Exercise Pattern Prediction

Mustofa Ahmed 180104005@aust.edu
Samia Sabrina Nabi 180104006@aust.edu
Sadia Mobasshira 180104014@aust.edu
Department of Computer Science and Engineering Ahsanullah University of Science and Technology
Dhaka,Bangladesh

*

I. Abstract

As the world is modernizing, we are spending more time sitting at desks. It can lead to various other health-related problems, which are prevented by exercising. These exercises include walking, jogging, weight lifting etc. Exercises like weight lifting can lead to multiple sprains, strains, or fractures if they are not done correctly. Learning how to do them correctly requires supervision from gym experts. But in our busy lives, we barely have time to go to the gym. So in our study, we are exploring whether we can correctly predict if a person is exercising correctly or not by analyzing data collected from various sensors attached to the user. Previous authors haven't explored the deep learning approach, so in our model, we set out to find if using deep learning can be a feasible option or not for predicting if a user is correctly exercising or not. Among the model we have used, we found fully connected network achieved the highest accuracy of 72.85%.

II. Introduction

Exercising regularly leads to numerous benefits, both physically and mentally. It reduces the risk of high blood pressure, heart disease, diabetes, obesity, and more. Apart from the physical welfare, it also helps combat stress and anxiety [5]. Weight lifting is one of the most common exercises a person can frequently do at home. Weight lifting allows us to keep our bones strong and healthy as we start losing bone density by a tiny fraction after we become older than 30. It also helps improve posture, sleep, mood, and energy levels [6]. Free weight exercises are the cause of 90.4 % of injuries associated with weight training. Incorrectly lifting weights can lead to sprains, strains, fractures, and other painful injuries [7].

A possible way of avoiding such injuries is hiring a personal instructor who will point out if the users are making any mistakes. This mistake can be in body movement, posture, etc. Hiring an instructor can be very expensive and is also not always available. In our study, we will try to predict if a user performs Unilateral Dumbbell Biceps Curl passive sensors correctly. This avenue is much cheaper than scalable than hiring instructors. Previous studies have used four sensors attached to the equipment to classify if the user is correctly performing the weight lift. Attaching many sensors may make the users uncomfortable and unable to exercise freely. So, we are proposing to use less number of sensors and obtain a good model. The previous study has also not used any deep learning-based approach, allowing us to test how deep learning works well on this data. The first model we used is the Fully Connected Neural Network(FCC). The fully connected neural network resulted in an overfitting problem if we included all data. So we overcame the problem by reducing the size of the data. Finally, the obtained accuracy was 74%. The second model we used is the Convolutional Neural Network model (CNN). We passed time series data to this model. We didn't pass and derived features to this model and used time series data only as CNN can work by detecting patterns in time series data. The accuracy achieved by CNN was 61%. As our third model, we have used the Long Short Term Memory (LSTM) model. The LSTM performed very poorly and obtained a result of 50%.

III. Background Study

- **Fully Connected Layer:**Fully connected layers in a neural network are those layers where every activation unit in the layer above is connected to every input in the layer above it.
- **Convolutional Neural Network 1D:**At least one convolutional layer is contained in a CNN. The convolutional layer is locally linked, unlike the fully-connected layer in MLP.
- **Long-Short Term Memory:**LSTM is an abbreviation for long short-term memory networks.Apart from single data points such as images, LSTM has feedback connections, which means it can process the entire sequence of data.
- **Accuracy:**Accuracy is the most basic model evaluation metric. The number of correct predictions divided by the total number of predictions produced.
- **Precision:**Precision refers to the quality of a positive prediction and is one of the indications of the machine learning model's success. It is the ratio of True Positives to all the positives predicted by the model.
- **Recall:**Recall indicates the percentage of true positives that the model correctly classifies. It's also known as Sensitivity. It's the proportion of true positives in your dataset to all positives.

- **F1score:**The F1 Score is the outcome of a trade-off between recall and precision. In certain circumstances, high precision is required, while in others, high recall is required.

IV. Related Works

There are few study publications on this topic from researchers that sought to investigate this area.

Very little research has examined how an activity is carried out "well." The exercise used in Velloso et al. [1] was a unilateral dumbbell biceps curl. They have used data from various sensors to forecast how successfully the activity is performed. Six men in total were enlisted to help with data collection. The users' bodies and their equipment both have sensors attached to them. The sensor's readings included roll, pitch, yaw, total acceleration, etc. Numerous statistical features that define the sensor readings of a window were generated using this data. A machine learning classifier, Random Forest, was used. The overall accuracy was 80%. However, the study has lots of scope for improvement. Include female participants to create a more generic classifier. The design, implementation, and assessment of FEMO, a passive RFID-based free-weight activity monitoring system, are presented by Ding et al. [2]. In this work, the training tools are passive RFID-tagged dumbbells. The four fundamental modules that make up the implementation of FEMO are Doppler value pre-processing, activity segmentation, activity recognition, and activity assessment. The outcome of in-depth tests conducted on 15 volunteers shows that FEMO may be used for a range of free-weight exercises and can provide helpful feedback for users' activity correction. The algorithm of FEMO cannot deal with non-stop activities. As a result, the accuracy of the segments will reduce to some extent.

Lee et al. [3] demonstrated a single tri-axial accelerometer-based PLL (Personal Life Log) device that can recognize human activity and generate information on exercise. Using subject-independent and subject-dependent recognition for the six everyday activities of lying, standing, walking, going up and down stairs, and driving, the PLL system was evaluated on a total of twenty subjects. Electronic tools like video sensors (like cameras) and motion sensors are used to record life logs. Artificial neural networks (ANNs), based on the feed-forward back-propagation algorithm, are used as a classifier after computing the feature vector for each level. Subject-independent recognition accuracy was 94.43 %, while overall subject-dependent recognition accuracy was 96.61 %. Even though they utilized a laptop for their feasibility study, it would be preferable to develop a PLL system on mobile phones that is more portable and less constrained. Kotsev et al. [4] analyzed a dataset collected from thousands of users. They have shown graphs amount of user activity changes throughout the year. They have shown simple graphs of how activity levels across various countries. They have tried to conclude why these changes occur from the results of these graphs. Using the data, they have also tried to predict the future activity level of the users based on time series. Their results are very poor, and the highest accuracy was only 53% using the Logistic Regression Model.

V. Dataset

The dataset has 40000 data and 160 features.

A. *Readings from sensors:*

Pitch is a rotation around the vertical axis.Roll is a rotation around the front-to-back axis.Yaw is a rotation around the side to side axis. Gyroscope is a device that can measure and maintain records of an object's orientation and angular velocity. Magnet represents the magnitude of the geomagnetism and magnetism generated by a magnet. Total acceleration refers to how quickly the sensors move. Six participants will wear accelerometers on their belts, forearms, arms, and dumbbells to collect data and They provide several types of measurement. Number of windows refers to how many times participants are lifting dumbbells. Five different ways to complete a set of 10 unilateral dumbbell biceps curls were requested of the participants. When a participant performs exactly as required by the specification, class A is assigned. Class B is assigned when participants in group throw their elbows forward. Lifting the dumbbell only halfway those are assigned Class C. Those who only lower the dumbbell halfway are given a Class D. And participants in a group are given Class E when they throw their hips forward. Class A represents how the activity should be carried out, whilst the other 4 classes often make mistakes. The raw time is calculated using unix time.A system for displaying a point in time is known as Unix time. It is the number of times that has passed from January 1st, 1970 in seconds.

We derived various attributes from the readings of the sensors.

B. *Derived Features*

Maximum: Maximum sensor value of a particular sensor reading belonging to a specific window.

Minimum: Minimum sensor value of a particular sensor reading belonging to a specific window.

Standard deviation: Standard deviations are used to determine how variable a distribution is. The standard deviation for each attribute is measured.

Average: Average is a number that represents the center or typical value in a group of data, specifically the mode, median, or mean, which is derived by dividing the total number of values in the set by their sum.

Amplitude: Amplitude is the difference between maximum and minimum values. All the maximum and minimum values are calculated from sensor value.

Kurtosis: The combined weight of a distribution's tails in relation to its center is measured by kurtosis. A bell peak with the majority of the data being within three standard deviations (plus or minus) of the mean can be seen when a collection of essentially normal data is graphed using a histogram.

Skewness: Skewness is a measure for a distribution's asymmetry. When the left and right sides of a distribution are not mirror reflections, it is asymmetrical. Right (or positive), left (or negative), or zero skewness can all apply to a distribution.

C. Dataset Visualization

This section contains some visualizations of the data.The diagrams are time series graph.

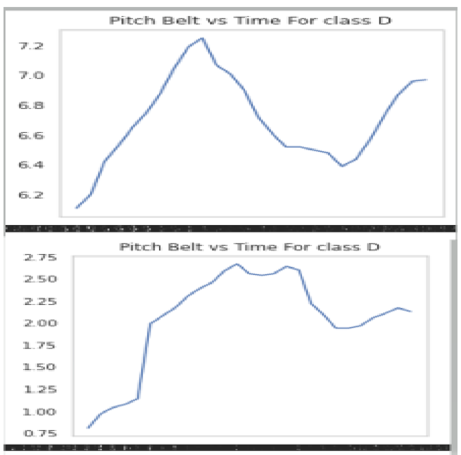

Fig. 1. Pitch vs Time Series for class D

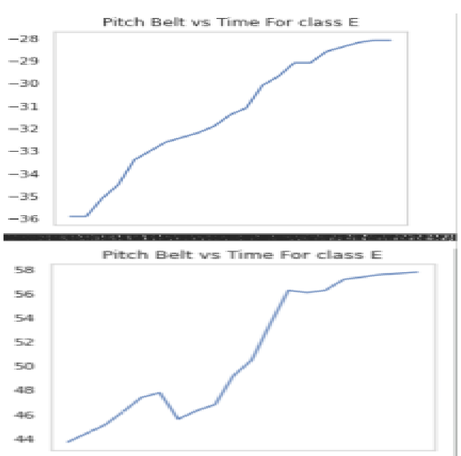

Fig. 2. Pitch vs Time Series for class E

VI. Methodology

This section discusses the data preprocessing steps and the deep learning models we have used. As part of the deep learning models, we have used a Fully Connected Neural Network a Convolutional Neural network and a Long Short Term Memory.

A. Workflow

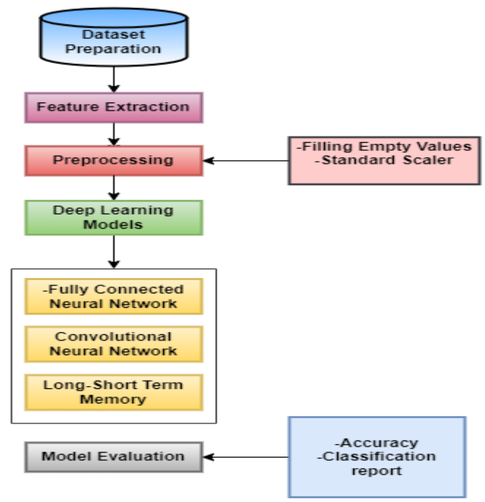

Fig. 3. Working Procedure

B. Data preprocessing

In the first part of preprocessing, we removed all the data associated with the 'arm' and 'forearm' sensors because our target was to achieve a carryout experiment with data from 2 sensors only. For FCC, we first filled the empty values by recomputing the statistical features for all features. Then we applied the standard scaler to scale all values to -1 to 1. For the CNN and LSTM models, we had to convert the data into windows of data. Using the NumPy library, we reshaped a window of data and stacked them on top of each other. We took the size of the window to be 27. Each slot of data had 27 values from each sensor. There were a total of 30 different readings from 30 different sensors. So, the size of the slot was (30, 26).

C. Deep Learning Models

This section contains description of the deep learning models that we have used. It contains details about the number of layers, sequence, and the number of nodes in each layer.

- **Fully Connected Network**
  For our model, we have used nine layers of a fully connected network. The activation function for each layer contained Relu and Tanh activation functions interchangeably.We played around with the activation functions a bit.

Layer	Output Shape	Param
Dense	(None,1200)	106800
Dropout	(None,1200)	0
Dense	(None,1500)	1801500
Dropout	(None,1500)	0
Dense	(None,900)	1350900
Dropout	(None,900)	0
Dense	(None,700)	630700
Dense	(None,350)	245350
Dense	(None,5)	1755

- **Convolutional Neural Network**

  Here, we used the readings from the sensor values directly like window number instead of deriving new features from these readings. Using CNN allows us to detect patterns in these readings and the model to train from these findings.

Layer	Output Shape	Param
Conv1D	(None, 29, 6)	106800
Conv1D	(None, 29, 9)	0
Dropout	(None, 29, 9)	1801500
Maxpooling1D	(None, 29, 9)	0
Conv1D	(None, 14, 12)	1350900
Dropout	(None, 14, 12)	0
Dropout	(None, 168)	630700
Flatten	(None, 400)	245350
Dense	(None, 400)	106800
Dropout	(None, 700)	0
Dense	(None, 700)	106800
Dropout	(None, 1000)	0
Dense	(None,1000)	106800
Dropout	(None,6)	0

- **Long Short Term Memory**

  Here, we used four layers of a long short term memory.We used RuLu as activation function.The learning rate for the model was set to 0.001, and the Adam optimizer was used.

Layer	Output Shape	Param
LSTM	(None,22)	4312
Dense	(None,512)	11776
Dense	(None,256)	131328
Dense	(None,6)	1542

VII. Result

This section contains our analysis of the performance of all three deep learning models we used. According to our results, FCC performed better than LSTM and CNN.However, this is comparison is not justified as we used two types of data for these models. For CNN and LSTM, we have used times series data. In other words, we used raw sensor data for LSTM and CNN. This is because LSTM and CNN can pick up differences in these time series. They are capable of observing the differences in corresponding samples of data. On the other hand, this data cannot be fed to the FCC because FCC is not built for time-dependent values. A row is not connected to the next row for FCC. This connection between samples exists for CNN and LSTM. So, this is why time series data was used for CNN and LSTM. We have provided details about the CNN layers used in the methodology section. Here we discuss some of the parameters we have used. The dataset had varying window sizes. It varied from 5-72. So our first intuition was to use padding with value 0. We padded all the data that contained window sizes less than 72. However, the performance of this was horrible. Padding with the value 0 did not help the model learn something meaningful. So our second intuition was to take a smaller window size. So we experimented with various window sizes, and the best-performing window size was 27. The accuracy obtained using this window size was 61.22%. The other window size's performance is shown in the table. For window sizes smaller than 27, we dropped them. For larger than 27, we turned them into two windows if the values are larger than 54.

CNN	20	25	27	29	30
Without Padding	28.89%	49.30 %	**61.22%**	59.48%	55.85 %

TABLE I
PERFORMANCE OF CNN MODEL

For the LSTM model, the accuracy obtained was only 50%. So for our dataset, CNN performed much better. We even experimented with a hybrid approach of using CNN and LSTM. The obtained accuracy is 58.3%, and this is close to the CNN model. Our findings on CNN and LSTM models have been summarized in figure X.

CNN+LSTM	Padding	Window Size=27
	29.56%	**58.29%**

TABLE II
PERFORMANCE OF CNN LSTM MODELS

For FCC, the data we used was statistical data. From the window sizes, many statistical values were derived, which can be used to give a description of the raw sensor data of a particular window. This data was fed to the FCC, and it outperformed both CNN and LSTM. We performed oversampling, but the accuracy did not increase. So we did not continue with oversampling. Training test split in the ratio of 70 to 30 resulted in better performance. Since the size of our dataset is small, we could not build a large model. Our findings on the FCC model have been summarized in Figure Y.

FCC	Test train split 80/20	Test train split 70/30
Oversampling	72.22%	72.62%
Without Sampling	71.9%	**72.85%**

TABLE III
PERFORMANCE OF FCC MODEL

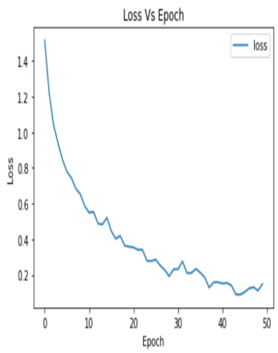

Fig. 4. Target Class Distribution

Classes	Precision	Recall	F1-score	Support
E	**0.93**	**0.87**	**0.90**	47
D	0.55	0.68	0.61	41
C	0.68	0.57	0.62	44
B	0.74	0.73	0.74	48
A	0.77	0.78	0.77	**72**

TABLE IV
CLASSIFICATION REPORT FOR FULLY CONNECTED NEURAL NETWORK

VIII. FUTURE WORK

We'll use a bluetooth sensor for a more appropriate usage. The mobile phone will get sensor data through bluetooth. Through a mobile phone, the user may observe how to lift a dumbbell properly. We'll use more sensors, including magnetometers, gyroscopes, and orientation sensors. A device that measures magnetic dipole moment is called a magnetometer. Gyro sensors are tools that detect angular velocity. A device's orientation in relation to an orthogonal coordinate frame is determined via an orientation sensor. Users will receive more precise output from these sensors. Many attributes may be derived from those sensors.

IX. CONCLUSION

Regular exercise has a variety of advantages for the body and mind. Lifting weights improperly can result in fractures, sprains, strains, and other painful injuries. Four sensors that were attached to equipment in earlier trials were used to determine if the operator was carrying out the weight lift appropriately. The users may get uncomfortable and unable to exercise freely if several sensors are attached. We propose using fewer sensors to get a good model.The previous study has not used any deep learning-based approach, allowing us to test how deep learning works well on this data. The first model we used is the Fully Connected Neural Network(FCC). FCC resulted in an overfitting problem if we included all data. So we overcame the problem by reducing the size of the data. The second model we used CNN and passed time series data to this model. As our third model, we have used the LSTM model.

REFERENCES

[1] Velloso, E., Bulling, A., Gellersen, H., Ugulino, W., Fuks, H. (2013, March). Qualitative activity recognition of weight lifting exercises. *In Proceedings of the 4th Augmented Human International Conference (pp. 116-123).*

[2] Ding, H., Shangguan, L., Yang, Z., Han, J., Zhou, Z., Yang, P., ... Zhao, J. (2015, November). Femo: A platform for free-weight exercise monitoring with rfids. *In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (pp. 141-154).*

[3] Lee, M. W., Khan, A. M., Kim, T. S. (2011). A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. Personal and Ubiquitous Computing, 15(8), 887-898.

[4] Kotsev, G., Nguyen, L. T., Zeng, M., Zhang, J. (2014, November). User exercise pattern prediction through mobile sensing. *In 6th International Conference on Mobile Computing, Applications and Services (pp. 182-188). IEEE.*

[5] Importance of Exercise. (n.d.). Importance of Exercise Essay. Retrieved July 23, 2022, fromhttps://www.toppr.com/guides/essays/importance-of-exerciseessay/https://www.toppr.com/guides/essays/importance-of-exerciseessay/

[6] https://globalnews.ca/news/3513498/8-reasons-why-weighttraining-is-incredible-for-yourhttps://globalnews.ca/news/3513498/8-reasons-why-weighttraining-is-incredible-for-your

[7] https://www.mayoclinic.org/healthy-lifestyle/fitness/indepth/weight-training/art-2004584https://www.mayoclinic.org/healthy-lifestyle/fitness/indepth/weight-training/art-2004584