Falls Risk Classification of Older Adults Using Deep Neural Networks and Transfer Learning

Matthew Martinez, Phillip L. De Leon, *Senior Member*

***Abstract*—Prior research in falls risk classification using inertial sensors has relied on the use of engineered features, which has resulted in a feature space containing hundreds of features that are likely redundant and possibly irrelevant. In this paper, we propose using fully convolutional neural networks (FCNNs) to classify older adults at low or high risk of falling using inertial sensor data collected from a smartphone. Due to the limited nature of older adult inertial gait data sets, we first pre-train the FCNN models using a publicly available data set for pedestrian activity recognition. Then via transfer learning, we train the network for falls risk classification. We show that via transfer learning, our falls risk classifier obtains an area under the receiver operating characteristic curve of 93.3%, which is 10.6% higher than the equivalent model trained without the use of transfer learning. Additionally, we show that our method outperforms other standard machine learning classifiers trained on features developed in prior research.**

***Index Terms*—Multi-layer neural networks, machine learning, accelerometers, gyroscopes**

I. Introduction

Each year in the United States, 2.8 million adults over the age of 65 are treated for fall-related injuries, which include broken bones, hip fractures, and traumatic brain injuries [1]. Because of the quality of life and economic impacts related to falling, considerable research effort has focused on falls prevention [1], [2]. Extrinsic factors contributing to an increase in falling include the home environment, medications, and footware [3]. Intrinsic risk factors include gait and balance disorders which have been found to contribute most to a heightened falls risk [3]. Although there has been substantial research focus on falls risk assessments [4], [5], their intended use is for outpatient services [6] and do not allow for continuous monitoring and measurement of gait. Gait can be measured with 3-D motion capture systems or pressure sensitive walkways, however, these systems can be intrusive and/or expensive to deploy for long term use [7]. Instead, low cost sensor platforms such as depth cameras [8], radio signal based methods [7], and inertial sensors [9] may be better suited for gait monitoring. In particular, inertial sensors are attractive due to the low cost of micro electro-mechanical systems (MEMS) based inertial measurement units (IMUs) which are now standard in smartphones and wearable devices. IMUs have been shown to be useful for gait analysis [10] and effective for falls prediction [9].

Deep learning has recently been applied to gait analysis [11], [12] and gait disorder classification [13], [14]. Machine learning has also been shown to be effective for falls prediction[1]. A review of inertial sensors and machine learning for falls risk assessment can be found in [9]. From the papers surveyed, the authors identified 130 unique features [9]. The classification methods used in these studies include decision trees, logistic regression, and support vector machine (SVM). More recently, [16] used a wavelet transform of acceleration measurements and barometric pressure to classify individuals as fallers or non-fallers. In [17], the authors used 74 features extracted from acceleration measurements of walking and turning to classify individuals as non-fallers and prospective fallers, i.e. a person who experienced a fall six-month after data collection. The best performance in this study was achieved using random forests [17].

Three challenges exist in the current falls prediction research. First, current research has focused on feature engineering, which has resulted in a high dimensional feature space where many features are likely redundant [15], [18]. Additionally, feature engineering requires expert domain knowledge, often fails to generalize to unseen data, and does not transfer well to other domains and tasks [19]. Second, retrospective falls history has been used by many researchers for label assignment and the resulting models fail to capture changes in gait mechanics leading up to a fall [9], [15], [18]. Third, many models are trained on a low number of examples, which is a result of the cost and time limitations of data collection [15] leading to models that are less robust to noisy features and labels, and are sensitive to overfitting.

In this paper, we propose using deep neural networks (DNN) and transfer learning in order to address the three challenges above for falls risk classification of older adults. We treat the falls risk classification problem as a time series classification

Matthew Martinez is with Sandia National Laboratories, Albuquerque, NM, 87185 USA and the Klipsch School of Electrical & Computer, New Mexico State University, Las Cruces, New Mexico, 88003 USA email: mtmart@sandia.gov

Phillip L. De Leon is with the Klipsch School of Electrical & Computer, New Mexico State University, Las Cruces, New Mexico, 88003 USA email: pdeleon@nmsu.edu

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

[1]In this paper we make the clear distinction between "falls prediction" and "falls risk." The former refers to using data to predict whether a person has fallen in the past and the latter refers to using data to assess a person's risk of falling in the future. For the purpose of data labeling, falls prediction uses the terms "faller" and "non-faller". These labels are associated with retrospective falls history [15] and result in a mapping that is representative of past events. Falls risk uses the terms "low risk" and "high risk" for data labeling and results in a mapping that is representative of the risk of a future fall.

task where the goal is to classify each measurement of gait as being either low or high falls risk. Our contributions are as follows. First, we show how to use readily-available pedestrian activity data to pre-train a model which learns generalized feature representations related to human pedestrian activities. This partially addresses the problem of small data sets due to the collection expense for falls risk prediction. Second, via transfer learning we use the pre-trained fully convolutional neural network (FCNN) as a feature extractor for falls risk classification. Finally, from these models, we are able to directly classify falls risk from inertial gait measurements from a smartphone which may prove to be more economical than more complex sensing systems for falls risk assessment.

The remainder of this paper is organized as follows. In Section II, we describe the time series classification problem, FCNNs and their elements, and the network architectures used for this work. In Section III, we provide an overview of data sources and processing including the publicly-available data set used for pre-training, smartphone data, data labeling, and data augmentation for time series data. In Section IV, we describe the training and evaluation of each model and provide classification results in Section V. In Section VI, we compare our DNN for falls risk classification to other feature based classifiers. Finally, we provide a discussion in Section VII and conclusions in Section VIII.

II. Time Series Classification

In a time series classification task, the goal is to classify a given vector sequence, $\mathbf{x}_t$ to one of C classes. In this paper, we treat the IMU measurements as a vector sequence where each element is a sample from an independent sensor axis. The application of DNNs to human activity recognition has focused on the use of deep convolutional neural networks (CNN) [20]; recurrent neural networks (RNN) [21], [22]; and hybrid CNN-RNN architectures [23]. We elect to use FCNNs [24] which have been shown to outperform RNNs on the same task [25], have less trainable parameters, and have achieved state-of-the-art results [26], [27]. Like RNNs, FCNNs can efficiently learn local patterns at different time scales by using dilated convolutions [25], [28] and can process arbitrary length sequences since the fully connected layer used in deep CNNs is omitted from the network architecture.

A. Dilated Causal Convolutions

One way of increasing the receptive field without an additional increase in computational cost, is through the use of a dilated causal convolution [28]

$$(h *_d x)_n = \sum_{k=0}^{N-1} h_k \cdot x_{n-dk} \tag{1}$$

where h is a filter of length N, x is the input sequence, and $d \in \mathbb{Z}^+$ is the dilation factor. Thus an effectively longer filter is achieved by inserting zeros between filter coefficients as illustrated in Fig. 1 and by increasing d exponentially with increasing network depth, the network is able to learn representations at longer time histories. Although this is similar to pooling or strided convolutions, dilated causal convolutions allow the output to remain the size as the input [28].

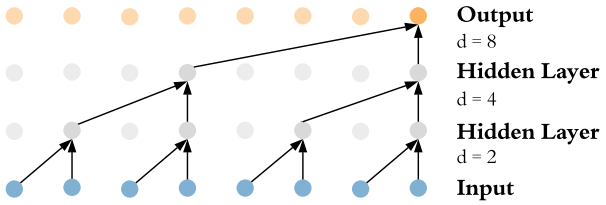


Fig. 1. Visualization of stacked dilated causal convolutional layers where the value at the output layer depends only on the prior time points at longer time scales. Adapted from [28].

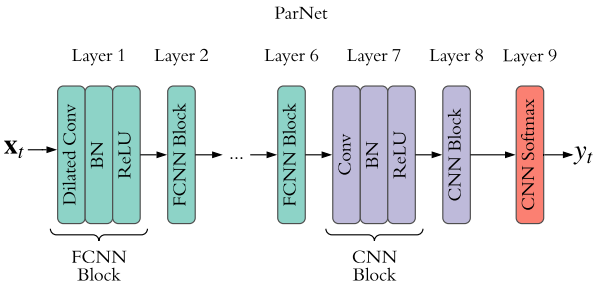


Fig. 2. ParNet architecture consists of 6 FCNN blocks, 2 CNN blocks, and a CNN softmax output layer. Each FCNN block is constructed from a dilated causal convolutional layer, a BN layer, and a ReLU activation. The CNN block uses the same elements but without dilation. Each FCNN layer learns 128 filters with a kernel width of 5 and 2 CNN blocks learn 256 filters with a kernel width of 3. The CNN softmax output layer has a kernel width of 1.

B. ParNet: Pedestrian Activity Recognition Network

We modify the FCNN in [27] for pedestrian activity recognition using the architecture in Fig. 2 and refer to this architecture as ParNet (Pedestrian Activity Recognition Network). ParNet consists of six causal dilated convolutional blocks (FCNN block), two causal convolutional blocks without dilation (CNN block), and a convolutional output layer with softmax activation which allows the network to output a scalar sequence of class predictions. This is similar to the FCNN architecture used for semantic image segmentation [24]. Each dilated causal convolutional block learns 128 filters with a kernel width of 5. The dilation rate for each FCNN block is 2^{p-1} where p is the layer number, e.g. the first convolutional block has a dilation rate of 1 and the sixth convolutional block has a dilation rate of 32. The two convolutional blocks each learn 256 filters with a kernel width of 3. The convolutional output layer uses a kernel width of 1. Each FCNN block consists of a 1-D dilated causal convolutional layer followed by a batch normalization (BN) layer [29] and a Rectified Linear Unit (ReLU) [30]. The CNN blocks use the same architecture except without dilation. In Section V-B, we apply transfer learning to ParNet for falls risk classification.

III. Data

A. Human Activity Sensing Consortium Corpora

For this work we train ParNet using the Human Activity Sensing Consortium (HASC) Pedestrian Activity Corpus 2016

TABLE I
PROPORTION (IN %) OF EACH ACTIVITY IN HASC-PAC2016 FOR SMARTPHONE PLACEMENT AT ALL LOCATIONS AND WAIST ONLY.

Placement	Activity					
	Stay	Walk	Jog	Skip	Up	Down
All (%)	17.9	17.3	17.1	17.2	15.4	15.2
Waist (%)	17.8	16.9	16.9	16.8	15.9	15.7

TABLE II
SUBSETS OF THE AUGMENTED HASC-PAC2016 DATA SET USED FOR TRAINING PARNET. THE DATA SET IS PARTITIONED BY SENSOR PLACEMENT LOCATIONS AND TYPE.

Model	Placement	Sensor Type(s)	Examples
ParNet(All, Accel)	All	Accel	47,125
ParNet(All, Accel + Gyro)	All	Accel + Gyro	47,125
ParNet(Waist, Accel)	Waist	Accel	12,882
ParNet(Waist, Accel + Gyro)	Waist	Accel + Gyro	12,882

(HASC-PAC2016) [31], which consists of IMU measurements from body-mounted smartphones and other devices manufactured by Samsung (58%), Apple (16%), ATR (13%), LG (9%), and others (4%) [31]. The sensor measurements include Walk, Jog, Skip, Upstairs, Downstairs, and Stay (no activity). Sensor measurements of each activity were made using 3-axis accelerometers, 3-axis gyroscopes as well as other sensors placed on the waist (front or back pants pocket) (48%), arm (18%), chest (11%), and other locations (23%) [31].

Measurements were collected from 510 individuals (120 female/390 male) ranging in age from 20 to 60 years (28.2 ± 12.9) with no health restrictions [31]. These are approximately 20 s in duration and are intended to be used as training examples [31]. All together the data set contains 111,027 examples for all sensors types. We only consider examples where both accelerometer and gyroscope measurements are available, and have a sampling rate of 100 Hz. We also consider measurements from all placement locations and sensor measurements collected at the waist. In total, there are 23,345 available examples for all measurements from all locations and 5,970 examples from measurements collected at the waist. Table I provides the proportions of activity examples from the total examples for the placements under consideration.

We post-processed the HASC-PAC2016 inertial data according to [31], applied additional filtering for noise reduction, and data augmentation. For each activity, we removed the first 2 s and last 5 s of signal. We then filtered out the gravity component present in each axis with a zero-phase, fourth-order Butterworth filter [32]. We found that a cutoff frequency of $f_{\text{cutoff}} = 0.15$ Hz worked best for HASC-PAC2016. Next we applied a Savitsky-Golay filter (frame length = 51, order = 3) to the inertial signals for noise reduction. Finally, we perform data augmentation as follows. For each vector sequence, we randomly select a subsequence with a uniformly distributed duration from 3 s to 10 s (6.5 s mean), which improves the network robustness to short variable length time series and increases the size of our training set. Second, we randomly rotate the data by either 0° or 180° on the x- and y- axes and 0° or 90° on the z-axis in order to simulate various smartphone orientations. This also improves the network's robustness to linear transforms of the data. The data set is partitioned by sensor placement (all placement locations or waist only placement) and sensor types (acceleration and gyroscope or acceleration only). In total, the augmented data set contains 47,125 examples (all placement locations) or 12,882 examples (waist only placement). We train our models using the architecture described in Section II-B, with one of the post-processed HASC-PAC2016 data subsets shown in Table II.

B. Older Adult Smartphone Based Gait Data

Our data set of older adult gait data was collected in partnership with the Electronic Caregiver® Company (ECG) using the Mobile Falls-Risk Assessment Unit and is used for falls risk classification. The average age of participants was 77.1 ± 7.4 years for female participants, 76.5 ± 6.6 years for male participants, and 76.9 ± 7.1 years for all participants.

The sensor platforms used to collect data include a TekScan® Walkway™ System and two Apple® iPhone® 6 smartphones. The walkway is a low-profile floor mat that measures kinetic, timing, and physical measurements of gait and as described below, assists in labeling smartphone data as low or high risk of falling. The two smartphones were used to collect inertial measurements, sampled at 100 Hz, of an individual's gait using a custom data logging app [33]. The use of the walkway and inertial sensor data for secondary data analysis was approved by the New Mexico State University Internal Review Board under reference number 15405.

During data collection, we mounted each smartphone near the individual's left and right hip bone using a gait belt and holster clip. We recorded inertial gait data for 30 s while the individual walked down the walkway, performed a turn at the end of the walkway, and finally walked back down the walkway returning to their starting location. We manually removed the standing and turning segments from the data and retained walking segments greater than 3.5 s which is sufficient, based on our walkway measurements, to capture three gait cycles. Overall, we collected 657 examples (436 female/221 male) of inertial gait data.

We label the smartphone gait segments as low or high risk of falling as follows [34]. In prior research, five gait variables (gait speed, cadence, stride length, percentage of gait cycle spent in swing phase, and percentage of gait cycle spent in double support) have been shown to contribute to an increase in the risk of falling [35]. These gait variables for each participant were directly measured using the walkway. The data were modelled with a two component Gaussian mixture model (GMM) where each component represents the low or high falls risk class. The two responsibilities are computed for each gait vector and a label is assigned to each gait segment using the threshold in [34]. For more details on our data labeling procedure, we refer the reader to [34]. Of the smartphone gait segments, 422 examples were labeled as low risk of falling and 235 were labeled as high risk of falling. We address the imbalance of low and high risk examples in Section V-B.

TABLE III
CLASSIFIER ACC FOR EACH ACTIVITY IN HASC-PAC2016.

Model	ACC (%)					
	Stay	Walk	Jog	Skip	Up	Down
ParNet(All,Accel)	99.6	95.6	97.4	98.2	98.0	93.0
ParNet(All,Accel+Gyro)	99.8	96.5	98.0	98.4	97.7	94.5
ParNet(Waist,Accel)	99.7	97.7	98.7	97.6	97.6	98.7
ParNet(Waist,Accel+Gyro)	99.4	98.3	99.7	98.5	98.7	98.4

TABLE IV
SUMMARY OF PARNET PERFORMANCE (%). PARNET($\star$, ACCEL + GYRO) OUTPERFORMED PARNET($\star$, ACCEL) AND PARNET(WAIST, $\diamond$) OUTPERFORMED PARNET(ALL, $\diamond$)

Model	Macro-PRC	Macro-RCL	Macro-F1	ACC
ParNet(All, Accel)	97.0	97.0	97.0	97.0
ParNet(All, Accel + Gyro)	97.4	97.5	97.4	97.4
ParNet(Waist, Accel)	98.3	98.4	98.3	98.3
ParNet(Waist, Accel + Gyro)	**98.8**	**98.8**	**98.8**	**98.8**

IV. PEDESTRIAN ACTIVITY RECOGNITION

A. Training

ParNet was implemented using the PyTorch deep learning framework [36]. Model training and evaluation was performed using two Nivida GeForce® GTX 980 GPUs. Each model in Table II was trained for a pedestrian activity classification task in a fully supervised manner for 250 epochs. Network parameters were optimized using mini-batch gradient decent to minimize cross-entropy loss [19], where each mini-batch consisted of 64 examples. Optimization was performed using the AMSGrad [37] variant of the Adam optimization method [38] using a learning rate of 10^{-5} and an epsilon value of 10^{-4}. Additionally, to improve network generalization we employed L^2 regularization with a regularization coefficient of 10^{-2}.

B. Evaluation

Each sequence is classified as one of six activities described in Section III-A by calculating the mode over the sequence of class predictions. Classification accuracy (ACC) for each activity for each model is given in Table III. Each model is assessed using macro-precision (PRC), macro-recall (RCL), macro-F1, and ACC, where macro scores are the average of individual scores for the six classes. From Table IV, the overall best model was ParNet(Waist, Accel + Gyro), with ACC of 98.8%. ParNet($\star$, Accel + Gyro)[2] outperformed ParNet($\star$, Accel) and ParNet(Waist, $\diamond$) outperformed ParNet(All, $\diamond$).

In [31], the authors provide a benchmark for the HASC-PAC2016 data set using a random forest classifier [39] and features extracted from 2 s and 4 s frames of acceleration data. Using this method, the authors achieved an ACC of 73.4% [96.9% (stay), 62.2% (walk), 87.1% (jog), 81.4% (skip), 57.5% (up), 51.9% (down)] for all sensor locations and 81.4% [96.0% (stay), 69.0% (walk), 91.3% (jog), 91.0% (skip), 76.2% (up), 56.0% (down)] when the sensor was placed on the waist (right pocket). Additionally, the authors in [22] achieved an ACC of 95.4% using a RNN [19]. The network was trained on acceleration measurements from the HASC 2011 corpus [40]. Based on the publicly available benchmarks, our FCNN architecture outperforms the random forest and is comparable to the RNN. Thus, ParNet provides a foundation for transfer learning for falls risk classification.

[2]We use $\star$ to denote both sensor placement options, i.e. all and waist only and $\diamond$ to denote both sensor combinations, i.e. accelerometer only and accelerometer and gyroscope.

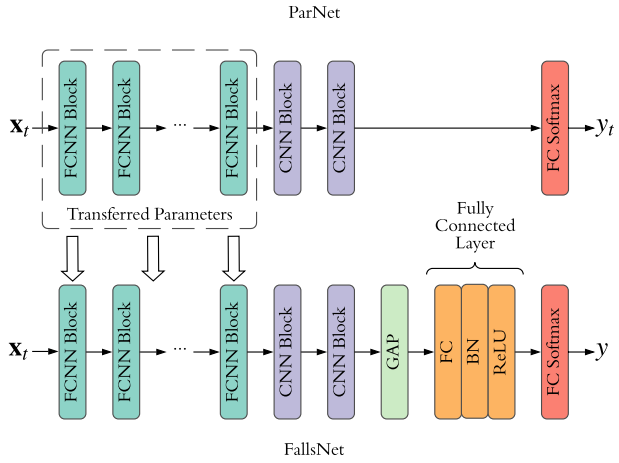


Fig. 3. Illustration of the development of FallsNet classifier. FallsNet has a similar architecture to ParNet with the exception of an additional GAP layer after the CNN blocks and fully connected layer. The learned parameters from the FCNN blocks in ParNet are transferred to FallsNet prior to training the remaining layers with the older adult gait data. Whereas ParNet makes a human activity classification decision every sample, FallsNet makes a low or high risk of falling decision for every input sequence.

V. TRANSFER LEARNING FOR FALLS RISK CLASSIFICATION

Transfer learning is a machine learning technique that uses a model trained on one task for an auxiliary task, where the data is in a different domain or the distribution differs from the source task [19]. In deep learning, the architecture is first trained with a sufficiently large data set and then a subset of layers are re-trained using data from the auxiliary task. The hypothesis of transfer learning as applied to DNNs is that the initial layers learn generalized feature representations and as the network depth increases, the representations become more task specific [19]. Transfer learning can be useful when the data from the auxiliary task is insufficient for the architecture to generalize well and has been shown to be effective for computer vision applications [41].

A. FallsNet: Falls Risk Classification Network

We modify ParNet (Fig. 2) for the falls risk classification task as illustrated in Fig. 3. First, after the CNN block in Layer 8, we add a global average pooling (GAP) layer [42] and a fully connected layer with 256 hidden units. Second, we replace the CNN softmax output layer (Layer 9) with a fully connected softmax output layer, which produces a binary decision of low or high risk of falling. The GAP layer provides dimensionality reduction prior to classifying each sequence and the fully connected layer uses a ReLU activation and

incorporates dropout [19]. All parameters for layers 1 - 8 are the same as those given in Section II-B. We refer to this architecture as FallsNet (Falls Risk Classification Network).

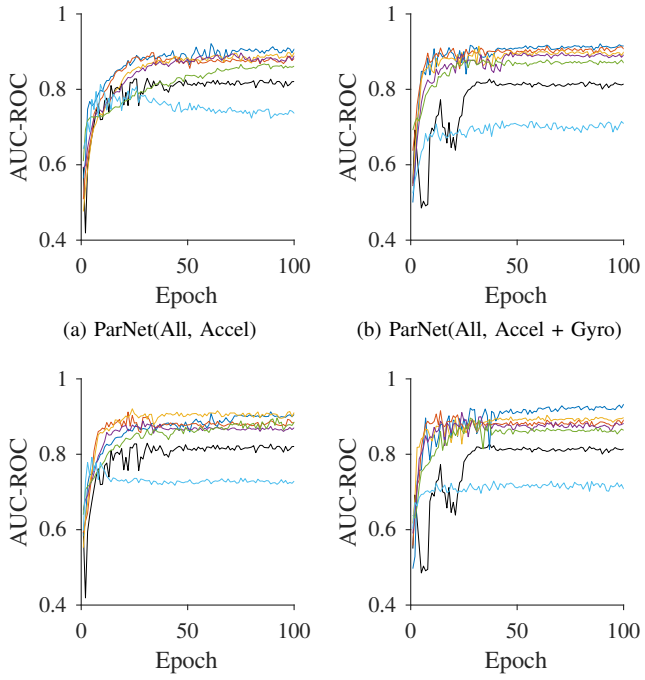
B. Transfer Learning Experiments and Training

We apply transfer learning by transferring the first l layers in ParNet to the first l layers of FallsNet; the remaining $6-l$ layers of FallsNet are randomly initialized. Then, FallsNet is trained in a supervised fashion by only backpropagating through the randomly initialized layers. We then vary l from one to six to assess which model in Table II and how many layers to transfer, provides the most generalized features and ultimately the best model performance. In total, 24 models were trained. In addition, two baseline models (accelerometer only and accelerometer and gyroscope) using the older adult gait data were trained without the use of transfer learning, i.e. $l = 0$. Due to a class imbalance in the older adult gait data, stratified sampling was performed during construction of each mini-batch. This ensures that the number of positive and negative examples in each mini-batch had the same proportion as the training set. Additionally, we used a learning rate scheduler that decreased the learning rate by 10^{-3} once the validation loss did not continue to decrease for 10 epochs. A stratified resampling scheme was used to partition the older adult gait data into training and validation sets, where 80% (525) of the examples were used for training and 20% (132) were used for validation. Otherwise, training optimization and regularization for FallsNet was identical to ParNet.

C. Evaluation

The four ParNet models in Table II and the transfer learning experiments in Section V-B are evaluated using the area under the receiver operating characteristic curve (AUC-ROC) [39]. Fig. 4 shows the AUC-ROC scores versus epoch for the validation set (only the first 100 training epochs are shown since there is only minor improvement after approximately 50 epochs). The black curve in each figure represents the baseline model, i.e. model trained without transfer learning. In each plot, all FallsNet models had higher starting AUC-ROC scores, higher slopes, and obtained a higher AUC-ROC score than the baseline FallsNet models with the exception when all ParNet layers were transferred to FallsNet as denoted by (—) in Fig. 4.

The best AUC-ROC scores obtained by FallsNet when trained without transfer learning (baseline) were 82.9% (acceleration only) and 82.7% (acceleration and gyroscope). From Table V, the best results from FallsNet were obtained by transferring the first layer weights ($l = 1$) from ParNet(Waist, Accel + Gyro). This model achieves an AUC-ROC score of 93.3%, which is a 10.6% increase in the AUC-ROC score over the equivalent baseline model. With the decision threshold set for equal error rate (EER), this model achieves 86.4% ACC, 85.1% sensitivity (SENS), and 87.1% specificity (SPEC). From these results we observe that transfer learning provides improvement in model performance.



(a) ParNet(All, Accel) (b) ParNet(All, Accel + Gyro)

(c) ParNet(Waist, Accel) (d) ParNet(Waist, Accel + Gyro)

Fig. 4. AUC-ROC versus epoch for each ParNet model in Table II. In each plot, the curves as denoted by (—), (—), (—), (—), (—), (—) correspond to number of layers transferred $l = 1, 2, 3, 4, 5, 6$, respectively. Additionally, (—) denotes the baseline FallsNet model trained without transfer learning. All FallsNet models have higher starting AUC-ROC values, higher slopes, and obtained higher AUC-ROC scores than the baseline FallsNet models, with the exception of when all ParNet layers are transferred to FallsNet.

TABLE V
AUC-ROC SCORES (%) FOR PARNET MODELS IN TABLE II VERSUS NUMBER OF LAYERS TRANSFERRED FROM PARNET TO FALLSNET. THE OVERALL BEST FALLSNET MODEL WAS OBTAINED USING PARNET(WAIST, ACCEL + GYRO) FOR $l = 1$.

Model	Layers Transferred, l, to FallsNet					
	1	2	3	4	5	6
ParNet(All, Accel)	92.1	89.7	90.4	89.7	87.0	**81.3**
ParNet(All, Accel + Gyro)	92.1	91.7	91.5	90.1	87.9	71.9
ParNet(Waist, Accel)	91.3	91.2	**92.1**	88.8	**89.7**	79.2
ParNet(Waist, Accel + Gyro)	**93.3**	**91.5**	90.3	**91.1**	86.9	73.5

VI. FEATURE BASED FALLS RISK CLASSIFICATION

In this section, we compare the proposed DNN and transfer learning approach for falls risk classification to standard machine learning approaches. Thus, we develop other falls risk classifiers (logistic regression, random forest, and SVM) [39], [43] trained on features identified from prior work.

A. Feature Extraction

In [9], the authors determined that features extracted from the magnitude spectrum of acceleration measurements of gait are the most discriminating for falls prediction. These are features 11-13 and 19-33 from Table I in [44] and are extracted according to [45]. The features include the fundamental frequencies for the x-, y-, and z-axis; ratios of the area under the first harmonic (fundamental, second, third, and fourth) to the sum of the area under the first six harmonics; ratio of the

sum of the area under the first six harmonics to the sum of the remaining area under the spectrum; and ratio of the sum of the area under the even harmonics to the sum of the area under the odd harmonics.

B. Training and Evaluation

Using the features described above, we train a logistic regression classifier [39], random forest classifier, and SVM with a radial basis function kernel [43] for falls risk classification [39]. Each model was trained using a Monte Carlo simulation for 100 iterations, where each iteration uses 80% of the data for training and 20% for validation. During training, we first optimize hyperparameters using Bayesian optimization with 5-Fold cross validation [39]. We then re-train the model using the best hyperparameters and determine the decision threshold for an EER. For validation, we evaluate each model using the AUC-ROC and the threshold found during training.

The ACC, SENS, and SPEC for all models are presented in Table VI. These results are comparable to those in [18] where the authors analyzed feature selection methods for predicting prospective fall occurrence not falls risk which while similar, is a more challenging task. Their best model with feature selection achieved an ACC of 65% and SENS of 59%. Without feature selection their best model achieved an ACC of 56% and SENS of 42%. When compared to either FallsNet which uses acceleration and gyroscope measurements and FallsNet which uses acceleration measurements only, these classifiers perform worse than our models in Section V-C. Although the work in [9] did not find gyroscope measurements discriminating, FallsNet appears to leverage this additional information in a beneficial way.

TABLE VI
COMPARISON OF FALLS RISK CLASSIFIERS USING THE OLDER ADULT GAIT DATA. FALLSNET IS FOR THE BEST MODELS OBTAINED IN SECTION V-C. SVM, LOGISTIC REGRESSION, AND RANDOM FORESTS WERE TRAINED USING FEATURES DESCRIBED IN SECTION VI-A.

Model	ACC (%)	SENS (%)	SPEC (%)
FallsNet (Accel + Gyro)	**86.4**	**85.1**	**87.1**
FallsNet (Accel)	**82.6**	**83.0**	**82.4**
Logistic Regression	58.1	56.6	59.0
Random Forests	63.8	43.9	74.8
SVM	59.6	53.9	62.8

VII. DISCUSSION

Based on the results in Table IV, we observe that training with both accelerometer and gyroscope measurements yields better low risk/high risk falls classification results than training with accelerometer only measurements. By training with just acceleration measurements, FallsNet is only capable of learning features related to gait dynamics [46], [47]. However, by including rotation rate measurements, the network is capable of learning additional features related to joint rotation [48]. Additionally, models trained on data from a single sensor placed on the waist generalized better than models trained on measurements from all sensor placement locations. This is a direct result of the network only having to learn feature representations for a single placement location, instead of having to learn feature representations for various placement locations. This is consistent with the results published in [31].

In terms of training and subsequent evaluation of FallsNet, we find that ParNet(Waist, Accel + Gyro) has the best transfer learning abilities, since the placement location of the HASC-PAC2016 waist only data is in the proximity of the smartphone placement location of the older adult gait data. We hypothesize that due to similar sensor placements for both data sets near the hip, we need only transfer the first layer of ParNet to FallsNet and train subsequent layers with the older adult gait data in order to achieve the highest AUC-ROC scores. Thus, layer 1 appears to sufficiently learn generalized features related to human gait. FallsNet performance, as observed in Table V, begins to degrade when the weights from deeper layers of ParNet, which are more task specific, are transferred to FallsNet. These results are consistent with the findings in [49] which analyzed transfer learning for activity recognition.

There are at least three limitations of this study. First, our research is based solely on inertial measurements of gait. However, there are many factors both intrinsic and extrinsic which contribute to falls risk [3]. By incorporating additional, non-gait based factors the real-world predictive accuracy of our method maybe improved. Second, the older adult gait data was collected in a laboratory like setting and with smartphones carefully placed on a gait belt near the hip. Thus, greater variability in the gait data under normal walking conditions could be expected. Third, at this time we are unable to follow up with study participants to determine whether or not they had experienced a fall after data collection, however, our work presents a methodology that can be applied to future longitudinal studies of falls risk classification.

VIII. CONCLUSIONS

In this paper, we have proposed a method for classifying older adults at either low or high falls risk using inertial gait data acquired from a smartphone. First, we trained a DNN composed of eight convolutional layers (including six with dilated causal convolutions) and a convolutional softmax output layer using the HASC-PAC2016 corpus. Using this network (ParNet), we achieved an ACC of 98.8% for pedestrian activity classification which is comparable to prior results. Second, we modified ParNet to include a global average pooling layer, fully connected layer, and a fully connected softmax output layer to construct our falls risk classifier, i.e. FallsNet. Third, we applied transfer learning due to the limited number of examples in the older adult gait data. By transferring the first layer of ParNet to FallsNet and retraining the subsequent layers in FallsNet, we achieved an AUC-ROC of 93.3% which is 10.6% higher than without transfer learning. When compared to standard machine learning methods trained on features extracted from the magnitude spectrum of acceleration, FallsNet has a higher ACC, SENS, and SPEC for the falls risk classification task.

REFERENCES

[1] "Centers for Disease Control and Prevention, National Center for Injury Prevention and Control," www.cdc.gov/injury/wisqars, accessed: 2018-04-18.

[2] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *J. Am. Geriatr. Soc.*, vol. 66, no. 4, pp. 693–698, Apr. 2018.

[3] A. F. Ambrose, G. Paul, and J. M. Hausdorff, "Risk factors for falls among older adults: A review of the literature," *Maturitas*, vol. 75, no. 1, pp. 51 – 61, May 2013.

[4] J. A. Stevens and E. A. Phelan, "Development of steadi: A fall prevention resource for health care providers," *Health Promotion Practice*, vol. 14, no. 5, pp. 706–714, Sep. 2013.

[5] "Stopping elderly accidents, deaths, & injuries," www.cdc.gov/steadi, accessed: 2018-04-18.

[6] K. L. Perell, A. Nelson, R. L. Goldman, S. L. Luther, N. Prieto-Lewis, and L. Z. Rubenstein, "Fall risk assessment measures an analytic review," *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 56, no. 12, pp. M761–M766, Dec. 2001.

[7] C.-Y. Hsu, Y. Liu, Z. Kabelac, R. Hristov, D. Katabi, and C. Liu, "Extracting gait velocity and stride length from surrounding radio signals," in *Proc. 2017 CHI Conf. Human Factors Computing Syst.*, 2017, pp. 2116–2126.

[8] E. Stone and M. Skubic, "Evaluation of an inexpensive depth camera for in-home gait assessment," *J. Ambient Intell. Smart Environ.*, vol. 3, no. 4, pp. 349–361, May 2011.

[9] J. Howcroft, J. Kofman, and E. D. Lemaire, "Review of fall risk assessment in geriatric populations using inertial sensors," *J. Neuroengineering Rehabil.*, vol. 10, no. 91, pp. 1–12, Aug. 2013.

[10] S. Chen, J. Lach, B. Lo, and G. Z. Yang, "Toward pervasive gait analysis with wearable sensors: A systematic review," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 6, pp. 1521–1537, Sep. 2016.

[11] J. Hannink, T. Kautz, C. F. Pasluosta, K. G. Gaßmann, J. Klucken, and B. M. Eskofier, "Sensor-based gait parameter extraction with deep convolutional neural networks," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 85–93, Dec. 2017.

[12] O. Costilla-Reyes, P. Scully, and K. B. Ozanyan, "Deep neural networks for learning spatio-temporal features from tomography sensors," *IEEE Trans. Ind. Electron*, vol. 65, no. 1, pp. 645–653, Aug. 2018.

[13] D. Slijepcevic, M. Zeppelzauer, A. Gorgas, C. Schwab, M. Schüller, A. Baca, C. Breiteneder, and B. Horsak, "Automatic classification of functional gait disorders," 2017, arXiv:1712.06405.

[14] T. T. Um, F. M. J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interaction*, 2017, pp. 216–220.

[15] T. Shany, K. Wang, Y. Liu, N. H. Lovell, and S. J. Redmond, "Review: Are we stumbling in our quest to find the best predictor? over-optimism in sensor-based models for predicting falls in older adults," *Healthc. Technol. Lett.*, vol. 2, pp. 79–88(9), Aug. 2015.

[16] A. Ejupi, M. Brodie, S. R. Lord, J. Annegarn, S. J. Redmond, and K. Delbaere, "Wavelet-based sit-to-stand detection and assessment of fall risk in older people using a wearable pendant device," *IEEE Trans. Biomed. Eng*, vol. 64, no. 7, pp. 1602–1607, Oct. 2017.

[17] D. Drover, J. Howcroft, J. Kofman, and E. D. Lemaire, "Faller classification in older adults using wearable sensors based on turn and straight-walking accelerometer-based features," *Sensors*, vol. 17, no. 6, Jun. 2017.

[18] J. Howcroft, E. D. Lemaire, and J. Kofman, "Prospective elderly fall prediction by older-adult fall-risk modeling with feature selection," *Biomed. Signal Process. and Control*, vol. 43, pp. 320 – 328, Aug. 2018.

[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[20] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proc. 24th Int. Conf. Artificial Intell.*, 2015, pp. 3995–4001.

[21] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," 2016, arXiv:1604.08880.

[22] M. Inoue, S. Inoue, and T. Nishida, "Deep recurrent neural network for mobile human activity recognition with high throughput," *Artificial Life and Robotics*, vol. 23, no. 2, p. 173–185, Jun. 2017.

[23] F. J. Ordóñez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, Jan. 2016.

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[25] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, arXiv:1803.01271.

[26] R. Mittelman, "Time-series modeling with undecimated fully convolutional neural networks," 2015, arXiv:1508.00317.

[27] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 Int. Joint Conf. on Neural Networks*, 2017, pp. 1578–1585.

[28] A. van den c, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. W. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," 2016, arXiv:1609.03499.

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, arXiv:1502.03167.

[30] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learning*, 2010, pp. 807–814.

[31] H. Ichino, K. Kaji, K. Sakurada, K. Hiroi, and N. Kawaguchi, "HASC-PAC2016: Large scale human pedestrian activity corpus and its baseline recognition," in *Proc. 2016 ACM Int. Joint Conf. Pervasive Ubiquitous Computing*, 2016, pp. 705–714.

[32] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *21st European Symp. on'Artificial Neural Networks, Computational Intell. Mach. Learning*, 2013, pp. 437–442.

[33] M. Martinez and P. L. De Leon, "A smartphone-based gait data collection system for the prediction of falls in elderly adults," in *Proc. Int. Telemetering Conf.*, 2015.

[34] M. Martinez, P. L De. Leon, and D. Keeley, "Bayesian classification of falls risk," *Gait & Posture*, vol. 67, pp. 99–103, Jan. 2019.

[35] J. Verghese, R. Holtzer, R. B. Lipton, and C. Wang, "Quantitative gait markers and incident fall risk in older adults," *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 64A, no. 8, pp. 896–901, Aug. 2009.

[36] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *31st Conf. Neural Inform. Process. Syst.*, 2017.

[37] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in *Int. Conf. Learning Representations*, 2018.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.

[39] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, Heidelberg, Dordrecht, London: Springer, 2013.

[40] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, Y. Sumi, and N. Nishio, "HASC challenge: Gathering large scale human activity corpus for the real-world activity understandings," in *Proc. 2nd Augmented Human Int. Conf.*, 2011, pp. 27:1–27:5.

[41] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1717–1724.

[42] M. Lin, Q. Chen, and S. Yan, "Network in network," 2013, arXiv:1312.4400.

[43] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[44] Y. Liu, S. J. Redmond, N. Wang, F. Blumenkron, M. R. Narayanan, and N. H. Lovell, "Spectral analysis of accelerometry signals from a directed-routine for falls-risk estimation," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2308–2315, Aug. 2011.

[45] M. Martinez, P. L. De Leon, and D. Keeley, "Novelty detection for predicting falls risk using smartphone gait data," in *Proc. IEEE Int. Conf. Acoustics, Speech & Signal Proc. (ICASSP)*, 2017.

[46] A. Godfrey, R. Conway, D. Meagher, and G. ÓLaighin, "Direct measurement of human movement by accelerometry," *Med. Eng. Phys.*, vol. 30, no. 10, pp. 1364–1386, Dec. 2008.

[47] J. J. Kavanagh and H. B. Menz, "Accelerometry: A technique for quantifying movement patterns during walking," *Gait & Posture*, vol. 28, no. 1, pp. 1–15, Jul.. 2008.

[48] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, pp. 2255–2283, Feb. 2012.

[49] F. J. Ordóñez and D. Roggen, "Deep convolutional feature transfer across mobile activity recognition domains, sensor modalities and locations," in *Proc. ACM Int. Symp. Wearable Computers*, 2016, pp. 92–99.

Matthew Martinez received the B.S. in Electrical Engineering from New Mexico State University in 2007, the M.S. in Electrical Engineering from Colorado State University in 2011, and the Ph.D. in Electrical Engineering from New Mexico State University in 2018. He is currently at Sandia National Laboratories, where he works in statistical signal processing, machine learning, and deep neural networks.

Phillip L. De Leon (SM'03) received the B.S. Electrical Engineering and the B.A. in Mathematics from the University of Texas at Austin, in 1989 and 1990 respectively and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Colorado at Boulder, in 1992 and 1995 respectively. Currently, he serves as Associate Dean of Research in the College of Engineering and is a Professor in the Klipsch School of Electrical and Computer Engineering at New Mexico State University. His research interests include audio and speech processing, machine learning, and time-frequency analysis.