IoT Assignment 2

Human Activity Recognition (HAR) through wearable sensors with deep CNN-LSTM with a self-attention model

---

Name: Ankit Sharma
Roll No.: 2020CSB1072

Abstract:

Human Activity Recognition (HAR) is a field of study that aims to identify and classify human movements or actions based on sensor data. HAR has many applications in domains such as health care, sports, entertainment, and smart homes. One of the familiar sources of sensor data for HAR is smartphones, which have built-in inertial sensors such as accelerometers, gyroscopes, and linear accelerometers.

These sensors can capture the 3-axial linear acceleration and angular velocity of the smartphone at a fixed sampling rate. In the paper, authors have used a "sensor data collector" android application to collect data and create their own dataset, namely "H Activity." The proposed model works exceptionally well on H-Activity, i.e., with 99.93% accuracy. Authors continue to benchmark the proposed model for other publicly available datasets like MHEALTH (98.76%) and UCI-HAR(93.11%).

My goal is to replicate the results for the MHealth dataset, which is publicly available since H-Activity is not accessible to the public. The MHealth dataset consists of data from 2 sensors namely attached to the left ankle and right-lower arm. Both sensors record triaxial linear acceleration and gyro. MHealth also consists of more categories than H-Activity for classification (12 compared to 4).

Through various experimentations, I was able to achieve exceptional accuracy of 99.69% for the MHealth dataset for 12-class classification.

Introduction:

Importance of HAR: HAR is important for fitness, security, etc. Be it an athlete who wants to keep track of his/her calories, he/she may get a wearable which, based on various parameters, predicts the activity and approximate calorie consumption, or it can be an elderly where his/her relatives want to keep track of his movements like stumbling or falling. IoT enables a variety of integrations for a wide range of use cases. Human activity recognition and classification, among other things, aid in the identification of neurological problem patients and hemiparetic patients and the examination of sports-person activity patterns.

Novelty in the solution: Deep learning models can learn complex features from raw sensor data without requiring manual feature extraction or selection. This can lead to high accuracy and robustness in HAR tasks. However, deep learning models also face some challenges and limitations, such as high computational costs, large data requirements, interpretability issues, etc. However, with advancing technologies, those costs and physical overhead have considerably reduced, thus making it eventually possible to deploy these models even on some microcontrollers.

Modeling the problem: The bird's eye view of the problem is *given some data from sensors (mostly time series) to classify or predict which task the subject is performing*. Taking into account the form of input data, we must choose layers such that it captures both local features and global features of data during classification. It is known that CNNs and LSTMs work well for capturing local features and global features, respectively. Thus we reason to go ahead with some CNN-LSTM-based model architecture.

MHealth dataset exploration: As the rule of thumb for any deep-learning classification, we must explore the dataset. Following is the class description and available value count of data points. We must do this initial exploration in order to check if the dataset is not biased towards a particular category, as it may force our deep-learning model to guess that category to achieve max accuracy. I found that it is biased towards category 0. Thus I downsampled it to a reasonable number. (See **Table 2.1**)

Table 2.1: Class-wise distribution of data

Class category	Class Description	Value count
0	Nothing	30000
1	Standing Still	30720
2	Sitting and Relaxing	30720
3	Lying down	30720
4	Walking	30720
5	Climbing stairs	30720
6	Waist bends forward	28315

7	Frontal elevations of arms	29441
8	Knees bending (crouching)	29337
9	Cycling	30720
10	Jogging	30720
11	Running	30720
12	Jump front and back	10342

Related Work:

- Authors of this paper [1] used sensors to collect data for predicting the impact of greenhouse gas emissions on climate change. However, their data collection method was flawed and resulted in inaccurate measurements. Despite this limitation, their proposed deep learning model(based on LSTMs) was able to produce reasonable predictions and achieve high performance.
- To recognize the upper limb gesture in rehabilitation, the authors in this [2] used a fully connected deep learning approach. They compare their model to various machine learning algorithms and show that the proposed fully connected neural network outperforms them in gesture recognition.
- the authors of this paper [3] demonstrated that a category-aware gated recurrent unit model for the following POI category recommendation performs better than other baseline methods.
- In [4], the Authors have presented a combination of wearable sensor-based and Kinect sensor-based strategies for generating person stepping patterns and constitutive models of the work.

Methodology:

Author's Approach: The author used nine values, namely triaxial acceleration, gyro, and linear acceleration, to train their model. However, the problems encountered by the authors are similar to that of me in processing data. Different sensors have different sampling rates. Thus, some sensors record values at a faster pace than others. To resolve this ***data imbalance***, we must either under-sample or over-sample.

Time-series data thus obtained must be ***segmented*** into segments to avoid look-ahead bias. The data is split into train/validation/test sets. This simulates the situation in a production environment where after training a model, you evaluate data coming.

In the last step of preprocessing: output labels have been converted into ***one-hot encoding.*** That is, classes have been ***relabelled*** as numbers.

Preprocessed data is fed into the CNN-LSTM-based model, and we extract softmax embeddings that denote relative probability for activity classification.

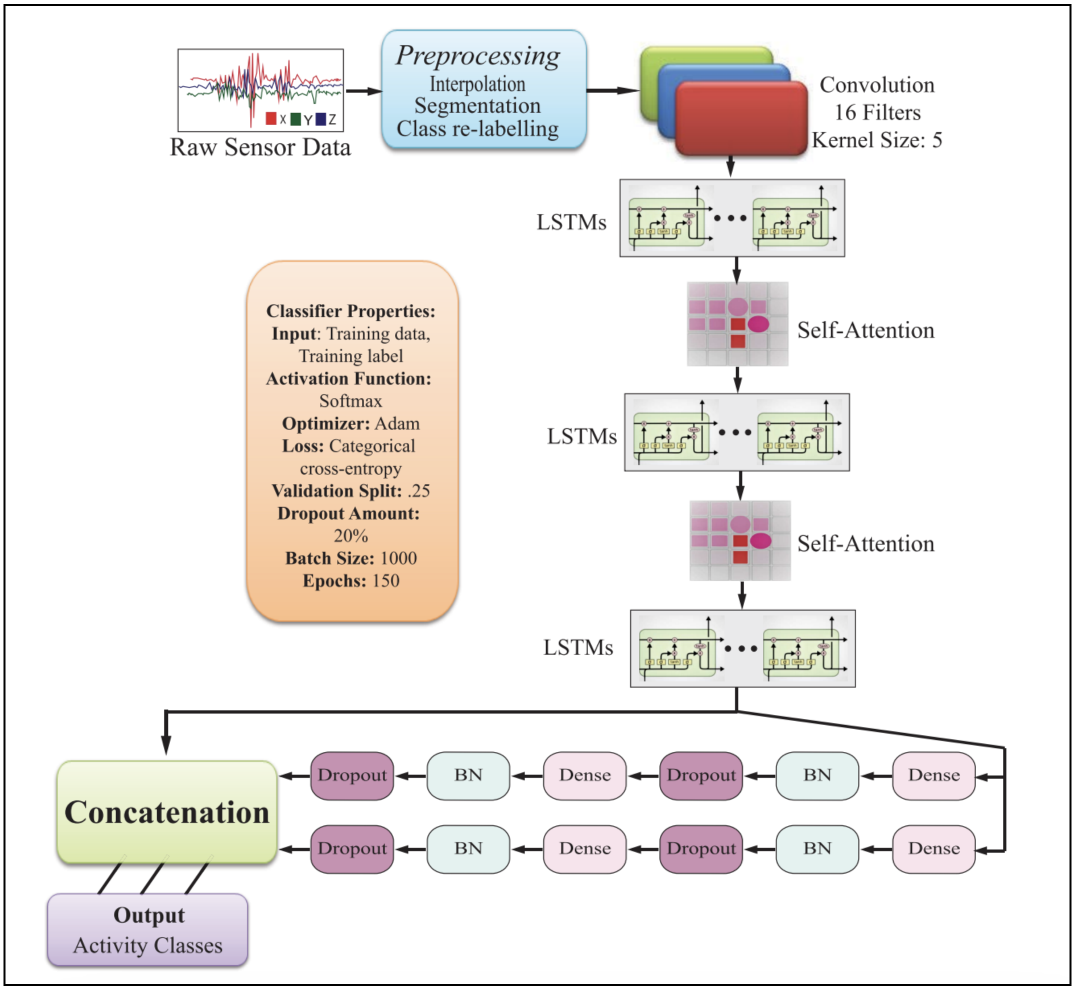
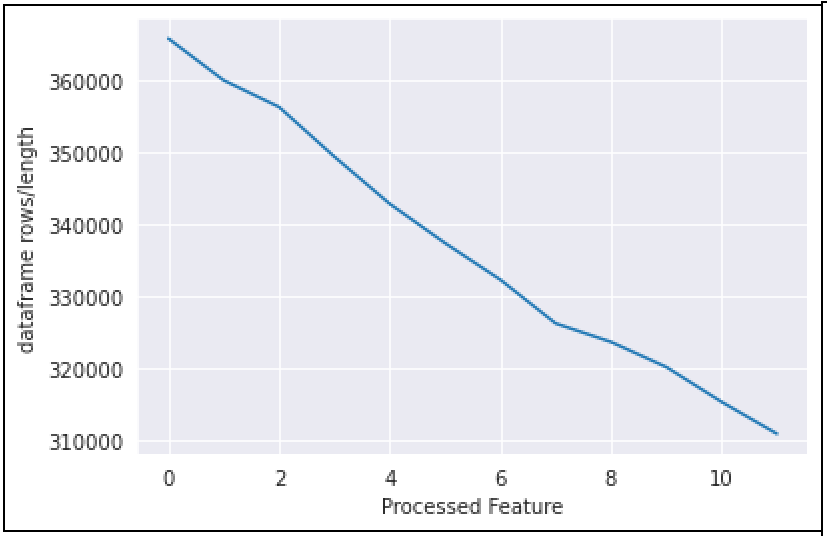


Fig 2.1: Original architecture in the paper.

My Approach: The decision of preprocessing and other things comes after observing the dataset carefully, so I started with Dataset Exploration:

- **Exploratory Analysis:** As observed through printing value counts for each category, the value count in category 0 is unbalanced. Thus we must take steps to preprocess the data to make sure to remove any bias. We pre-process data by limiting the category 0 value count to 30000 through random sampling. Sensor data sometimes may read noise. Thus, we process the outliers by taking only that data within a 98% confidence interval. This step accounts for uncertainty in the forecast.
- **Preprocessing:** Random Sampling: randomly sampled category 0 data points to limit their values to 30000. Alleviates class imbalance. (See Table 2.1 for modified data).The confidence interval of 98%: alleviates noise. (See Fig 2.2 and Fig 2.3)



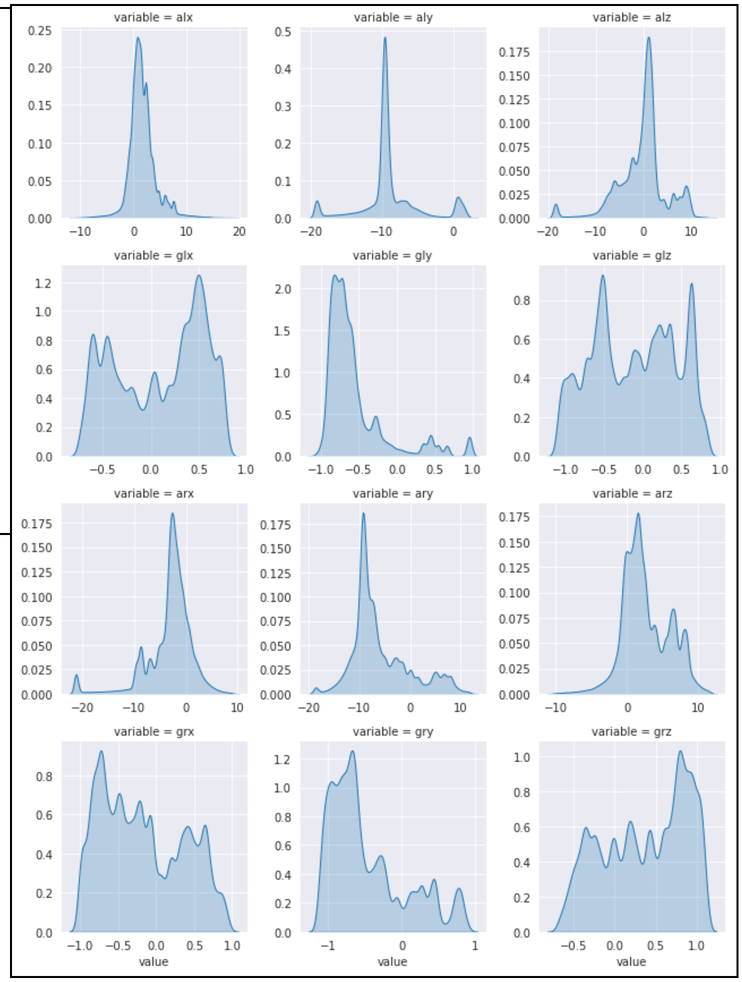


Fig 2.2: Processed row count(left) and data distribution after preprocessing(right)

- **Data Modeling and Process Overflow:** We model our data into two sets, i.e., the training and testing datasets. I divided my dataset into training data (246283 rows, 12 columns) and testing data (64423 rows and 12 columns), which was then segmented into time steps of 100 units.

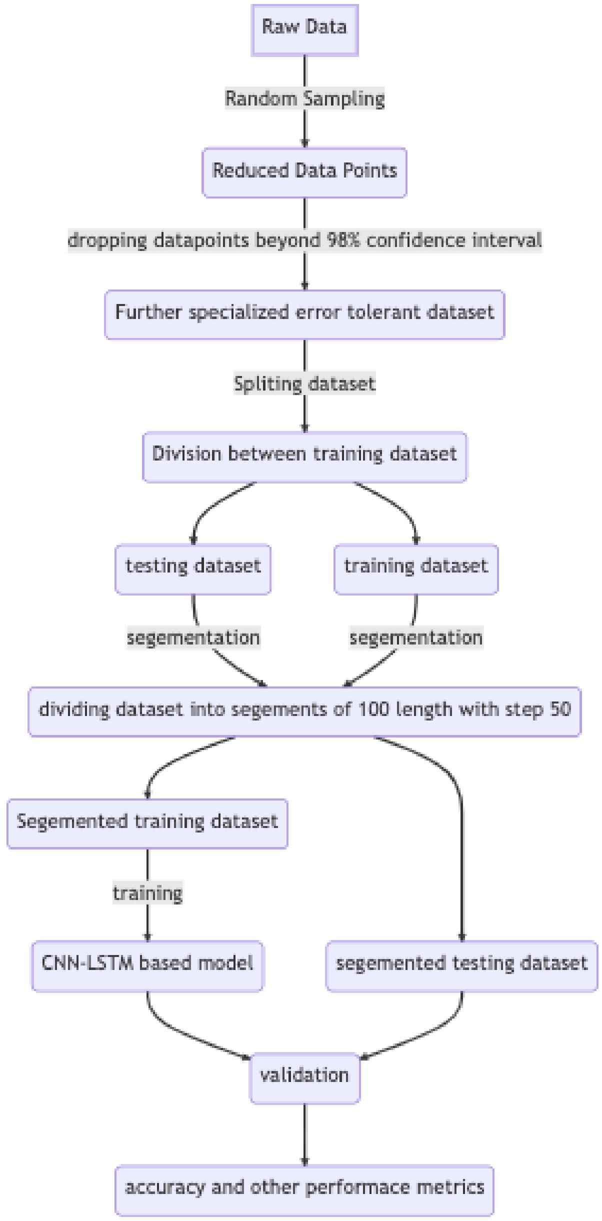


Fig 2.3: Process overflow of my approach

- **Hyperparameter tuning:** I have tested my code for various epochs, optimizers, and divisions, but the best result, that is, an accuracy of 99.69%, is achieved through the following hyperparameters:

Epochs: 100
Optimizer: Adam
Learning rate: 0.01
Loss function: Sparse Categorical cross-entropy

- Model architecture:

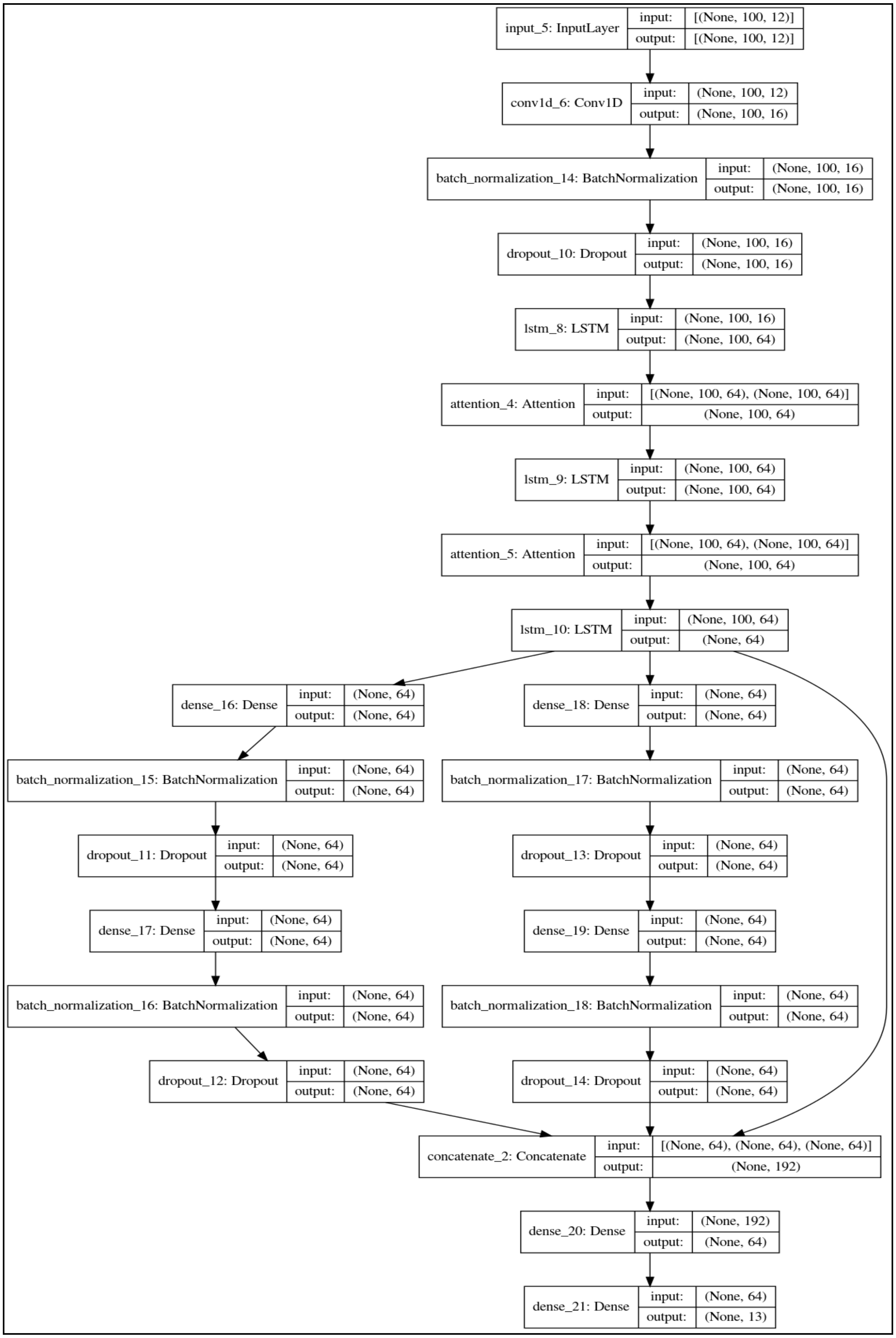


Fig 2.4: Schematic diagram of the proposed model that is employed for MHealth Dataset.

- ### Experiment Details:

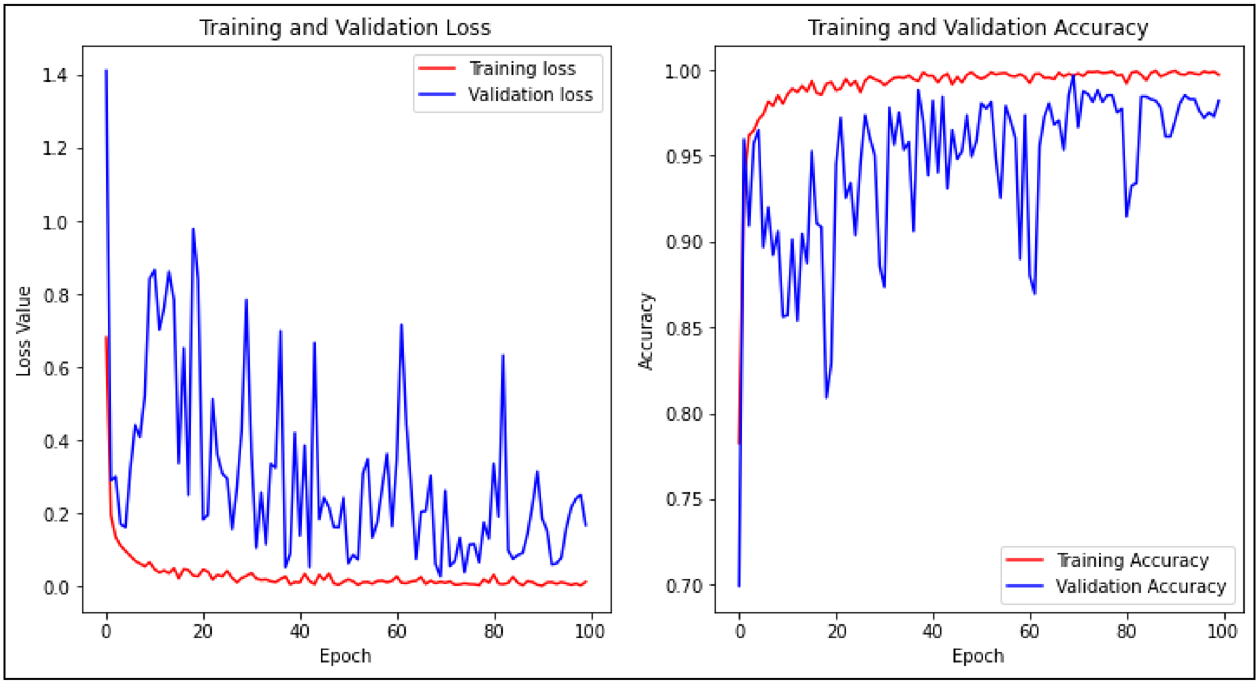


Fig 2.5: Loss vs. Epochs (Left) and Accuracy vs. Epochs (Right) plots for Best model (99.69% accuracy)

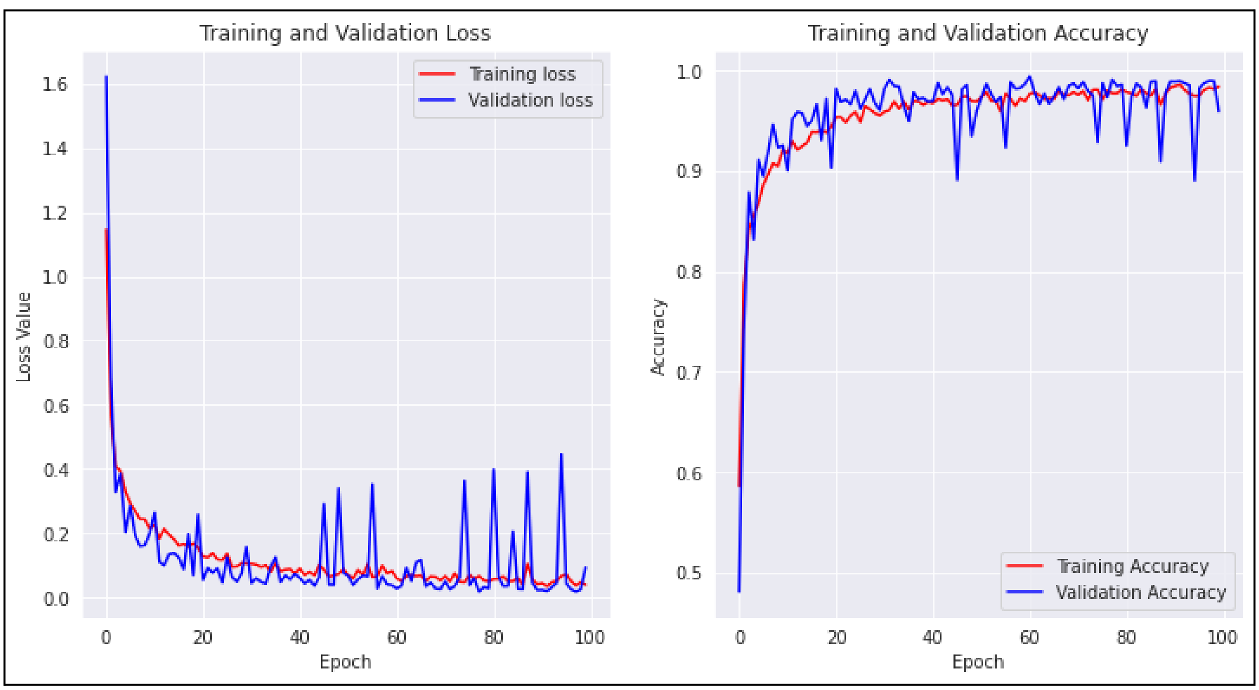


Fig 2.6: Loss vs. Epochs (Left) and Accuracy vs. Epochs (Right) plots for a model trained on data from Left Ankle only (98.92% accuracy)

Fig 2.7: Loss vs. Epochs (Left) and Accuracy vs. Epochs (Right) plots for a model trained on data from Right Elbow only (85.31% accuracy)

- **Code:** Code for the best model has been attached and named *99_69_best_model.ipynb*, other files like *99.69.h5(trained best model)* and *Generic.ipynb* are also attached for analysis.
- **Results:**
  Accuracy is the primary metric used to evaluate the performance of the model; however, through the compiled results, we can see other metrics:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	92
1	0.99	1.00	1.00	122
2	0.99	0.99	0.99	124
3	1.00	1.00	1.00	122
4	1.00	1.00	1.00	120
5	1.00	0.99	0.99	84
6	1.00	1.00	1.00	106
7	1.00	0.99	1.00	112
8	0.99	0.99	0.99	116
9	0.99	1.00	1.00	121
10	1.00	1.00	1.00	89
11	1.00	1.00	1.00	52
12	1.00	1.00	1.00	27
accuracy			1.00	1287
macro avg	1.00	1.00	1.00	1287
weighted avg	1.00	1.00	1.00	1287

Fig 2.8: Performance metrics for best model (99.69% acc.)

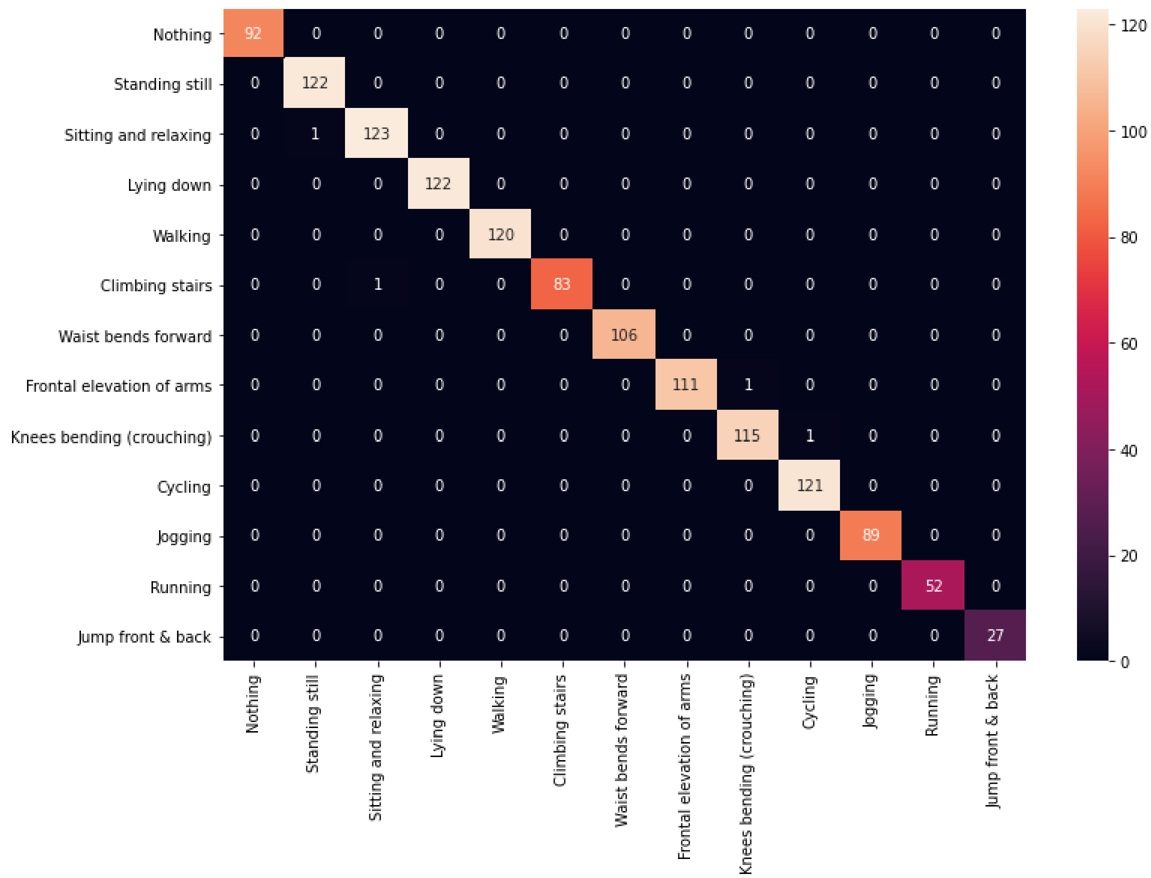


Fig 2.9: Confusion Matrix for best model (99.69% acc.)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	370
1	0.90	1.00	0.95	493
2	1.00	1.00	1.00	491
3	1.00	1.00	1.00	361
4	1.00	1.00	1.00	468
5	1.00	1.00	1.00	369
6	1.00	1.00	1.00	459
7	1.00	0.87	0.93	399
8	1.00	1.00	1.00	460
9	1.00	1.00	1.00	450
10	1.00	1.00	1.00	322
11	1.00	1.00	1.00	203
12	1.00	1.00	1.00	83
accuracy			0.99	4928
macro avg	0.99	0.99	0.99	4928
weighted avg	0.99	0.99	0.99	4928

Fig 2.10: Performance metrics for the best model with data from only left ankle (98.92% acc.)

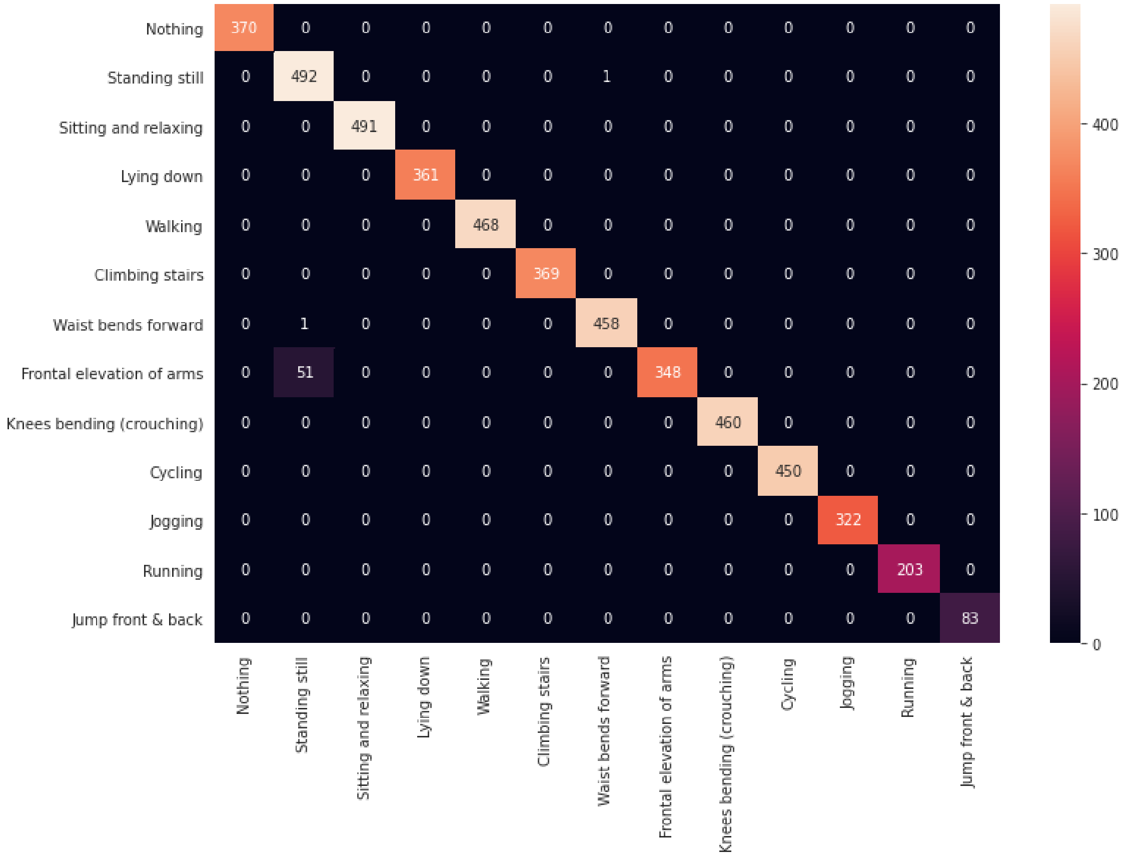


Fig 2.11: Confusion Matrix for the best model with data from only left ankle (98.92% acc.)

	precision	recall	f1-score	support
0	1.00	1.00	1.00	92
1	0.98	0.98	0.98	122
2	0.98	0.99	0.98	124
3	0.97	1.00	0.98	122
4	0.99	0.85	0.91	120
5	0.70	0.82	0.76	84
6	0.67	0.04	0.07	106
7	0.97	0.97	0.97	112
8	0.51	0.96	0.67	116
9	1.00	0.97	0.98	121
10	0.98	0.57	0.72	89
11	0.98	1.00	0.99	52
12	0.42	1.00	0.59	27
accuracy			0.85	1287
macro avg	0.86	0.86	0.82	1287
weighted avg	0.88	0.85	0.83	1287

Fig 2.12: Performance metrics for the best model with data from only the right elbow (85.31% acc.)

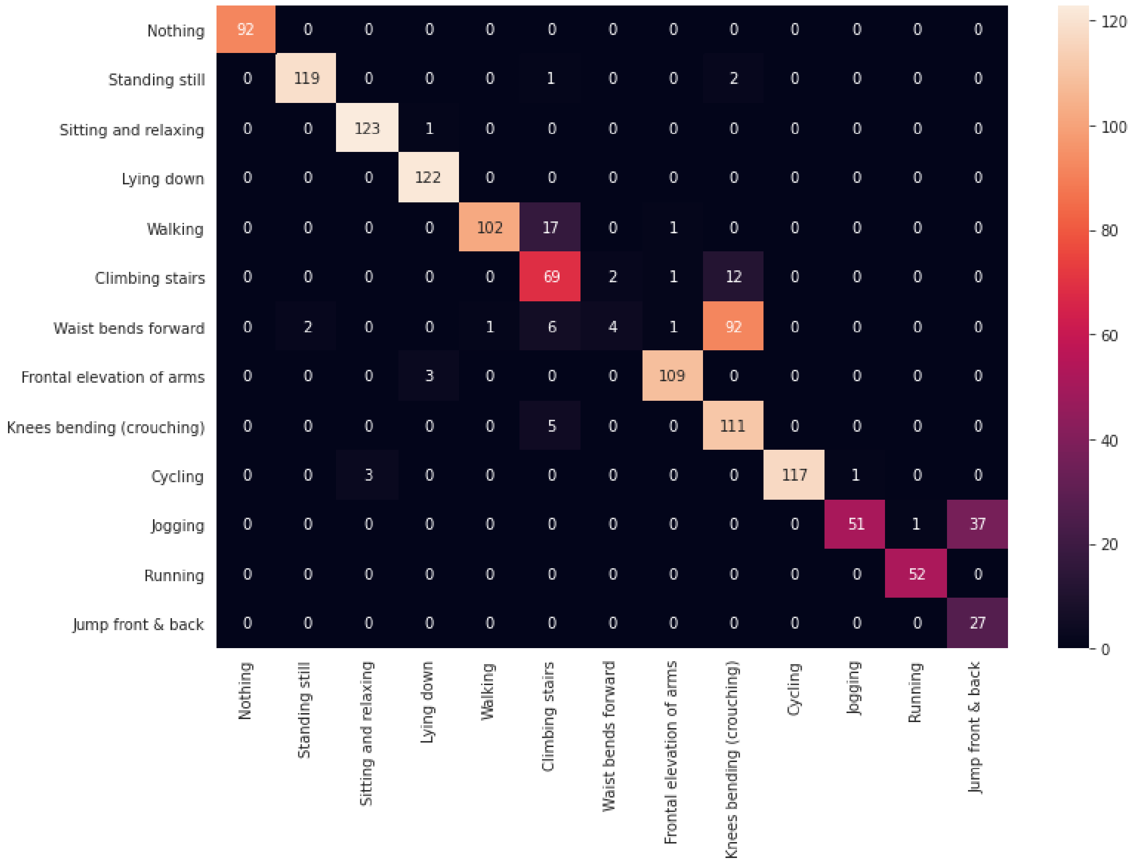


Fig 2.13: Confusion Matrix for the best model with data from only the right elbow (85.31% acc.)

Conclusion:

Based on my experiments with various splits of the datasets and different hyperparameters, I arrived at the following conclusions:

1. Using a sensor only on the right elbow is not a feasible option as it yields the lowest accuracy, regardless of the parameter settings. This can be clearly observed from the epoch versus accuracy graph. (**Fig** 2.7)
2. In addition to the model proposed by the authors of the original paper, I experimented with adding one to five dense layers to the network. The optimal performance was achieved by using a single dense layer, as shown in **Fig** 2.4.
3. The current approach to hyperparameter tuning relies on trial-and-error methods, which are inefficient and unpredictable. A possible direction for future work is to enhance the proposed model by applying more data augmentation techniques, optimizing the hyperparameters, and investigating other attention mechanisms or architectures such as self-attention, local/global attention, soft/hard attention, etc.
4. One of the limitations of the proposed model is that the attention weights are not easily interpretable or explainable, which may hinder our understanding of how the model works and why it makes certain decisions. Therefore, a possible direction for future

research is to explore methods that can enhance the explainability of attention-based models, such as visualizing the attention maps, generating natural language explanations, or incorporating human feedback.

References:

- [1] Y. Liu et al., "A long short-term memory-based model for greenhouse climate prediction," Int. J. Intell. Syst., vol. 37, no. 1, pp. 135–151, 2022.
- [2] Q. Liu et al., "A fully connected deep learning approach to upper limb gesture recognition in a secure FES rehabilitation environment," Int. J. Intell. Syst., vol. 36, no. 5, pp. 2387–2411, May 2021.
- [3] Y. Liu et al., "An attention-based category-aware GRU model for the next POI recommendation," Int. J. Intell. Syst., vol. 36, no. 7, pp. 3174–3189, Jul. 2021.
- [4] V. Bijalwan, V. B. Semwal, and T. K. Mandal, "Fusion of multisensor-based biomechanical gait analysis using vision and wearable sensor," IEEE Sensors J., vol. 21, no. 13, pp. 14213–14220, Jul. 2021.