and/or physical disabilities [2], [3]. Impairment of gait can have negative consequences on an individuals quality of life through limited mobility and independence [3]. Clinical gait analysis is used to detect movement abnormalities for the purpose of diagnosis and prescribing treatment for gait disorders. However, due to cost and time, its use in every day clinical practice is limited [2]. Clinical specialists typically rely on qualitative diagnostic methods, such as visual observations or verbal descriptions, to identify abnormal gait [4]. This is largely subjective based upon the individual clinicians experience and knowledge;

therefore, more accessible quantitative methods of analysing gait patterns could provide more objective measures [5].

Clinical diagnosis through identification of gait abnormalities is challenging due to the number of different symptoms which may be associated with multiple disorders. Various methods and classification systems have been proposed to facilitate movement disorder diagnosis and particular systems are often associated with specific movement disorders [6], [7]. Three-dimensional motion capture systems, with integrated force plates and electromyography (the gold standard) are very expensive and require large areas of space for operation. Furthermore, data capture and processing are time consuming, invasive and the acquired data must be interpreted by a trained professional. Perhaps even more costly to the patient is that this type of analysis only provides a snap-shot of an individuals gait, which may lead to the continuous adaptation of compensatory strategies to be overlooked as they may never be observed within the test time frame.

Various classification systems exist which again typically focus on gait abnormalities in specific disorders. Examples include the Gait Profile Score (GPS), the Gross Motor Function Classification System (GMFCS) and the Edinburgh Visual Gait Score (EVGS) [4], [8], [9]. A strong correlation exists between the GPS and the EVGS, furthermore, significant differences have been identified for the GPS at the different levels of the GMFCS demonstrating its efficacy [7]. However, [9] suggest that the use of qualitative systems such as the GMFCS are more prone to bias due to the subjective nature of assessment.

Although visual observations and grading criteria are an appropriate way of diagnosing and treating disorders, the identification of minor gait deviations resulting in random error are not accounted for when treating the individual [10]. For instance, less pronounced movement abnormalities in an individual are more challenging to identify and classify than larger deviations from normal gait. Additionally, certain gait abnormalities may not be present at the time of diagnosis, but can reappear during the patients normal routines. Previous work demonstrated a proof of concept that non-invasive in-shoe data capture systems along with high performance computing and machine learning have the capacity to provide qualitative

classifications of movement alterations during gait [11]. The efficacy of deep learning architectures has been shown to be proficient in sequence identification and pattern recognition in image classification, signal processing and feature extraction [12], [13]. Long term short term memory networks (LSTMs) are a machine learning architecture which have the ability to learn, both long and short term, large complex data sets from varying time periods can be classified. It is therefore possible for the LSTM to create its own interpretation of gait function using only the raw unprocessed data. The use of LSTMs for this work is based on a direct comparison to convolutional neural networks in the aforementioned previous work, which demonstrated the superiority of the LSTMs, as they were significantly more effective at classifying the raw under-foot pressure data [11].

The ultimate goal of this study was to combine deep learning techniques and a robust, easy to use and affordable in-shoe sensor technology to create an accurate gait analysis tool, which can be used outside of a clinical environment. In order to progress beyond a proof of concept, a more user friendly and cost effective in-shoe measurement system is needed. The F-scan (Tekscan, Boston, USA), which was used in our previous study [11] although lightweight, requires the use of ankle cuffs to connect the insoles to the telemetric device, a belt to hold the battery and telemetry components, and cables running from the ankle cuffs to the belt. Furthermore, the in-shoe sensors are fragile and can be easily damaged. All of which makes the use of such a system challenging for the general public. The concept of the Smart (Moticon GmbH, Munich, Germany) insole is that each insole can simply be slipped into the users shoe and linked to an Android phone. The greatest challenge therefore lies in the difference of the amount of data available from the two systems. Each insole of the F-scan system can provide up to 1260 channels, generating 252,000 data samples per second across both feet, whereas the Moticon system consist of only 16 pressure sensors per foot and 3200 data samples per second across both feet. However in addition, each insole contains an x,y,z accelerometer producing an extra 3 channels of data.

The primary aim of this research was therefore to explore the capacity of combining wearable technology with limited sensors and a deep learning architecture to identify gait abnormalities. The first objective was to examine whether LSTMs can be used to classify gait alterations with significantly less data and relatively cheaper devices to allow accessibility to the wider population. A secondary objective was to understand if this technology could be used to detect interaction gait abnormalities, rather than just focusing on a single type. Additionally, we wanted to provide further evidence that non-invasive wearable technology combined with deep learning techniques can accurately and reliably identify gait alterations outside of a clinical setting. This research aimed to establish the capacity for underfoot pressure sensors and LSTMs to identify artificial gait perturbations derived from altered kinematics as well as under shoe perturbation conditions.

II. Related Work

There has been a significant amount of work done in the application of wearable sensors to provide more effective treatment and diagnostics in healthcare setting [14], [15]. However, within the field of analysing gait, the two areas that receive the most attention are that of Parkinsons' disease [16], [17], [18] and cerebral palsy [19], [20], [7].

A systematic review was conducted in [2] which surveyed 32 papers relating to analysis of movement using wearable sensors. 29 of these were applied to the analysis of gait. Until recently, machine learning is infrequently used in the analysis of gait compare to more traditional measures [19], [20], [7], [2]. However, more recently the use of machine learning to analysie gait in combinations of settings and with various modalities has become increasingly popular [21], [25], [26], [22], [23], [24]. While much progress has been observed, especially over the recent years, on predicting assessment aspects, there is little effort to assist healthcare professionals [26]. Clinical gait analysis for disease diagnosis and treatment planning without a trained professional present remains rare.

In 2018, a subset of the research team involved with this project released a paper where the Tekscan system was used to analyse minor gait perturbations [11]. There are two significant areas that this work builds upon. The first is the difference in data acquisition, where we use the Motion wireless insole instead of the tekscan system. The Moticon system is superior to the Tekscan system primarily in terms of outright cost. Additionally, the running costs are less as it doesn't require replacement of the insoles over a short period of time or a PC to download the data. The interactions with the Moticon system can be achieved using an Android enabled phone. However, the Moticon system produces approximately 1% of the data when compared to the Tekscan system which reduces the possibilities in terms of machine learning. The second difference in this work is that we use this data to detect different interacting gait perturbations simultaneously in both the feet and the knees, instead of just focusing on a single perturbation. This is to demonstrate the efficacy of such a system in a more real world environment, where it is uncommon for large sets of people to have a single, well defined non-interacting gait abnormality.

III. Long Terms Short Term Memory Networks for Sequence Classification

Long Terms Short Term Memory Networks (LSTMs) are a type of deep neural network [27], [28], [29] which is derived from the recurrent neural network [30] and is able to exhibit temporal dynamic behaviour . Deep neural networks have been responsible for many of the advancements in machine learning including object detection, sentiment detection in text and advanced signal processing [31], [32], [33], [30]. The LSTM is comprised of LSTM units, which are particularly well suited to analysing complex time series sequence data, and are significantly better at this than the recurrent neural network (RNN), the model in which they were based upon.
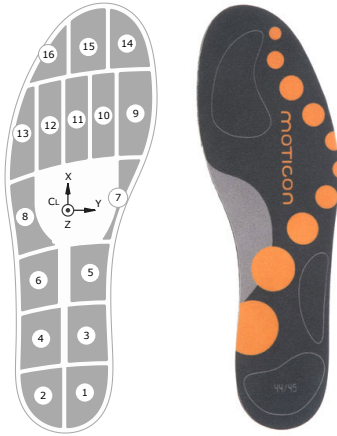
Fig. 1. The moticon insole which is used to generate the data to link the gait perturbations to their classification. Each insole has 16 pressure sensing pads and an x,y,z accelerometer. In total there are 22 readings from the insole, the three accelerometer readings and total force, center of pressure (x) center of pressure (y) alongside the pressure readings. Although variable, all of the sampling throughout this work will be done at 100hz.

The LSTM unit can be seen in Figure 2. The unit is made up of many different structures, where one of the key structures in this work is the forget gate. This allows the LSTM unit to build representations of data over long time scales whilst also being able to exclude any representations that prove not to be useful. This allows for a robust efficient method in acquiring and processing the most relevant features for the data [30]. For the type of data in this work, it has been previously shown that LSTMs are better suited to classification than convolution neural networks which are more typically used for visual systems and for tasks such as object detection [11].

A further benefit of LSTM's in this work is that they are capable of deriving their own representation of different classes of data. And rather than relying on a set of specified human generated rules to specify a classification, it can identify a range of criteria some of which might not be available to clinicians as it is not easy to visualise. Moreover, it provides a probability of it's correctness, which could prove useful in real world classifications.

IV. METHODS

Eight able-bodied participants (21-36 years, 60-95 kg, 6-11 UK shoes size) were recruited to complete 12 walking trials around a figure of 8 walkway (40 m in length) for 120 seconds per trial. The experimental conditions included 12 artificially induced gait perturbation conditions (PCs) (Table I). The artificially induced perturbations were generated using two methods. Firstly, square compressible rubberised pads were affixed to the sole of each shoe (3.5 cm^2 with a depth of 1.5 cm) located under 1) the lateral border of the forefoot, 2) the medial border of the forefoot and 3) the heel (Figure 3) . Additionally, a lockable knee brace on the left leg was used to restrict knee movement by limiting the amount of extension available to the individual. Two different levels of
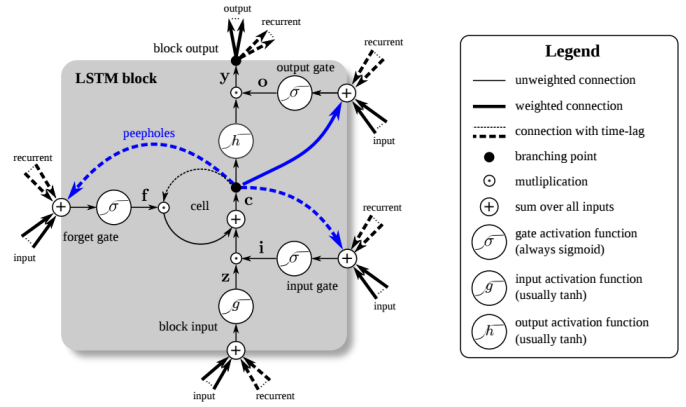

Fig. 2. An illustration of an LSTM unit which makes up the LSTM network derived from [30]. Note the forget gait which allows these units to disregard irrelevant information.

restriction were used, a minor (25 degree) and a major(45 degrees) restriction depicted in Figure 4. Upon arrival to the laboratory, participants provided their written informed consent prior to any testing. All participants wore standardized trainers provided by the researchers for all trials (size specific) to ensure consistency during data collection. Smart pressure sensing insoles (Moticon GmbH, Munich, Germany) were placed inside the trainers and were zeroed prior to the individual walking for each condition. This was achieved by asking the participant to sit down and lift their feet off of the floor, thereby removing all body-weight from the sensors. Body mass of the participants was not required for the processing of data in the current study as each participants data was normalised to values between 0 and 1 based on the minimum and maximum pressure values recorded.

Foot Pattern	Brace Setting	Classification
None	None	1
None	25°	2
None	45°	3
1	None	4
1	25°	5
1	45°	6
2	None	7
2	25°	8
2	45°	9
3	None	10
3	25°	11
3	45°	12

TABLE I
THE ARTIFICIAL GAIT ABNORMALITIES AND THEIR CLASSIFICATION. ALL OF THE FOOT STRIKE COMBINATIONS FROM FIGURE 3 ARE USED IN COMBINATION WITH THREE KNEE BRACE SETTINGS; FULL MOVEMENT, 25° AND 45° (FIGURE 4). EACH OF THESE COMBINATIONS OF SETTINGS IS GIVEN ITS OWN CLASSIFICATION NUMBER. THE PRIMARY ORDER IS BASED UPON THE FOOT PATTERNS AND THE SECONDARY ORDERS IS BASED UPON THE ANGLES OF THE KNEE BRACE. WHEN THE KNEE BRACE IS SET TO 'NONE' THE BRACE IS STILL WORN, BUT NO SETTING IS ENGAGED.

Two sets of Smart insoles (sizes 7 and 10), designed to function one UK shoe size above or below their defined
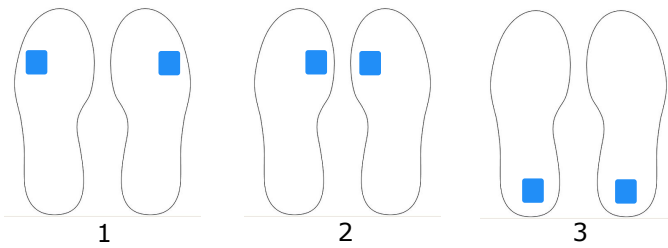
Fig. 3. Underfoot locations of the rubberised pads used to generate the foot strike pattern perturbations. Each rubberised pad was 3.5 cm(2) with a depth of 1.5 cm and deformed under load by approximately 1 cm. One of these perturbations was used at a time except when no perturbation was used.

size, were used for all data collection. The smart in shoe pressure sensors consist of an insole containing 16 pressure sensors per foot, a tri-axial accelerometer (x,y,z), a battery, and a wireless telemetry to an Android App. (Moticon Science version 03.06.02 (28)) (Figure 1). Data were recorded at 100 Hz over 19 channels per insole (16 pressure sensors and 3 accelerometer vectors) as well as a calculated total ground reaction force (vertical component) and calculated underfoot centre of pressure (COP) vectors (x = medio-lateral, y = anterio-posterior). This generated 1.15 million data vectors for all participants, over all perturbations with experiments running for 120 seconds. Each of which contained 42 variables from the two insoles. The vectors generated were combined into five second data instances consisting of input from both insoles; a graphical representation of one such instance can be seen in Figure 5. In order for an LSTM to be able to classify data it must first be trained. Data from this study were split into two categories, training data and testing data. Participants were randomly assigned to either the training or testing category; the data from six participants was used to train the LSTMs whilst the remaining two participants data where withheld for testing purposes. The testing process enabled the evaluation of each LSTM using unseen data from participants not involved in training the LSTM, thus providing information about the capacity of the LSTM to classify new data / participants that had not been used in the training process. This was to promote more real world statistical analysis, where the models are used on patients which were not used in the design of the models. Each instance of data was generated with seven one 100ths of a second gap between its predecessor and itself. This ensured a comprehensive sampling of the data without being prohibitively costly in the amount of time it took to train the LSTMs (Figure 6). The LSTMs in this work will contained 21 cells and were shaped according to table 2 and optimised using the Adaptive Moment Estimation optimiser with a mini batch size of 128 over 50 epochs. These parameters were found to be suitable through exploratory testing.

V. RESULTS AND ANALYSIS

The overall results of how well the LSTMs classified the perturbations outlined in table I can be seen in figures 7 and 8. Figure 7 presents data from the best performing LSTM in this work based on classification accuracy. The network presented
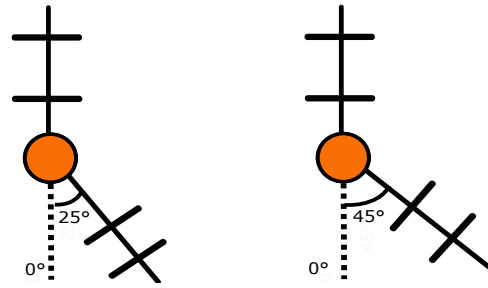

Fig. 4. The restrictions of extension of the knee designed to artistically perturb the gait of a healthy person. A restriction of 25° is considered small and had little effect on peoples gait. It is to be noted that whilst 25° of restriction via the use of a knee brace is significant, a large proportion of that restriction is absorbed through the tissue of the participant and results in a minor gait change. Each participant specified that this perturbation had little / if any restriction on their gait. This 45° of restriction via the use of a knee brace is more substantial. Each of the participants specified that this perturbation has a significant affect on their gait and was very restrictive of movement.

Layer	Type	Description
1	Sequence Input	Length of 42
2	Bi-LSTM Layer	21 Cells
3	Dropout Layer	50% dropout likelihood
4	Fully Connected Layer	12 Outputs
5	Softmax Layer	Convert the outputs to classification
6	Classification Layer	Output vector

TABLE II

THE CONFIGURATION OF THE LSTM USED IN THIS WORK. THE LSTM CONTAINS 39 CELLS AND PRODUCES AN OUTPUT VECTOR DESCRIBING THE PROBABILITY OF EACH OF THE 12 CLASSIFICATIONS.

in Figure 7 was trained over eleven epochs, resulting in an accuracy rate of 58.8%. This is for each individual time step of data, however, when grouped together as a sequence as seen in figures 7 and 8 - the classification for the entire sets of data for each condition is 83% and 92% respectively. The confusion matrix provides the overall accuracy as well as information on how the LSTM classified each PC. Based on the 12 different PCs, a random classification rate of each PC would be 8.3%. The diagonal running from the top left of the confusion matrix to the bottom right represents the number (and percentage) of correct classifications of each PC. Perturbation condition 11 was the most accurately classified with a correct classification over 97% of the time. In contrast to this PC 7 was the least accurately classified with correct classification occurring at a rate of only 15%. This LSTM correctly classified the exact condition more frequently than any other, for 10 out of the 12 PCs.

Figure 8 presents data from an LSTM trained over 23 epochs with an overall classification accuracy of 46.8%. Although the network presented in Figure 8 achieved an overall lower classification accuracy than that of Figure 7, the LSTM was able to correctly classify the right PC on aggregate over 11 of the 12 PCs.

Figure 9 presents performance data for the LSTMs over time based on training and testing. As the number of epochs increases up to approximately ten, both training and testing show a steady improvement with a testing peak reached at

Fig. 5. Graphical interpretation of the time separation of each instance of data used to train the LSTM networks. Each horizontal bar represents an instance of data consisting of 500 consecutive data points (depicted in Figure 4). A 7 data point gap was used to reduce the amount of data required to train the LSTM networks whilst maintaining comprehensive sampling. In terms of pressure, from 0 being the lowest and 1 the highest, the blue colors represent values closer to 0, and the yellow colors represent colors closer to 1. Each column consists of a single time step of data for both insoles, and the accelerator data can be seen by the two bands of data (consisting of the x,y,z accelerator values and the ground reaction force (vertical component) and calculated underfoot centre of pressure (COP). This data is represented by two bands at the top and the middle of the image. The rest of the image shows the pressure data from the insoles.

Fig. 6. A representation of how the data from each individual step is collated into an instance of data used to train the LSTM. Each horizontal bar can be considered 1 instance of data consisting of 500 time steps of data (information from both insoles). There is a 7 time step gap in between each instance of data (0 time step gap would be perfect overlap). The reason for this gap is that it gave a comprehensive sampling of the data, without being prohibitively costly to train in terms of time.

Predicted class	1	2	3	4	5	6	7	8	9	10	11	12		
1	2400	191			224	478	28	179	40	203	2		64.1%	35.9%
2		2922	87	1076	858	575	499	1315	137	103			38.6%	61.4%
3		138	2990		215	419		31	1315	29			58.2%	41.8%
4	946	38		1806	41	36	3		2				62.9%	37.1%
5		2		313	1492	444		2	7				66.0%	34.0%
6		17	14	53	195	1707							86.0%	14.0%
7		1		1			496	13					97.1%	2.9%
8		2	26		6	29	1629	1888	12				52.6%	47.4%
9		13	127			36	500	6	1561				69.6%	30.4%
10			43							1452	88		91.7%	8.3%
11					13					1644	3188	1750	48.3%	51.7%
12			36									1378	97.5%	2.5%
	71.7%	87.9%	90.0%	51.8%	45.4%	52.1%	15.0%	57.3%	48.2%	45.0%	97.3%	44.1%		
	28.3%	12.1%	10.0%	48.2%	54.6%	47.9%	85.0%	42.7%	51.8%	55.0%	2.7%	55.9%		
	1	2	3	4	5	6	7	8	9	10	11	12		

Actual Class

Fig. 7. Presentation of data from the best performing LSTM in this work based on classification accuracy. The network gave an overall accuracy rate of 58.8%. The confusion matrix however provides the overall accuracy as well as information on how the LSTM classified each PC. Based on the 12 different PCs, a random performance woukld be 8.3%. The diagonal running from the top left of the confusion matrix to the bottom right represents the number (and percentage) of correct classifications of each PC. Perturbation condition 11 was the most accurately classified with a correct classification over 97 % of the time. In contrast to this PC 7 was the least accurately classified with correct classification occurring at a rate of only 15 %. This LSTM correctly classified the exact condition more frequently than any other, for 10 out of the 12 PCs. As a general trend, it can be seen that the model is classifying a large proportion of the data accurately, which can be seen by the values within the blue diagonal.

11 epochs (the LSTM presented in Figure 7). At 15 epochs there was a substantial drop in training accuracy which did not appear to affect testing accuracy. However, the decrease in training accuracy is reversed by 20 epochs. The network presented in Figure 8 is represented by the blue square in Figure 9, although it is evident that after 25 epochs the training accuracy is maintained at near 100%, the testing accuracy begins to decline suggesting that training the LSTM beyond this point holds no benefit and becomes time costly.

Figure 11 highlights the activations of the biLSTM layer within the network and shows that the activations for the top half of the figure which correspond to the data from the left foot are much more varied and dynamic than those from the right foot. Broadly speaking this shows that the data from the left foot has more effect on the dynamics of the network and its behavior than that of the right foot. The difference is striking, suggesting that the data from the left foot is overwhelmingly responsible for the output of the network, which is the side the knee brace is attached. Although surprising that the data from the right foot was not used more by the network, the data in

the right foot was broadly a mimic of a subset (without the information from the knee brace as that was on the left leg) of data from the left foot, and was therefore not as valuable to the decision making process.

VI. DISCUSSION

The primary goal of this research was to determine if an affordable and easy to use gait analysis tool could be developed using in-shoe pressure sensors and deep learning

Predicted class	1	2	3	4	5	6	7	8	9	10	11	12		
1	1697	405			21	23	5			185	16		72.2%	27.8%
2	22	1223	83	1133	758	448	420	194	93	9			27.9%	72.1%
3		636	2258	27	755	647	202	1259	1058	53			32.7%	67.3%
4	1627	1020		1761	227	90							37.3%	62.7%
5				169	1419	657		5					63.1%	36.9%
6			605	393	104	1195		5					51.9%	48.1%
7						49	34		57				24.3%	75.7%
8		40	41	1		10	1017	1261	242				48.3%	51.7%
9			319	2	1	155	1627	581	1592				37.2%	62.8%
10		17				1				1508	456		76.1%	23.9%
11										1134	2756	1360	52.5%	47.5%
12										510	64	1768	75.5%	24.5%
	50.7%	36.8%	68.0%	50.5%	43.2%	36.5%	1.0%	38.3%	49.2%	46.7%	84.1%	56.5%		
	49.3%	63.2%	32.0%	49.5%	56.8%	63.5%	99.0%	61.7%	50.8%	53.3%	15.9%	43.5%		
Actual Class	1	2	3	4	5	6	7	8	9	10	11	12		

Fig. 8. Presentation of data from an LSTM trained over 23 epochs with an overall classification accuracy of 46.8%. Although the network presented achieved an overall lower classification accuracy than that of Figure 7, the LSTM was able to correctly classify the correct PC more often than any other with 11 of the 12 PCs correctly classified overall. The overall accuracy of this second network was lower based on a greater number of incorrect classifications being more evenly spread, resulting in a lower number of correct classifications overall, yet the highest classification values were attributed to the correct class for 11 of the 12 conditions.
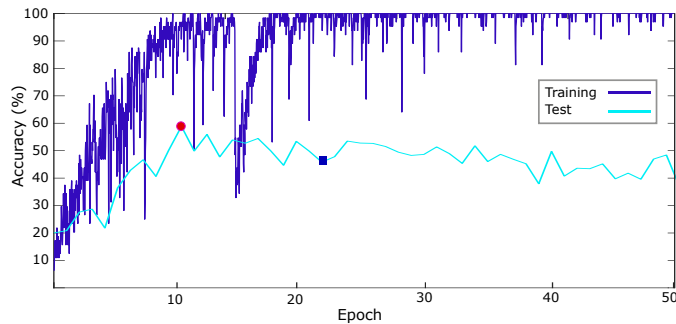
Fig. 9. Graphical representation of the performance of the LSTMs over time during training and testing. The red dot identifies the best performing LSTM (Figure 7) and the blue square denotes the LSTM which presented with the highest number of correct classifications (11 out of 12, Figure 8).

tools. The first objective was to explore the impact of fewer channels of data available to the LSTMs based on a relatively limited number of pressure sensors within the insoles, whilst increasing both the number and complexity of the artificially induced gait perturbations. The results show that a peak accuracy of 58.8% for individual time steps of data was achieved using the Moticon sensors (Figure 1) and a five second data instance (Figure 6) on unseen data (data not used for training purposes). This network correctly classified 10 of the 12 perturbations correctly providing an accuracy of 83% on aggregate.

The confusion matrices presented in this study show that the LSTMs used were able to form patterns pertaining to the classification of the PCs. Both Figures 7 and 8 clearly illustrate that some PCs were correctly classified more often than others (eg. PC 11 in both experiments) and that some PCs were misclassified more often as one or two other (similar) PCs such as PCs 1 and 12, whilst some were misclassified over a broader spectrum (eg. PC 6). Moreover, some PCs were never misclassified as others for example, PC 12 was never misclassified as anything other than PC 11, and PC 11 was never misclassified as anything other than PC 10 (Figure 7). Some of these patterns can be logically explained based on the characteristics of the conditions. Table 1 and Figure 1 presents those factors which dictate each PC. PCs 10 to 12 all consist of an under-heel rubber pad. It is therefore possible that the networks identified the foot strike pattern easily. This was evidenced by limited misclassifications of any of the under-heel PCs to each other and minimal numbers of misclassifications as PCs 3 and 4 (Figure 7). In Figure 8 only two other PCs (3 and 7) were misclassified as PC 10 and none were misclassified as PCs 11 or 12. The LSTMs were clearly capable of differentiating between the three underfoot perturbation conditions, irrespective of the knee angle restrictions. However, the two forefoot conditions were quite often misclassified as the no under-foot condition but very rarely as each other. The most obvious example of this is evident in Figure 7 where the medial border underfoot perturbation was only misclassified as a lateral foot PC a total of 14 times, whereas it was misclassified as one of the no underfoot conditions 3719 times out of a possible 9838 instances.

A secondary objective of this study was to understand to what extent underfoot pressure sensors, accelerometry data and LSTMs would be able to identify multiple interacting artificial gait perturbations derived from altered kinematics, as induced at the knee joint, as well as those generated by under shoe perturbations. Unlike the clear pattern seen relating to the under-shoe perturbations described above, it was much more challenging to identify a pattern within the LSTM output in the confusion matrices related to the kinematic perturbations. There are several plausible reasons for this: firstly, it is possible that the impact of the restricted knee extension did not present with substantially altered underfoot pressure for the LSTMs to easily detect. This may have placed a greater reliance on the accelerometry data. Secondly the method used to induce the restricted knee extension was a locking knee brace, which may have produced different kinematic responses from each individual based on the fitting and physical properties of each participant. Although the brace was secured as tightly as possible to each participant, any slack in the system would translate into different movement characteristics. Furthermore, the different volumes of adipose tissue and muscle mass within each participants leg may have altered the fitting. It is therefore likely that the angular position of the knee is less distinguishable by the networks. However, the present data suggest that with a greater number of training instances and participants this issue could be alleviated as the LSTMs did correctly identify the kinematic conditions 1) (freely movable) 45.9% of the time, 2) (25° restriction) 72.0% of the time and 3) (45° restriction) 58.6% of the time.

Figure 10 illustrates how the LSTM (represented in Figure 7) interpreted data to make a classification. For each PC, a matrix (44 x 50) of zeros was established and at random a single zero was replaced with a one. If the addition of the randomly assigned one improved the probability of the
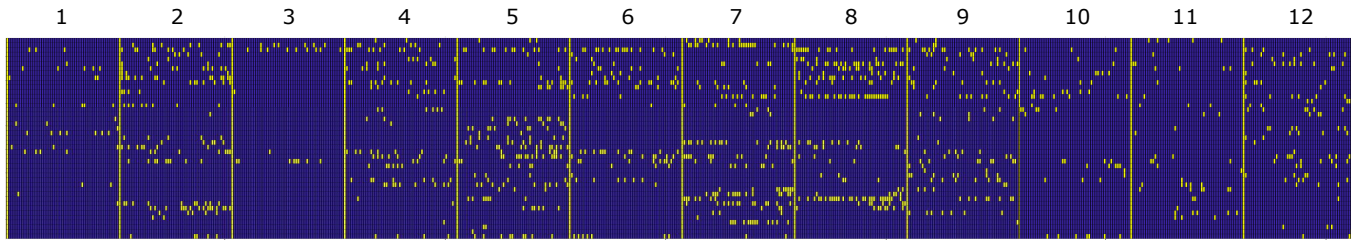
Fig. 10. A representation of how the LSTM from figure 7 interprets data to make classifications. This was built by starting off with a 44 * 50 matrix of zeros, and at a random position a 1 is added. If this addition improves the probability of the network assigning a certain classification (this is repeated through all the classifications) then the change stays. If it worsens the probability of said classification it is removed. This process is repeated for each of the 12 classifications so that the LSTM is over 99.9% sure each of the matrices is a specific classification. Then a final search is done to remove the 1's whist not reducing the 99.9% probability for the classification. What this more generally means is that for each of the 12 classifications, what is the minimum amount of data required for the LSTM to be over 99.9% sure that that data represents a given classification. As can be seen in the image, each classification requires different elements of data to be sure of an accurate classification.
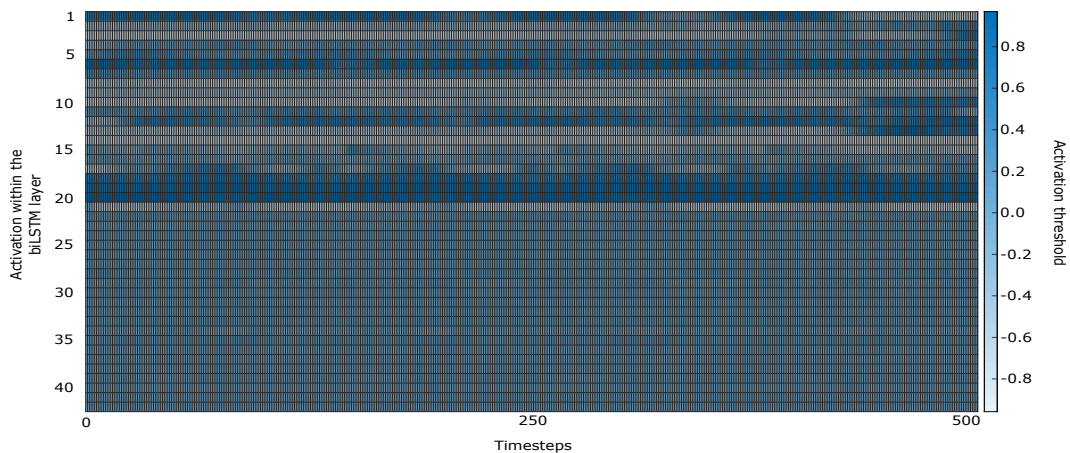
Fig. 11. A representation of the activations within the LSTM from figure 8 during a single instance of test data. What can be seen is that there are a range of different activations in the top half of the graph which corresponds to the data from the left foot. The higher and more varied the activations show that broadly those data channels are being used dynamically to alter the behaviour of the network. The data from the bottom half of the graph is more monotone, and shows that the data from the right foot is not being used to dynamically alter the state of the network as much as the top half. This suggests that the LSTM is placing more significance on the data from the left foot, which was the side of the body the knee brace was used.

network correctly classifying the PC, then the alteration to the matrix remained. If the probability was not altered or decreased the one was removed. This process was repeated for all 12 PCs until the LSTM was able to classify a specific PC with each matrix at over 99.9% probability. Finally, a search of each of the 12 matrices was completed to remove the randomly assigned ones, without reducing the classification probability in order to reveal the minimum amount of data required for a 99.9% classification probability for each of the 12 PCs. As can be seen in Figure 10, each PC requires different elements and amounts of data to be accurately classified. Perturbation condition 3 clearly required minimal data for a classification to be made (Figure 10). Moreover, the data required for a PC 3 classification exists on two horizontal bands which correspond to the locations of the accelerometry data within the matrix. This suggests that for PC 3 classification (which was classified with 90% accuracy in figure 7), only the accelerometry data may be required and pressure data was superfluous in this instance. A similar trend was evident in PC 6, although the division between the upper and lower band was not as clear (Figure 10) and some elements of pressure data were evident in

the matrix locations. All other classifications appear to require data from both the accelerometers and the pressure sensors for the LSTM to be able to determine the appropriate PC.

Figure 11 builds upon this by showing that in more real world situations, where classifications by the networks are rarely over 99% certain, the bulk of the decision making process is achieved from the data from the left foot - corresponding to the side of the body the leg brace was used. It shows that over time, the data from the left foot results in high activations and low activations meaning that certain parts of the data are used more variably then others over a single instance of data containing 500 time steps. The data from the right foot in the lower half of the figure rarely contains any activation outside of 0 ± 0.2. This suggests that for the majority of the decision making, only one of the Moticon insoles was required - and that both insoles were not treated equally in the decision making process.

Overall and on balance, it can therefore be acknowledged that the use of an appropriately trained LSTM can be used in combination with the Moticon insoles to distinguish between interacting artificially induced gait perturbations.

VII. CONCLUSION

This work has demonstrated that non-invasive wireless insoles, in combination with LSTMs can be used to classify interacting gait perturbations related to underfoot pressure and kinematic restrictions. The separation of training and testing data also demonstrated that the LSTM is capable of accurate classification on unseen data from participants not used to train the LSTM. This suggests that such technology has the potential of future deployment to better inform movement disorder diagnosis with implications to reduce the financial and time burdens in clinical care settings.

REFERENCES

[1] Paul Fiolkowski, Denis Brunt, Mark Bishop, and Raymond Woo. Does postural instability affect the initiation of human gait? *Neuroscience Letters*, 323(3):167–170, 2002.

[2] Shanshan Chen, John Lach, Benny Lo, and Guang-Zhong Yang. Toward pervasive gait analysis with wearable sensors: A systematic review. *IEEE journal of biomedical and health informatics*, 20(6):1521–1537, 2016.

[3] Walter Pirker and Regina Katzenschlager. Gait disorders in adults and the elderly. *Wiener Klinische Wochenschrift*, 129(3-4):81–95, 2017.

[4] Chandrasekar Rathinam, Andrew Bateman, Janet Peirson, and Jane Skinner. Observational gait assessment tools in paediatrics–a systematic review. *Gait & posture*, 40(2):279–285, 2014.

[5] Maria del Pilar Duque Orozco, Oussama Abousamra, Chris Church, Nancy Lennon, John Henley, Kenneth J Rogers, Julieanne P Sees, Justin Connor, and Freeman Miller. Reliability and validity of edinburgh visual gait score as an evaluation tool for children with cerebral palsy. *Gait and posture*, 49:14–18, 2016.

[6] Geruza P Bella, Nádia BB Rodrigues, Paola J Valenciano, Luciana MAE Silva, and Regina CT Souza. Correlation among the visual gait assessment scale, edinburgh visual gait scale and observational gait scale in children with spastic diplegic cerebral palsy. *Revista Brasileira de Fisioterapia*, 16(2):134–140, 2012.

[7] LW Robinson, N Clement, M Fullarton, A Richardson, J Herman, G Henderson, JE Robb, and MS Gaston. The relationship between the edinburgh visual gait score, the gait profile score and gmfcs levels i–iii. *Gait & posture*, 41(2):741–743, 2015.

[8] Robert J Palisano, Lisa Avery, Jan Willem Gorter, Barbara Galuppi, and Sarah Westcott McCoy. Stability of the gross motor function classification system, manual ability classification system, and communication function classification system. *Developmental Medicine & Child Neurology*, 60(10):1026–1032, 2018.

[9] Erich Rutz, Oren Tirosh, PAM Thomason, Alexej Barg, and H Kerr Graham. Stability of the gross motor function classification system after single-event multilevel surgery in children with cerebral palsy. *Developmental Medicine & Child Neurology*, 54(12):1109–1113, 2012.

[10] Jennifer L McGinley, Richard Baker, Rory Wolfe, and Meg E Morris. The reliability of three-dimensional kinematic gait measurements: a systematic review. *Gait & posture*, 29(3):360–369, 2009.

[11] Alexander Turner and Stephen Hayes. The classification of minor gait alterations using wearable sensors and deep learning. *IEEE Transactions on Biomedical Engineering*, 66(11):3136–3145, 2019.

[12] Alexander Rampp, Jens Barth, Samuel Schülein, Karl-Günter Gaßmann, Jochen Klucken, and Björn M Eskofier. Inertial sensor-based stride parameter calculation from gait sequences in geriatric patients. *IEEE transactions on biomedical engineering*, 62(4):1089–1097, 2014.

[13] Julius Hannink, Thomas Kautz, Cristian F Pasluosta, Karl-Günter Gaßmann, Jochen Klucken, and Bjoern M Eskofier. Sensor-based gait parameter extraction with deep convolutional neural networks. *IEEE journal of biomedical and health informatics*, 21(1):85–93, 2016.

[14] Siddharth Srivastava, Sumit Soman, Astha Rai, and Praveen K Srivastava. Deep learning for health informatics: Recent trends and future directions. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1665–1670. IEEE, 2017.

[15] Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21, 2016.

[16] Cosmin Stamate, George D Magoulas, Stefan Küppers, Effrosyni Nomikou, Ioannis Daskalopoulos, Marco U Luchini, Theano Moussouri, and George Roussos. Deep learning parkinson's from smartphone data. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 31–40. IEEE, 2017.

[17] Johannes CM Schlachetzki, Jens Barth, Franz Marxreiter, Julia Gossler, Zacharias Kohl, Samuel Reinfelder, Heiko Gassner, Kamiar Aminian, Bjoern M Eskofier, Jürgen Winkler, et al. Wearable sensors objectively measure gait parameters in parkinsons disease. *PloS one*, 12(10):e0183989, 2017.

[18] Antonio Suppa, Ardian Kita, Giorgio Leodori, Alessandro Zampogna, Ettore Nicolini, Paolo Lorenzi, Rosario Rao, and Fernanda Irrera. L-dopa and freezing of gait in parkinsons disease: Objective assessment through a wearable wireless system. *Frontiers in neurology*, 8:406, 2017.

[19] Susan A Rethlefsen, Gideon Blumstein, Robert M Kay, Frederick Dorey, and Tishya AL Wren. Prevalence of specific gait abnormalities in children with cerebral palsy revisited: influence of age, prior surgery, and gross motor function classification system level. *Developmental Medicine & Child Neurology*, 59(1):79–88, 2017.

[20] Robert Palisano, Peter Rosenbaum, Stephen Walter, Dianne Russell, Ellen Wood, and Barbara Galuppi. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine & Child Neurology*, 39(4):214–223, 1997.

[21] Fabian Horst, Sebastian Lapuschkin, Wojciech Samek, Klaus-Robert Müller, and Wolfgang I Schöllhorn. Explaining the unique nature of individual gait patterns with deep learning. *Scientific reports*, 9(1):1–13, 2019.

[22] Shaohua Wan, Yan Liang, Yin Zhang, and Mohsen Guizani. Deep multilayer perceptron classifier for behavior analysis to estimate parkinsons disease severity using smartphones. *IEEE Access*, 6:36825–36833, 2018.

[23] Ahnryul Choi, Hyunwoo Jung, Ki Young Lee, Sangsik Lee, and Joung Hwan Mun. Machine learning approach to predict center of pressure trajectories in a complete gait cycle: a feedforward neural network vs. lstm network. *Medical & Biological Engineering & Computing*, 57(12):2693–2703, 2019.

[24] Jade He, Kevin Lippmann, Najia Shakoor, Christopher Ferrigno, and Markus A Wimmer. Unsupervised gait retraining using a wireless pressure-detecting shoe insole. *Gait & posture*, 70:408–413, 2019.

[25] A. S. Alharthi and S. U. Yunas and K. B. Ozanyan. Deep Learning for Monitoring of Human Gait: A Review *IEEE Sensors Journal*, 19(21):75–82, 2019.

[26] Caldas, Rafael and Fadel, Tariq and Buarque, Fernando and Markert, Bernd. Adaptive predictive systems applied to gait analysis: A systematic review *Gait & Posture*, 77:75–82, 2020.

[27] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

[28] David Snyder, Daniel Garcia-Romero, Daniel Povey, and Sanjeev Khudanpur. Deep neural network embeddings for text-independent speaker verification. In *Interspeech*, pages 999–1003, 2017.

[29] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In *international conference on machine learning*, pages 2847–2854, 2017.

[30] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.

[31] Xi Li, Liming Zhao, Lina Wei, Ming-Hsuan Yang, Fei Wu, Yueting Zhuang, Haibin Ling, and Jingdong Wang. Deepsaliency: Multitask deep neural network model for salient object detection. *IEEE transactions on image processing*, 25(8):3919–3930, 2016.

[32] Dumitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. Scalable object detection using deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2147–2154, 2014.

[33] Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615, 2016.