IASc-INSA-NASI

Summer Research Fellowship 2022
4 week Report

Vrushali Pandit
Application No. **: ENGS3583**

Under the supervision of:
Prof. Ashok Kumar Dasmahapatra
Department of Chemical Engineering and Centre of Nanotechnology
Indian Institute of Technology, Guwahati (IITG)

Guided by:
Shivani Gupta
Research Scholar
Department of Chemical Engineering and Centre of Nanotechnology
Indian Institute of Technology, Guwahati (IITG)

Project Title:

Prediction of cattle behaviour based on real time, on body sensor data using ML or DL models

Motivation:

Manual monitoring of cattle is done based on visual characterisation of an animal's gait/ behaviour. Other standard health monitoring processes involve measuring rectal temperature, and collection of bodily fluids for laboratory testing. Performing these manual evaluations and diagnostic tests for an entire flock is time and effort consuming. Moreover, natural movements of animals may be influenced by the presence of a human observer. Wearable sensors (such as accelerometers, gyroscopes, compass, GPS, and pressure sensors) may provide a possible solution by enabling measurements of acceleration and gyrations of the cattle in real time, 24*7. This is an example of Precision Agriculture (PA) which pays attention to the welfare of individual animals instead of a herd based approach. Such behaviour analysis can possibly provide insights into:

1. Animal Health
2. Feed intake
3. Heat/ Oestrus - The onset of the ovulation cycle in female mammals is indicated by changes in behaviour such as increased restlessness and smaller rumination times.

Insights from Literature Review:

Various studies have taken this approach of precision agriculture by using sensor data to predict cattle activities. Old studies have used decision trees for classifying data obtained from various types and positions of sensors (Robert et al. (2009), Gonzalez et al. (2015)). Dutta et al. (2015) has used an ensemble of binary classifiers to apply the one-vs-all approach for 5 activities, achieving 96% accuracy. Many studies have compared different algorithms such as DT, SVM, RF, kNN, Naive Bayes and HMMs against each other (Abell et al. (2017), Benaissa et al. (2019), Diosdado et al. (2015)). Khahn et al. (2020) utilised Gradient Boosted Decision trees. The first usage of deep learning based CNN model was done by Kasfi et al. (2016). Further Peng et al (2019) used RNN and LSTM, Rahman et al. (2016) used autoencoders to auto-extract features and SVM for the classification, the first work in this field of hybrid learning. Pavlovic et al. (2021) did an extensive hyperparameter training of CNNs to predict cattle behaviour, along with significant model pruning to make it light weight.

Relatively little work is done on how to handle class imbalance of activities. There has also been only one study which uses Deep Hybrid Learning. Further a study had applied a divide and conquer approach using 1D CNNs to use a model to first classify high level activities and then use a second model to classify lower level activities for human activity recognition (Cho et al. (2018)). However, this approach has not yet been applied in cattle activity prediction.

Objective of further work:

I am currently working on a publicly available dataset of sensor data of cattle. Based on thorough discussion with my guide, I plan to benchmark four different kinds of class imbalance handling techniques (random oversampling and undersampling, SMOTE and near miss algorithm). The performance of these techniques will be based on accuracy after using a logistic regression classifier. As is standard to every time series problem, I will also benchmark different window lengths. I will try out two different classification methods. These are a divide and conquer based CNN model and a Deep hybrid model where a CNN extracts features while a boosted tree performs the classification.