[Team 17]Proj-C: Terrain Type Classification using Time Series Data

Pramodini Karwande
M.S. in Statistics
North Carolina State University
Durhum, North Carolina
pkarwan@ncsu.edu

Md Farhamdur Reza
PhD in Electrical Engineering
North Carolina State University
Raleigh, North Carolina
mreza2@ncsu.edu

Rageeni Sah
M.S. in Electrical Engineering
North Carolina State University
Raleigh, North Carolina
rsah@ncsu.edu

***Index Terms*—Convolution Neural Network, LSTM, Terrain Classifier, Imbalance Classes**

I. METHODOLOGY

Terrain Identification is a typical classification task to find different terrains from time series data. This project is done using accelerometer data and gyroscope data obtained from different subjects walking on four types of terrain and it is inspired from [1]. The idea for this project is to classify terrain correctly to automate robotic prosthetic using machine learning models. The data-set provided has two major challenges. The first challenge is the mismatch of the frequency of the data from accelerometer and gyroscope with the frequency of the assigned class level data. The second challenge is that the target variable consists of four imbalanced classes as shown in Fig. 1. Class 1,2, & 3 are relatively minority classes and class 0 is majority class. It is found that sampling rate of the accelerometer and gyroscope data are 40 hertz and the sampling rate of the assigned class level is 10 hertz. To sync sampling rate either up-sampling of the measured data or down-sampling of the target is a must step and we tried both the options to obtain better performance.

In the first phase of the submission, the approach followed was to down-sampled the data as 4:1 by aggregating 4 consecutive readings of the accelerometer and gyroscope to a single value. With trying multiple estimators (average, median and random samplings) to aggregate values, average estimator performance was relatively best of the three techniques that are applied through the features to down-sampled to 10 hertz. In the subsequent submissions, deep learning architectures are explored to improve baseline performance. Training a fully connected neural network using averaging technique performed better than other two techniques, and outperforms baseline Random Forest model performance.

As a part of further improvement plan, we used one dimensional convolution neural network model with upsampling the target variable. CNN-1D is an effective technique to handle human activity recognition (HAR) data [2]. By analysing the measured data, it is observed that the average of the walking cycle on different terrain is about 1.2 seconds. So, a preprocessing of data is done using the walking cycle. To train the CNN-1D model, the approach used is to stack 48 consecutive rows along with all six features making a 48×6 feature matrix and the target variable is considered as the target corresponding to last row of the matrix. The f1 score with this approach is 0.78 on validation set. Further improvement was achieved using Long Short Term Memory (LSTM) model when trained with same preprocessed dataset. In addition, class weights are estimated to handle the imbalanced classes. The f1 score using LSTM model is 0.80 on validation set.

Two dimensional convolution neural network (CNN2D) are more effective in prediction on HAR data as CNN2D

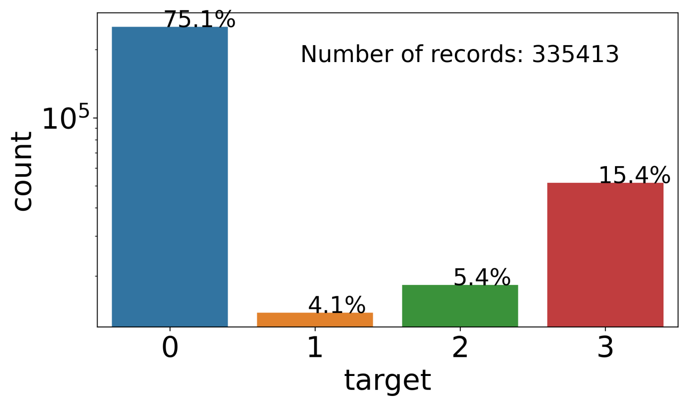

Fig. 1: Target Class Distribution: (0) indicates standing or walking in solid ground, (1) indicates going down the stairs, (2) indicates going up the stairs, and (3) indicates walking on grass.

TABLE I: Number of sessions participated by each subject

Subject ID	Number of sessions
001	08
002	05
003	03
004	01
005	03
006	03
007	04
008	01

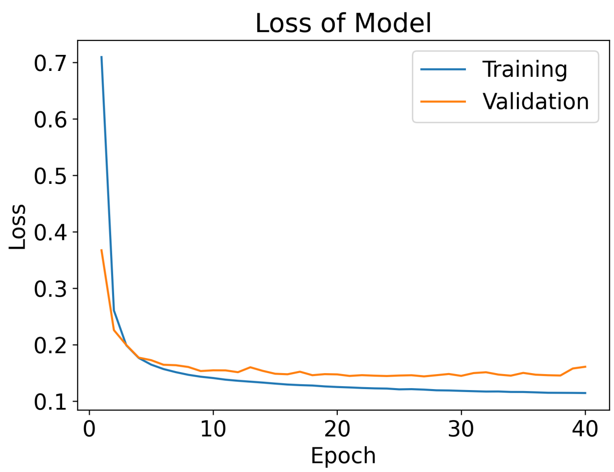

Fig. 2: Training and Validation Loss

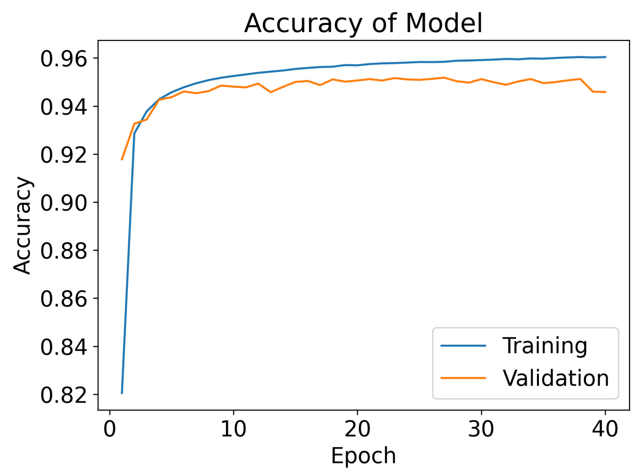

Fig. 3: Training and Validation Accuracy

captures both temporal information and interraction effects between features [3]. Following the same, in the final attempt, CNN2D model is trained on given data resulting to the best performance on validation set. To train this model, further preprocessing of the data is done. We have reshaped the data used in CNN-1D as $1 \times 48 \times 6$ tensor before feeding to the model.

II. MODEL TRAINING AND SELECTION

The training data-set consists of readings of accelerometer and gyroscope of eight subjects where each of the subjects has data collected across multiple sessions as given in table I. A predictive classifier model is built using 2-dimensional convolution neural networks which is also called popularly as CNN2D. The training data is split into several 2D numpy array of size 48X6 and transformed to 1X48X6 tensors. Google Colab is used for training models.

A. Model Training

The total data is split into training and validation sets in the ratio of 7:3. Accelerometer and gyroscope readings are split into 1X48X6 tensor objects. A number of experiments are conducted using CNN2D. The architecture of the best model is shown in Fig. 4. There are total four hidden layers: first two are convolution layers (pooling layers are not used) and next two layers are fully-connected layers. The features are batch normalized. Adam with default parameters is used as optimizer. Model is run for total 40 epoch with 5000 images in each of the batch.
The best model f1-score is 96.85% for class 0, 93.07% for class 1, 92.78% for class 2 and 88.16% for class 3 on the validation set. Overall validation accuracy is 95% and class wise accuracy is as follows: class-0: 97%, class-1: 91%, class-2: 93%, & class-3: 85%. The training and validation loss curve of the network is shown in Fig. 2 and the accuracy curve is as shown in Fig. 3.

B. Model Selection

A model with higest F1-score on validation set is considered the best. Starting experiments using classical predictive models such as BaggingClassifier, GradientBoostingClassifier, DecisionTreeClassifier and, RandomForestClassifier from python Sci-kit learn library followed by convolution neural networks (1D and 2D) and LSTM, the best validation f1-score and accuracy is achieved using 2-dimensional convolution neural network (CNN2D). There are drastic improvement in prediction accuracy using deep learning networks compared to classical models. Table II compares f1-score of the models experimented in the project. LSTM models validation score are highest however, they did not perform well on test datasets. The reason is not quite clear at this stage, however, it is possibly over-fitting.

III. EVALUATION

The predictive model performance was evaluated using validation set which was 30% of the total available data. There are total 335413 records. F1-score is used as the metric to evaluate model performance. A model with optimal training and validation f1-score is decided to be optimal model.The f1 - score of the final model is 96.85% for class 0, 93.07% for class 1, 92.78% for class 2 and 88.16% for class 3 on the validation set. Table III shows accuracy comparison for experiments conducted and flags out experiments where minority class handling were attempted, although the results were not effective using current class weight balancing techniques.

IV. FUTURE IMPROVEMENT SCOPE

The current model performance can be enhanced further by using advanced techniques to handle imbalanced classes and applying regularization to avoid model over-fitting along with exploring wider parameter space. Using techniques to enhance features information are also options to uplift current performance. Fast Fourier Transformation (FFT) can be applied on features matrices. Due to lack of high speed infrastructure a

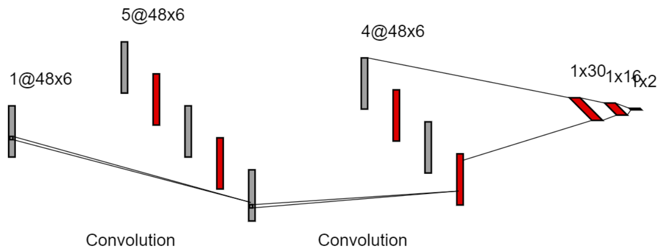

Fig. 4: Best Model Architecture - CNN2D. Four hidden layers (2 CNN layers + 2 FC layers). Pooling is not used in the architecture. FC1 has 30 nodes and FC2 has 16 nodes. Input layer is 1 X 48 X 6 torch object.

TABLE II: F1-Score on Validation Set; (*) epoch, (+) learning rate = 0.01, remaining models default learning rates.

Models	f1-score: 0	f1-score: 1	f1-score: 2	f1-score: 3
Random Forest	0.78	0.43	0.65	0.43
CNN-1D	0.98	0.97	0.95	0.92
LSTM (1OO*)	0.96	0.90	0.89	0.87
LSTM (2OO*)	0.99	0.98	0.97	0.97
LSTM (1OO*+)	0.99	0.97	0.97	0.97
CNN-2D (40*)	0.97	0.93	0.93	0.88

lot of the parameters space are not fully explored. Using high speed infracture and training models for higer epoch sides can also help to explore global minimum which seems to be missed in current study. It is observed imbalanced classes handling improves overall performance. Proposed CNN2D with class weight balancing can be studied.

TABLE III: Overall Validation Accuracy; (*) epoch, (+) learning rate = 0.01, remaining models default learning rates. Class Weight Balancing - Upsampling of minority classes.

Models	Validation Accuracy (%)	Class Weight Balancing
Random Forest	55	Y
CNN-1D	97	N
LSTM (1OO*)	93	Y
LSTM (2OO*)	94	Y
LSTM (1OO*+)	98	Y
CNN-2D (40*)	95	N

References

[1] B. Zhong, R. L. d. Silva, M. Li, H. Huang and E. Lobaton, "Environmental Context Prediction for Lower Limb Prostheses With Uncertainty Quantification," *in IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 458-470, April 2021, doi: 10.1109/TASE.2020.2993399.

[2] D. Li, J. Zhang, and X. Wei, "Classification of ECG signals based on 1D Convolution Neural Network," *19th International Conference on* $e_{H}ealthNetworking, ApplicationsandServices$, 2017.

[3] M.Gholamrezaii, S. AlModarresi, "A time-efficient convolutional neural network model in human activity recognition," *Multimed Tools Appl*, vol. 80, pp. 19361–19376, 2021, https://doi.org/10.1007/s11042-020-10435-1.