

Frailty Identification using Heart Rate Dynamics: A Deep Learning Approach

Maryam Eskandari, Saman Parvaneh, Hossein Ehsani, Mindy Fain, Nima Toosizadeh

Abstract—Previous research showed that frailty can influence autonomic nervous system and consequently heart rate response to physical activities, which can ultimately influence the homeostatic state among older adults. While most studies have focused on resting state heart rate characteristics or heart rate monitoring without controlling for physical activities, the objective of the current study was to classify pre-frail/frail vs non-frail older adults using heart rate response to physical activity (heart rate dynamics). Eighty-eight older adults (≥ 65 years) were recruited and stratified into frailty groups based on the five-component Fried frailty phenotype. Groups consisted of 27 non-frail (age= 78.80 ± 7.23) and 61 pre-frail/frail (age= 80.63 ± 8.07) individuals. Participants performed a normal speed walking as the physical task, while heart rate was measured using a wearable electrocardiogram recorder. After creating heart rate time series, a long-short term memory model was used to classify participants into frailty groups. In 5-fold cross validation evaluation, the long-short term memory model could classify the two above-mentioned frailty classes with a sensitivity, specificity, F1-score, and accuracy of 83.0%, 80.0%, 87.0%, and 82.0%, respectively. These findings showed that heart rate dynamics classification using long-short term memory without any feature engineering may provide an accurate and objective marker for frailty screening.

Index Terms—Long short-term memory, Heart rate variability, Frailty, Aging, Data augmentation, Classification, Machine learning, Deep learning

This work was supported by an award from the National Institute of Aging (NIA/NIH - Phase 2B Arizona Frailty and Falls Cohort 2R42AG032748-04).

Eskandari M. is with the Department of Computer Sciences, University of Arizona, Tucson, AZ 85719 USA (e-mail: maryameskandari@arizona.edu).

Parvaneh S. is with Edwards Life Sciences, CA, USA (e-mail: parvaneh@ieee.org).

Ehsani H. is with the Kinesiology Department, University of Maryland, College Park, MD, USA (e-mail: hossein.ehsani.bme@gmail.com).

Fain M. is with the Division of Geriatrics, General Internal Medicine and Palliative Medicine, Department of Medicine, University of Arizona, AZ, USA (e-mail: mfain@aging.arizona.edu).

Toosizadeh N. is with the Department of Biomedical Engineering and the Division of Geriatrics, General Internal Medicine and Palliative Medicine, Department of Medicine, University of Arizona, AZ, USA (email: ntoosizadeh@arizona.edu).

I. INTRODUCTION

THIS world's older population is rapidly growing, and it has been estimated that the population of adults aged 60 years or more will increase from 600 million individuals in 2000 to two billion by 2050 [1]. It is now more important than ever to identify and treat aging-related syndromes, among which, frailty is highly prevalent and clinically important. Frailty is a common geriatric syndrome characterized by an increased vulnerability to external stressors and results in hospitalization, cognitive decline, institutionalization, delirium, and mortality [2][3][4][5].

One cause of frailty is the homeostenosis phenomenon, which is the reduced ability to maintain homeostasis or physiological balance under stress [6]. To understand the underlying mechanism leading to frailty, several physiological systems have been studied, including the cardiovascular system. Evaluating resting-state heart rate (HR) showed an impaired cardiac autonomic nervous system (ANS) among frail older adults, which was represented by smaller HR variability (HRV) and reduced complexity [7][8][9]. The ANS, including the cardiac ANS, plays an important role in maintaining homeostasis by regulating physical activities [10][11]. This role becomes particularly important while performing essential activities of daily living [12]. Aging would cause impairments in the cardiac ANS performance, which can lead to impairments in HR response to physical activity. Theoretically, an impaired ANS, which already suffers from decreased physiological reserves due to frailty, is not able to restore cardiac output and maintain hemodynamic homeostasis under stress from the physical activity [13][14]. Sympathetic and parasympathetic activation regulate the HR by changing HR rate up and down, to reduce a negative or positive deviation from the equilibrium and keep HR homeostasis [15].

Impairment in cardiac autonomic control with frailty was demonstrated as an impaired orthostatic HR and a slowed recovery of systolic blood pressure during standing [8], which may relate to sinoatrial node dysfunction due to alterations in electrical conduction and action potential morphology [16]. Accordingly, frail elderly adults are more vulnerable to develop cardiovascular diseases and health complications including myocardial infarction [17].

In previous research machine learning approaches have been implemented to study the association between motor performance and frailty. Arshad et al. [18] implemented Convolutional Neural Networks (CNN) to study frailty, for which, gait data were collected using inertial measurement unit

(IMU) sensors. Segmented signals collected from sensors in this study were encoded as images and provided 85.1% accuracy in predicting frailty/pre-frailty. In another study by Aponte-Hao et al. [19] the performance of identifying frail participants was assessed from a sample of 5466 participants within the Primary Care Sentinel Surveillance Network (CPCSSN). Several machine learning approaches were utilized including elastic net logistic regression, support vector machine (SVM), random forest, XGBoost, and feedforward neural network (NN) models. Among eight models, XGBoost showed the best performance in identifying frail participants with an accuracy of 84.84%. In a larger longitudinal study, Tarekegn et al. used different machine learning models including artificial neural network (ANN), genetic programming (GP), SVM, random forest, logistic regression, and decision tree to develop predictive models for identifying frailty in older adults. For this study administrative health dataset of 1,095,612 elderly participants aged 65 or older was used, with 58 input variables including age, emergency department visits with red code, number of urgent hospitalizations, having disability, having diabetes, having anemia and 6 output variables (mortality, disability, urgent hospitalization, fracture, preventable hospitalization, and accessing the emergency department). Among all models ANN showed highest performance with 78% accuracy. Lastly, according to Friedrich et al. [20] deep learning approach (i.e., long short-term memory-LSTM) has better performance compared to machine learning approaches in frailty identification when time-series data from wearable sensor were implemented as input, for which the LSTM deep neural network was able to achieve 95% accuracy.

Although many studies have shown strong associations between frailty with motor function and resting-state HR characteristics [21], limited evidence exists to show the association between frailty and HR behavior during physical activity. In our series of investigations, we have previously observed a weaker and slower HR response to walking among pre-frail/frail older adults, compared to non-frail individuals [22].

The aim of the current study was to assess capability of electrocardiogram (ECG) assessment for identifying frailty status in older adults using LSTM model. The main hypothesis was that using HR response to physical activity and LSTM model, we would predict pre-frailty/frailty with an accuracy equal or better than what has been commonly observed using motor function ($\geq 80\%$ accuracy). The secondary hypothesis was that the LSTM deep learning approach would better predict frailty compared to engineered ECG parameters or shallow learning. For assessing the hypotheses, we implemented multiple LSTM architectures to classify pre-frail/frail versus non-frail older adults using RR intervals extracted from ECG signals. Of note, the sequence of Q, R, and S waves represent ventricular depolarization within the ECG signal. Machine learning and deep learning algorithms demonstrated great performance in pattern recognition, causal inferences, and classification of HR time series for frailty assessment [23][24][25][26], among which LSTM method have been outperformed other approaches [24][27]. Further, to confirm our secondary hypothesis, we compared LSTM result with classic machine learning approaches. We also studied the impact of 1)

converting unevenly sampled HR data to evenly sampled HR as an input to LSTM and 2) data augmentation on frailty identification to account for small number of participants and unbalance data set (unequal number of participants in each frailty group).

II. MATERIALS AND METHOD

A. Participants

Participants were recruited from the Arizona Frailty and Fall Cohort, including primary, secondary, and tertiary health care settings, community providers, assisted living facilities, retirement homes, and aging service organizations between October 2016 and March 2018. Inclusion criteria were being age 65 years or older and being able to walk a minimum distance of 9.14m (30 feet). Participants were excluded from the study for any of the following conditions: gait or mobility disorders (Parkinson's disease, multiple sclerosis, or recent stroke), cognitive impairment identified by a Mini-Mental State Examination (MMSE) score ≤ 23 [28], terminal illness, diseases/disorders that can directly influence HR (including arrhythmia and use of pacemaker), and usage of β -blockers or similar medications that can influence HR. This study was conducted according to the principles expressed in the declaration of Helsinki [29] and written informed consent obtained from eligible participants. The study was approved by the University of Arizona Institutional Review Board.

Eighty-eight participants were recruited, among whom, 27 (10 males, 17 females) were non-frail (age= 78.80 ± 7.23 years) and 61 (17 males, 44 females) were pre-frail/frail (54 pre-frail and 7 frail, age= 80.63 ± 8.07). None of the demographic information was significantly different between the pre-frail/frail and non-frail groups ($p > 0.07$, Table I).

B. Frailty assessment

Frailty was assessed based on five physical characteristics, including [30]: 1) self-reported unintentional weight loss of 4.54 kg (10 pounds) or more in the previous year; 2) weakness based on grip strength measurements from both left and right arms (adjusted with body mass index (BMI) and sex); 3) slowness based on the required time to walk 4.57m or 15 feet (adjusted with height and sex); 4) self-reported exhaustion based on a short two-question version of Center for Epidemiological Studies Depression (CES-D) scale; and 5) self-reported low energy expenditure based on a short version of Minnesota Leisure Time Activity questionnaire [31]. Participants were categorized as frail if they met at least three criteria out of five, pre-frail if they met one or two criteria, and non-frail if they met none of the criteria. Due to a small number of frail individuals in this study ($n=7$), participants were classified into two groups of "non-frail" and "pre-frail/frail". For assessing frailty, the Fried phenotype was used as the most validated and reliable tool, and because it incorporates measures associated with physical frailty (e.g., gait test and grip strength measures) [32].

C. Gait test and data acquisition

Gait assessment was conducted at participants' home under the normal speed condition, for a distance of 4.57m (15 feet). In addition to capturing HR during walking test, the baseline and recovery HR were collected while the participant was standing

TABLE I
DEMOGRAPHIC INFORMATION FOR FRAIL/PRE-FRAIL AND NON-FRAIL GROUPS

Demographic information	Non-frail (n=27)	Pre-frail/Frail (n=61)	p-value (Effect size)
Sex, male (% of the group)	10 (37%)	17 (29%)	0.24
Age, year (SD)	77.80 (7.23)	80.63 (8.07)	0.21 (0.30)
Height, cm (SD)	167.54 (11.4)	164.16 (9.89)	0.07 (0.63)
Weight, kg (SD)	73.30 (14.81)	74.56 (19.92)	0.60 (0.13)
Body mass index, kg/m ² (SD)	26.29 (6.05)	27.50 (6.12)	0.17 (0.32)
Engineered HR variables			
Time to peak HR, second (SD)	6.03 (2.34)	8.96 (3.70)	<0.01* (0.93)
HR recovery time, second (SD)	12.49 (4.15)	11.14 (3.63)	0.12 (0.35)
HR Percent increase, % (SD)	22.03 (19.36)	11.99 (8.67)	<0.01* (0.67)
HR Percent decrease, % (SD)	16.95 (11.60)	10.99 (7.24)	0.01* (0.62)
HR mean, beats per minute (SD) - Baseline	75.08 (11.98)	78.67 (15.41)	0.25 (0.26)
RR mean, second (SD) - Baseline	0.82 (0.12)	0.79 (0.15)	0.30 (0.22)
RR CV, % (SD) - Baseline	4.21 (5.56)	1.94 (2.15)	<0.01* (0.54)
RMSSD, millisecond (SD) - Baseline	43.43 (61.47)	21.12 (29.25)	0.01* (0.46)
P50, % (SD) - Baseline	19.46 (28.08)	5.69 (14.48)	<0.01* (0.62)
Poincare's SD1, millisecond (SD) - Baseline	31.95 (47.92)	16.65 (24.74)	0.03* (0.40)
Poincare's SD2, millisecond (SD) - Baseline	32.54 (38.65)	11.96 (11.55)	<0.01* (0.72)

SD: standard deviation

HR: heart rate

RR CV: coefficient of variation (SD divided by mean) of RR intervals

RMS: room-mean-square

P50: percentage of successive RR intervals with differences larger than 50 millisecond

still before and after the walking trial. Data for 5 and 10 seconds of quiet standing were used, respectively for baseline and recovery HR. HR data were collected using wearable electrocardiogram (ECG) and accelerometer sensors (360° eMotion Faros (Fig 1), Mega Electronics, Kuopio, Finland; sampling frequency = 1000Hz for ECG and 100Hz for accelerometer). For recording ECG two electrodes were used; one was located on the left side of the torso, and the other one under the rib cage on the left side (Fig. 1).

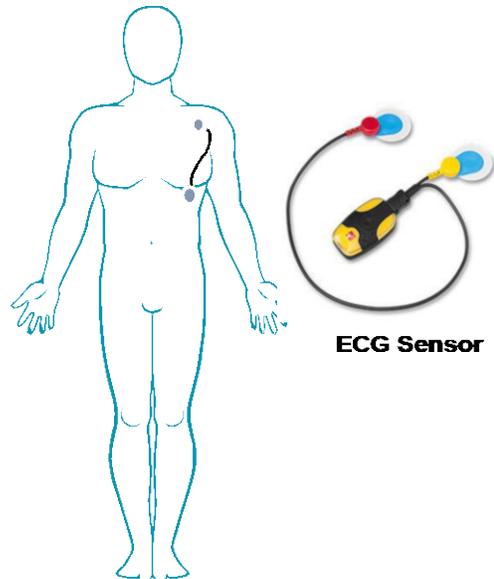


Fig. 1. Sensor's location on the body

D. Machine learning approach

The workflow of the implemented machine learning approach was as follows (Fig 2): 1) data preprocessing to create HR time series from the ECG data; 2) interpolation and resampling to make the HR data evenly sampled; 3) zero padding all HR time series to the same length; 4) data augmentation to increase and balance the number of samples in each frailty group; and 5) LSTM algorithm to predict frailty categories using the HR data and recording the performance of prediction using a five-fold cross-validation. The LSTM algorithm was implemented here because it has capability of storing and remembering information from earlier stages for a long time using long short-term memory, and this characteristic make it preferable for classification of physiological time series such as HR data [27]. For cross-validation the dataset was split up into five folds with equal number of participants in each fold. One fold was used for testing and the remaining four folds were used for training. This situation continued until all folds were used for testing. Finally, average of each performance metric was calculated from each training set and was used as the performance metric.

1) Preprocessing

For the preprocessing, first, using the synchronized ECG and accelerometer data and available examiner note, the exact starting and ending points of walking trials were selected (Fig 2). Corresponding segment of ECG data was then extracted from the ECG signals, including the period of five seconds before the walking (baseline) and 10 seconds after finishing the task (recovery). To extract RR intervals from the ECG data and create HR time series, QRS peaks were detected using the Pan-Tompkins algorithm [28]. To assure the accuracy of QRS detection, the selected peaks were manually checked and revised separately by two investigators (HS and ME). Third, ectopic beats were filtered by removing RR intervals that diverge more than 20% from the previous beat [29].

2) Machine learning approach

In the current current study logistic regression, multilayer

perceptron (MLP), and XGBoost algorithm were utilized for comparison with LSTM. In the logistic regression approach, ten extracted features were used to predict frailty, including: 1) time to peak HR: Elapsed time to reach maximum HR during the task with reference to minimum baseline HR; 2) HR recovery time: Elapsed time to reach minimum HR during the recovery with reference to maximum HR; 3) HR Percent increase: Increase in HR during the task compared to minimum baseline HR as the percentage of minimum baseline HR; 4) HR Percent decrease: Decrease in HR during the recovery compared to maximum HR during the task as the percentage of maximum HR; 5) HR mean; 6) beat-to-beat (RR) interval mean; 7) RR CV: The coefficient of variation (standard deviation divided by mean) of RR intervals; 8) RMSSD: Root mean square of successive heartbeat interval differences; 9) P50: Percentage of successive RR intervals with differences larger than 50 millisecond; and 10) Poincare's SD1 and SD2: Minor (SD1) and major (SD2) axis of the fitted ellipse to Poincare plot (plot of RR interval as a function of a previous RR interval) (based on our previous work – see Table I) [22]. In this approach, different solvers (SAGA

and LBFGS) and different regularization methods (11, 12) with stratified 5-fold cross validation strategy were implemented.

Multiple MLPs with different hidden layers have been deployed. Stochastic gradient descent (SGD), Adam optimizers (an adaptive moment estimation method based on gradient descent) were used to optimize the model and minimize the cost function. Also, hyperbolic tangent (Tanh) and rectified linear unit (ReLU) activation functions on different hidden layers (20, 15, 10) and (30, 20, 15, 10) were examined. Similar to logistic regression, stratified 5-fold cross validation strategy was utilized and the L1 regularization technique was added to the model.

For XGBoost approach, different max-depth, number of estimators, and different booster methods were applied. The number of estimators was the number of decision trees that were used within XGBoost. For the boosters, gbtrees, gblines, and dart were implemented. For the performance metrics, accuracy, sensitivity, specificity, precision, and F1-score for predicting pre-frailty/frailty (positive condition) were calculated and reported for these three algorithms.

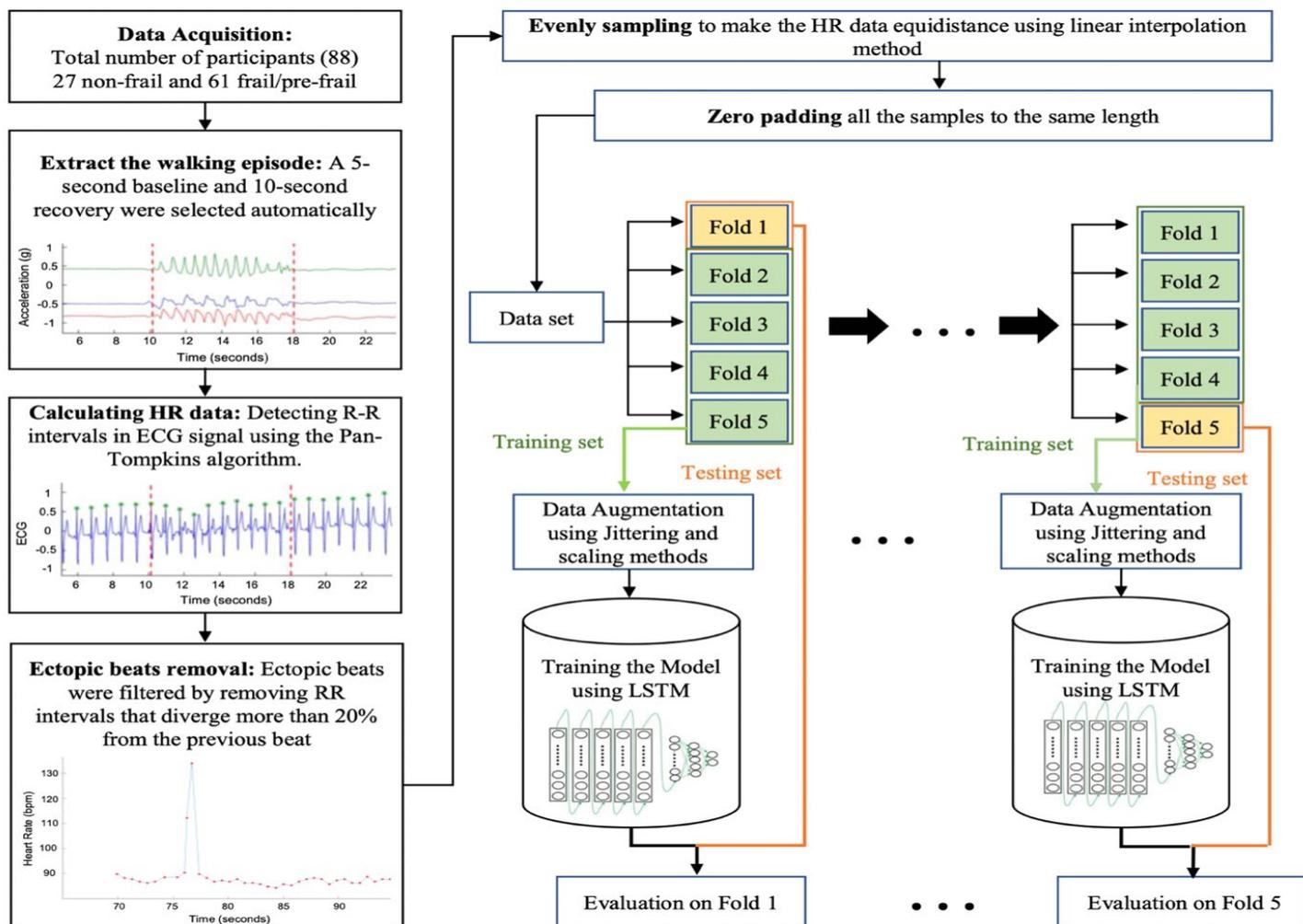


Fig. 2. Schematic representation of the study workflow

3) Interpolation, resampling, and zero-padding

HR time series is a non-uniformly sampled data as time interval between heart beats (QRS peaks) can change through recording in response to demand. LSTM model performance drops for analysis of non-uniformly sampled data. Since HR time series are unevenly sampled, the interpolation and resampling were used to create equidistant HR time series. A linear interpolation was used to form a continuous signal and resampling to three different frequencies of 3 Hz, 5 Hz, and 7 Hz was used to produce evenly sampled HR data with the length of 199, 332, and 466 samples [22]. For example, for 3 Hz setting, at every second, three data points were interpolated using linear interpolation. Of note, to assess the impact of evenly sampling on the performance of the LSTM model, both raw unevenly sampled and evenly sampled time series were used in the model development. Further, the number of samples within HR time series may differ between participants, because the walking duration were different between trials. Therefore, we implemented the “zero-padding” approach [33] by adding zeros to the end of the input sequence to provide an equal sample size for HR time series across participants. For this, the number of zeros that were added to the end of each HR data was equal to the difference between studies HR data and the longest HR data available. Of note, the longest and shortest HR vectors were 199 and 142, respectively.

4) Data augmentation and classification

In the current study we used the LSTM model, which is a variation of recurrent neural network (RNN) algorithm, best suited for analysis of time series data similar to HR [34]. To find the best LSTM model, several hyper-parameters were tuned. For tuning the algorithm structure, 10 settings with combination of different LSTM and dense layers were utilized, to explore the optimum number of layers for avoiding under- and over-fitting (Fig 2). For Training and optimization, the model hyperparameters was considered for testing, such as activation function, loss function, and learning rate. here, Sigmoid was applied as an activation function for the classification of the output layer, while, ReLU and Tanh were deployed for all other hidden layers as activation functions. Loss function was calculated using binary cross-entropy function, which is useful for binary classification tasks, and Adam optimizer was used to update weights during back-propagation. Based on the performance of the model on different sets of hidden layers, the optimum number of layers and nodes (best combination of the number of hidden layers and number of nodes in each hidden layer) was specified (Fig 3). Two classes of pre-frail/frail and non-frail labels were considered for the model.

Insufficient data may result in poor convergence of the LSTM model. Furthermore, imbalance dataset (smaller sample of non-frail compared to pre-frail/frail) can lead to a biased classification result [35]. Therefore, data augmentation was performed in this study to increase training sample size and provide balanced data. In this study, a combination of scaling and jittering methods [36] was applied, to create new HR training data. For scaling a random scalar value was multiplied to the vector of HR time series. Random value was returned from a normal distribution with a mean of one and standard deviations (SDs) of 0.05, 0.1, or 0.2 [37][14]. In the second step, Gaussian noise was multiplied to the scaled data set, including a vector of normal distribution with a zero mean and

SDs of 0.01, 0.05, or 0.1 [37][14]. For tuning the scaling and the Gaussian noise values, nine setting based on the different SDs in scaling and jittering were implemented and among them the best setting of scaling with SD of 0.2 and jittering with SD of 0.1 was selected. Using the above approach, the number of data were increased from 88 (before data augmentation) to 455 (after data augmentation, including 224 non-frail and 231 pre-frail/frail).

For model evaluation, five-fold cross-validation approach was used. LSTM model was trained on training dataset (four folds) and F1-score, accuracy, AUC, sensitivity, specificity, and

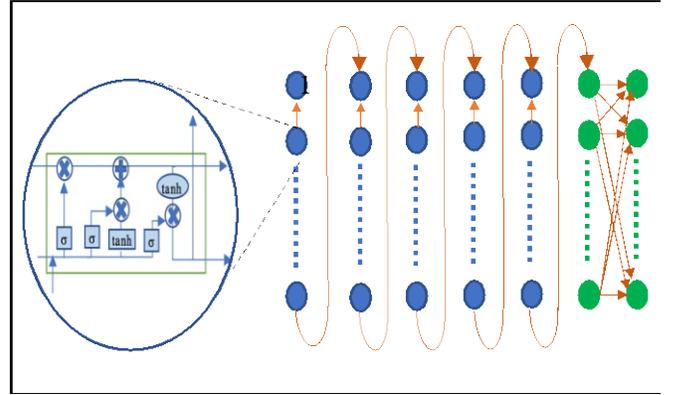


Fig. 1. The LSTM network with five LSTM layers and two fully-connected layers

precision were reported for the testing sets (one-fold). To study impact of data augmentation, model training was done on training data without and with data augmentation and performance was assessed only on original testing dataset (one assigned fold for testing without any data augmentation).

III. RESULT

A. Participants

As discussed in section 2, eighty-eight participants were recruited, among whom, 27 were non-frail and 61 were pre-frail/frail (age=80.63±8.07). Twelve pre-frail/frail participants out of 61 and six non-frail out of 27 participants were used for each fold. One fold was used as testing and the rest of data were used as training. Training sets were balanced using the data augmentation process. Of note, the data augmented were only used in the training sets (four folds in cross-validation approach). Because the number of training pre-frail/frail data was almost twice as training non-frail data (49 frail/pre-frail data compared to 21 non-frail data), for each pre-frail/frail data five new HR data and for each non-frail data 10 new HR data were produced to balance the data (total of 210 non-frail vs 245 frail/pre-frail training dataset for each fold). Using this approach, the number of training data was increased by a factor of 7.5, which was required for providing reliable training sample for the LSTM approach [37].

B. Frailty assessment using Machine learning

Result of logistic regression models based on multiple settings with different regularization methods and solver options are presented in Table II. The comparison between different

settings showed that saga as a solver without any regularization method for regularization provided the best result (accuracy and F1-score of 70.0% and 79.0%, Table II). Based on the precision values (between 29.0% to 48.0%), the results were biased toward pre-frailty/frailty class.

For MLP, maximum iteration was set on 500 and regularization was set on 0.001 as optimum values for the model. Using these settings and applying MLP, an accuracy of 73.0% and F-score of 82.0% were achieved (Table III). The best result was achieved using SGD as an optimizer, ReLU as an activation function, with three hidden layers of (20,15, 10) units (Table III).

Finally, for XGBoost multiple settings have been applied based on different maximum depth, number of estimators, and booster (gbtree, gblinear, and dart) (Table IV). The best result was obtained from maximum depth=4, number of estimator=200, gbtree as the booster, and L2 regularization, providing an accuracy of 72.0% and F-score of 82.0%.

C. Frailty assessment using LSTM

LSTM model, with evenly sampled HR time series (7 Hz) and data augmentation steps, showed classification results of 87.0%, 82.0%, and 0.87, respectively, for F1-score, accuracy, and AUC.

TABLE V

A) AVERAGE OF PERFORMANCE METRICS IN 5-FOLD CROSS VALIDATION USING LSTM WITH UNEVENLY SAMPLED DATA AND EVENLY SAMPLED DATA WITH LINEAR INTERPOLATION AND RESAMPLING FREQUENCIES OF 3, 5, 7 AS AN INPUT. HIGHEST PERFORMANCE METRIC IN EACH COLUMN IS HIGHLIGHTED IN BOLD

LSTM (55,50,40,30,20) +Dense (10,5) (Frequency)	Ave- F1-score (%)	Ave-accuracy (%)	Ave-sensitivity (%)	Ave-precision (%)	Ave- specificity (%)	Ave-AUC
1 Without Evenly Sampling	82.6	78.8	79.2	89.6	77.0	0.83
1 LSTM With Data Augmentation (3)	79.0	69.0	80.4	74.8	55.2	0.72
1 LSTM With Data Augmentation (5)	75.0	67.6	77.8	92.0	83.8	0.79
1 LSTM With Data Augmentation (7)	87.0	82.0	83.0	91.0	80.0	0.87
1 LSTM (number of units in each hidden layer)						
= + Dense (number of units in each hidden layer)						
or evenly sampled data with frequency 7						
1 LSTM(50,25)+Dense(10)	80.0	71.4	81.0	77.0	43.6	0.64
1 LSTM(50,25)+Dense(10,5)	80.4	76.2	66.8	94.8	94.0	0.88
1 LSTM(50,35,25)+Dense(10)	82.2	71.2	88.8	76.0	42.2	0.69
1 LSTM(50,35,25)+Dense(10,5)	86.2	79.8	96.2	80.0	52.4	0.71
1 LSTM(55,50,40,30,20)+Dense(10)	81.8	74.0	85.4	79.6	49.0	0.64
1 LSTM(55,50,40,30,20)+Dense(10,5)	87.0	82.0	83.0	91.0	80.0	0.87
1 LSTM(55,55,35,25,10)+Dense(10)	82.0	74.0	91.6	76.4	44.0	0.68
1 LSTM(55,55,35,25,10)+Dense(5)	84.8	79.2	85.0	84.8	58.4	0.77
1 LSTM(100,80,65,50,40,30,20)+Dense(5)	82.4	76.6	86.2	82.0	54.6	0.85
1 LSTM(100,80,65,50,40,30,20)+Dense(10,5)	82.4	67.8	88.0	70.6	38.6	0.62

TABLE IV

AVERAGE OF PERFORMANCE METRICS OF XGBOOST WITH DIFFERENT MAX NUMBER OF DEPTH, NUMBER OF ESTIMATORS, AND DIFFERENT BOOSTERS. HIGHEST PERFORMANCE METRIC IN EACH COLUMN IS HIGHLIGHTED IN BOLD

XGBoost (Max_depth, N_estimators, Booster)	Ave- F1-score (%)	Ave-accuracy (%)	Ave-sensitivity (%)	Ave-precision (%)	Ave- specificity (%)	Ave-AUC
XGB (4, 200, gbtree)	77.0	69.0	79.0	77.0	45.0	0.62
XGB (4, 300, gbtree)	72.0	63.0	71.0	74.0	44.0	0.57
XGB (6, 200, gbtree)	72.0	60.0	73.0	71.0	34.0	0.53
XGB (4, 200, gblinear)	78.0	68.0	85.0	74.0	31.0	0.58
XGB (4, 300, gblinear)	82.0	72.0	92.0	74.0	28.0	0.60
XGB (6, 200, gblinear)	81.0	59.0	90.0	74.0	29.0	0.59
XGB (4, 200, dart)	73.0	63.0	74.0	72.0	38.0	0.56
XGB (4, 300, dart)	69.0	60.0	66.0	74.0	49.0	0.58
XGB (6, 200, dart)	75.0	65.0	76.0	74.0	40.0	0.58

and AUC in detecting pre-frailty/frailty within the testing dataset; sensitivity, specificity, and precision were 83.0%, 80.0%, and 91.0%, respectively for predicting frail/pre-frail as the positive condition. Based on the current results, the best setting (maximum F1-score) was achieved using evenly sampled HR data (7 Hz), along with an architecture with five LSTM layers and two dense layers. The number of units for this architecture were 55, 50, 40, 30, and 20 for five LSTM layers and 10 and 5 for dense layers, respectively (see Table V for

IV. DISCUSSION

A. Frailty and HR dynamics

As hypothesized, we observed that solely based on characteristics of HR response and recovery to walking we were able to predict pre-frailty/frailty with an accuracy of 82.0% and F1-score of 87.0%. The observed strong association between HR behavior and frailty suggests that HR monitoring during physical activity, may be considered as a robust measure of frailty, and an additional measure of physiological reserve to, in combination with motor function assessment, improve frailty prediction. One cause of frailty is the homeostenosis phenomenon, which is the reduced ability to maintain homeostasis or physiological balance under stress [6] and ANS helps in maintaining homeostasis by regulating physical activities [10][11][12]. An impaired ANS would be incapable of maintaining hemodynamic homeostasis under stress from the physical activity of walking [13][14] and HR dynamics can be used for assessment of sympathetic and parasympathetic abnormalities during this physical activity [38]; here, the sympathetic (HR increase during activity) and parasympathetic (HR decline in the recovery phase) activities were directly included in HR dynamics measures used for LSTM modeling. In confirmation to this hypothesis, our previous findings showed significant associations between HRV baseline measures during resting and HR dynamic parameters [22].

B. HR data classification using LSTM model

In confirmation of our secondary hypothesis, our results suggest that the implemented deep learning approach can outperform shallow learning models for frailty identification. Machine learning results shows that the best performance was achieved using MLP with SGD as a solver, ReLU as an activation function and with 3 hidden layers with 20, 15, and 10 hidden units in each layer respectively. By comparing the result of MLP with the best result of LSTM with 5 LSTM layer with 55,50,40,30, and 20 hidden units and 2 Dense layers with 10 and 5 hidden units we can see the F1-score improved 5%, the accuracy improved 10%, precision improved 16%, and surprisingly, the specificity improved from 32% to 80%. The LSTM improved the specificity. The observed improvement in the model prediction here may be due to the fact that shallow learning approaches are useful for simple and more linear classification rather than complex nonlinear ones [39]. LSTM model, on the other hand, can learn more complex and nonlinear functions, and also provides a memory cell that is capable of storing and learning long range dependencies, which may be the case for HR response to physical activities [40].

Within the current sample, data augmentation improved the validation accuracy and F1-score in predicting frailty by 3.2%

details).

Further, interpolation and resampling approach enhanced the performance by 4.4% for F1-score, 3.2% for accuracy, and 4% for AUC compared to using unevenly sampled HR data as input to the LSTM model. The data augmentation approach affected the performance of the model by improving the F1score by 3.4%, accuracy by 3.2%, and AUC by 2.4%.

and 3.4%, respectively. This is because the performance of LSTM approach relies heavily on the training data sample size [37]; synthetic training data used for increasing sample size and balancing frailty groups in the training dataset here eliminated the bias in the trained model due to unbalanced number of two frailty groups in the original training data. Further, data augmentation technique improved the frailty prediction here by reducing the under-fitting problem [41][14].

We also observed that in the LSTM model, increasing the number of hidden layers from three to seven layers enhanced the performance of the model. But increasing the number to more than seven resulted in a reduced prediction accuracy, which may happen because of overfitting [42]. This finding was also corroborated with the Ye et al. [43] results, which found that as the LSTM model get deeper (increasing the number of hidden layers), the chance of overfitting in the model increases.

C. Limitation and future direction

In this study, 88 participants were included, among which less than one third were non-frail older adults. Although the small number of participants and unbalanced frailty group were, to some extent, remedied with augmentation, the sample size limitation still exists. Therefore, current findings need to be confirmed in future research and we expect a better performance of LSTM model in a larger sample of participants. Similar to Park et al. and Arshad et al. [44][18], we merged pre-frail and frail groups into one single group due to the limited number of frail participants within our cohort. This is an additional limitation of the current study as the trajectory of frailty progression was not investigated here. Specially, identifying frailty in early stages (i.e., pre-frailty) can enhance the rehabilitation planning, as frailty become less reversible in later stages.

The second limitation of the current study was related to unequal length of HR time series data. Within the current protocol we asked participants to walk for a certain distance, instead of a specific period. Since LSTM requires data of the same length for input, we used padding techniques to provide equal data sample for the analysis. Nevertheless, padding technique may affect the LSTM networks function, and subsequently impact the performance and accuracy of the model [45]. Therefore, it would be interesting to explore the accuracy of the proposed approach in future studies by collecting HR data for equal duration of walking. We used preprocessing technique (interpolation and resampling) to convert non-uniformly sampled HR data to uniformly sampled HR. Utilizing enhanced LSTM architecture [46] and the associated pros/cons needs to be explored in future studies. LSTM model with HR time series as input led to promising performance without any feature engineering that is required in traditional feature-based machine

learning methods. However, identifying characteristic in HR time series that was important to differentiate between frailty groups was not possible within the LSTM approach. Also, deploying the proposed algorithm in future on bigger datasets and other experimental settings (different datasets) will help in external validation of our approach, especially in comparison to other competing methods.

V. CONCLUSION

In this study, a deep learning model with HR time series as input was used for frailty identification. Using the LSTM model, participants were classified into their associated groups with a F1-score of 87.0% and an accuracy of 82.0%. Our results also demonstrated that data augmentation enhanced the model accuracy and F1-score by 3.2% and 3.4%, respectively. Further, current results showed the importance of converting unevenly sampled HR to evenly sampled data to improve model prediction (F1-score and accuracy by 4.4% and 4.0%, respectively). Comparing the LSTM with other machine learning approaches (i.e., logistic regression, MLP, and XGBoost) showed that LSTM outperformed other approaches by improving the F1-score and accuracy 5.0% and 9.0% on average.

REFERENCES

- [1] F. Buckinx, Y. Rolland, J. Y. Reginster, C. Ricour, J. Petermans, and O. Bruyère, "Burden of frailty in the elderly population: Perspectives for a public health challenge," *Archives of Public Health*, vol. 73, no. 1. BioMed Central Ltd., pp. 1–7, Apr. 10, 2015, doi: 10.1186/s13690-015-0068-x.
- [2] C. M. Boyd, Q. L. Xue, C. F. Simpson, J. M. Guralnik, and L. P. Fried, "Frailty, hospitalization, and progression of disability in a cohort of disabled older women," *Am. J. Med.*, vol. 118, no. 11, pp. 1225–1231, Nov. 2005, doi: 10.1016/j.amjmed.2005.01.062.
- [3] J. Sacha, M. Sacha, J. Sobon, Z. Borysiuk, and P. Feusette, "Is it time to begin a public campaign concerning frailty and pre-frailty? A review article," *Frontiers in Physiology*, vol. 8, no. JUL. Frontiers Media S.A., p. 484, Jul. 11, 2017, doi: 10.3389/fphys.2017.00484.
- [4] E. Dent, P. Kowal, and E. O. Hoogendijk, "Frailty measurement in research and clinical practice: A review," *European Journal of Internal Medicine*, vol. 31. Elsevier B.V., pp. 3–10, Jun. 01, 2016, doi: 10.1016/j.ejim.2016.03.007.
- [5] C. Welch *et al.*, "Delirium is prevalent in older hospital inpatients and associated with adverse outcomes: Results of a prospective multi-centre study on World Delirium Awareness Day," *BMC Med.*, vol. 17, no. 1, p. 229, Dec. 2019, doi: 10.1186/s12916-019-1458-7.
- [6] J. W. Shega, W. Dale, M. Andrew, J. Paice, K. Rockwood, and D. K. Weiner, "Persistent pain and frailty: A case for homeostenosis," *J. Am. Geriatr. Soc.*, vol. 60, no. 1, pp. 113–117, Jan. 2012, doi: 10.1111/j.1532-5415.2011.03769.x.
- [7] C. O. Weiss, H. H. Hoening, R. Varadhan, E. M. Simonsick, and L. P. Fried, "Relationships of cardiac, pulmonary, and muscle reserves and frailty to exercise capacity in older women," *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.*, vol. 65 A, no. 3, pp. 287–294, Mar. 2010, doi: 10.1093/gerona/glp147.
- [8] R. Romero-Ortuno, L. Cogan, D. O'shea, B. A. Lawlor, and R. A. Kenny, "Orthostatic haemodynamics may be impaired in frailty," *Age Ageing*, vol. 40, no. 5, pp. 576–583, Sep. 2011, doi: 10.1093/ageing/afr076.
- [9] S. Damanti, P. D. Rossi, and M. Cesari, "Heart rate variability: a possible marker of resilience," *European Geriatric Medicine*, vol. 10, no. 3. Springer International Publishing, pp. 529–530, Jun. 01, 2019, doi: 10.1007/s41999-019-00192-2.
- [10] R. Gordan, J. K. Gwathmey, and L.-H. Xie, "Autonomic and endocrine control of cardiovascular function," *World J. Cardiol.*, vol. 7, no. 4, p. 204, 2015, doi: 10.4330/wjc.v7.i4.204.
- [11] S. Parvaneh *et al.*, "Regulation of Cardiac Autonomic Nervous System Control across Frailty Statuses: A Systematic Review," *Gerontology*, vol. 62, no. 1, pp. 3–15, Jul. 2015, doi: 10.1159/000431285.
- [12] "Role of autonomic nervous system and hemodynamics in cardiovascular homeostasis after orthostatic stress - PubMed." <https://pubmed.ncbi.nlm.nih.gov/11220120/> (accessed Dec. 06, 2020).
- [13] F. Mourey, L. Brondel, V. Van Wymelbeke, M. Buchheit, D. Moreau, and P. Pfitzenmeyer, "Assessment of cardiac autonomic nervous activity in frail elderly people with postural abnormalities and in control subjects," *Arch. Gerontol. Geriatr.*, vol. 48, no. 1, pp. 121–124, Jan. 2009, doi: 10.1016/j.archger.2007.11.004.
- [14] K. M. Rashid and J. Louis, "Times-series data augmentation and deep learning for construction equipment activity recognition," *Adv. Eng. Informatics*, vol. 42, p. 100944, Oct. 2019, doi: 10.1016/j.aei.2019.100944.
- [15] Z. He, "The control mechanisms of heart rate dynamics in a new heart rate nonlinear time series model," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-61562-6.
- [16] M. Moghtadaei *et al.*, "The impacts of age and frailty on heart rate and sinoatrial node function," *J. Physiol.*, vol. 594, no. 23, pp. 7105–7126, Dec. 2016, doi: 10.1113/JP272979.
- [17] K. M. Flint, D. D. Matlock, J. A. Lindendorf, and L. A. Allen, "Frailty and the selection of patients for destination therapy left ventricular assist device," *Circ. Hear. Fail.*, vol. 5, no. 2, pp. 286–293, Mar. 2012, doi: 10.1161/CIRCHEARTFAILURE.111.963215.
- [18] "Gait-based Frailty Assessment using Image Representation of IMU Signals and Deep CNN - Google Search." <https://www.google.com/search?q=Gait-based+Frailty+Assessment+using+Image+Representation+of+IMU+Signals+and+Deep+CNN&oeq=Gait-based+Frailty+Assessment+using+Image+Representation+of+IMU+Signals+and+Deep+CNN&q=chrome.0.69i59j69i6112.631j0j4&sourceid=chrome&ie=UTF-8> (accessed Jan. 05, 2022).
- [19] S. Aponte-Hao *et al.*, "Machine learning for identification of frailty in Canadian primary care practices," *Int. J. Popul. data Sci.*, vol. 6, no. 1, Jan. 2021, doi: 10.23889/IJPDS.V6I1.1650.
- [20] B. Friedrich, S. Lau, L. Elgert, J. M. Bauer, and A. Hein, "A deep learning approach for tug and sppb score prediction of (Pre-) frail older adults on real-life imu data," *Healthc.*, vol. 9, no. 2, Feb. 2021, doi: 10.3390/HEALTHCARE9020149.
- [21] P. H. M. Chaves *et al.*, "Physiological Complexity Underlying Heart Rate Dynamics and Frailty Status in Community-Dwelling Older Women," *J. Am. Geriatr. Soc.*, vol. 56, no. 9, pp. 1698–1703, Sep. 2008, doi: 10.1111/j.1532-5415.2008.01858.x.
- [22] N. Toosizadeh *et al.*, "Frailty and heart response to physical activity," *Arch. Gerontol. Geriatr.*, vol. 93, p. 104323, Mar. 2021, doi: 10.1016/j.archger.2020.104323.
- [23] L. Wang and X. Zhou, "Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals," *Sensors (Basel)*, vol. 19, no. 7, Apr. 2019, doi: 10.3390/S19071502.
- [24] M. Radha *et al.*, "Sleep stage classification from heart-rate variability using long short-term memory neural networks," *Sci. Reports 2019 91*, vol. 9, no. 1, pp. 1–11, Oct. 2019, doi: 10.1038/s41598-019-49703-y.
- [25] S. Parvaneh and J. Rubin, "Electrocardiogram Monitoring and

- Interpretation: From Traditional Machine Learning to Deep Learning, and Their Combination,” doi: 10.22489/CinC.2018.144.
- [26] S. Idris and N. Badruddin, “Classification of Cognitive Frailty in Elderly People from Blood Samples using Machine Learning,” pp. 1–4, Aug. 2021, doi: 10.1109/BHI50953.2021.9508514.
- [27] M. Thirugnanam and M. S. Pasupuleti, “Cardiomyopathy -induced arrhythmia classification and pre-fall alert generation using Convolutional Neural Network and Long Short-Term Memory model,” *Evol. Intell.*, vol. 1, p. 3, Jul. 2020, doi: 10.1007/s12065-020-00454-0.
- [28] “Mini-mental state’. A practical method for grading the cognitive state of patients for the clinician - Google Search.” <https://www.google.com/search?q=Mini-mental+state'+A+practical+method+for+grading+the+cognitive+state+of+patients+for+the+clinician&oq=Mini-mental+state'+A+practical+method+for+grading+the+cognitive+state+of+patients+for+the+clinician&aqs=chrome..69i57.386j0j9&sourceid=chrome&ie=UTF-8> (accessed Jan. 05, 2022).
- [29] “World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects,” *JAMA - Journal of the American Medical Association*, vol. 310, no. 20, JAMA, pp. 2191–2194, 2013, doi: 10.1001/jama.2013.281053.
- [30] L. P. Fried *et al.*, “Frailty in older adults: evidence for a phenotype,” *J. Gerontol. A. Biol. Sci. Med. Sci.*, vol. 56, no. 3, 2001, doi: 10.1093/GERONA/56.3.M146.
- [31] R. A. Fieo, E. L. Mortensen, T. Rantanen, and K. Avlund, “Improving a measure of mobility-related fatigue (the mobility-tiredness scale) by establishing item intensity,” *J. Am. Geriatr. Soc.*, vol. 61, no. 3, pp. 429–433, Mar. 2013, doi: 10.1111/JGS.12122.
- [32] S. Sze, P. Pellicori, J. Zhang, J. Weston, and A. L. Clark, “Identification of Frailty in Chronic Heart Failure,” *JACC. Heart Fail.*, vol. 7, no. 4, pp. 291–302, Apr. 2019, doi: 10.1016/J.JCHF.2018.11.017.
- [33] J. Werth, M. Radha, P. Andriessen, R. M. Aarts, and X. Long, “Deep learning approach for ECG-based automatic sleep state classification in preterm infants,” *Biomed. Signal Process. Control*, vol. 56, p. 101663, Feb. 2020, doi: 10.1016/j.bspc.2019.101663.
- [34] R. DiPietro and G. D. Hager, “Deep learning: RNNs and LSTM,” in *Handbook of Medical Image Computing and Computer Assisted Intervention*, Elsevier, 2019, pp. 503–519.
- [35] P. Cao *et al.*, “A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation,” *Biomed. Signal Process. Control*, vol. 56, p. 101675, 2020, doi: 10.1016/j.bspc.2019.101675.
- [36] H. Lee and M. Whang, “Heart rate estimated from body movements at six degrees of freedom by convolutional neural networks,” *Sensors (Switzerland)*, vol. 18, no. 5, May 2018, doi: 10.3390/s18051392.
- [37] Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, and H. Xu, “Time Series Data Augmentation for Deep Learning: A Survey,” *arXiv*, Feb. 2020, Accessed: Nov. 21, 2020. [Online]. Available: <http://arxiv.org/abs/2002.12478>.
- [38] E. O. Nishime, C. R. Cole, E. H. Blackstone, F. J. Pashkow, and M. S. Lauer, “Heart rate recovery and treadmill exercise score as predictors of mortality in patients referred for exercise ECG,” *J. Am. Med. Assoc.*, vol. 284, no. 11, pp. 1392–1398, Sep. 2000, doi: 10.1001/jama.284.11.1392.
- [39] A. Allam, M. Nagy, G. Thoma, and M. Krauthammer, “Neural networks versus Logistic regression for 30 days all-cause readmission prediction,” *arXiv*. arXiv, Dec. 22, 2018, Accessed: Dec. 26, 2020. [Online]. Available: <https://arxiv.org/abs/1812.09549v1>.
- [40] S. Guney and C. B. Erdas, “A deep LSTM approach for activity recognition,” in *2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019*, Jul. 2019, pp. 294–297, doi: 10.1109/TSP.2019.8768815.
- [41] J. Nalepa, M. Myller, and M. Kawulok, “Hyperspectral Data Augmentation,” Mar. 2019, Accessed: Nov. 21, 2020. [Online]. Available: <http://arxiv.org/abs/1903.05580>.
- [42] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning.”
- [43] J. Zhu, H. Chen, and W. Ye, “A Hybrid CNN-LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar,” *IEEE Access*, vol. 8, pp. 24713–24720, 2020, doi: 10.1109/ACCESS.2020.2971064.
- [44] C. Park *et al.*, “Digital biomarker representing frailty phenotypes: The use of machine learning and sensor-based sit-to-stand test,” *Sensors*, vol. 21, no. 9, May 2021, doi: 10.3390/S21093258.
- [45] D. M. Reddy and S. Reddy, “EFFECTS OF PADDING ON LSTMS AND CNNs.”
- [46] S. O. Sahin and S. S. Kozat, “Nonuniformly Sampled Data Processing Using LSTM Networks,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 30, no. 5, pp. 1452–1461, May 2019, doi: 10.1109/TNNLS.2018.2869822.