**BYOP: PROJECT PROPOSAL**

**DroneGuard**

**Real-time Anomaly Detection in Drone Flight Logs Using LSTM and Deep Learning**

* **Contact Information:**

Name: **Archi**

Contact Number: **9016128573**

Branch: **Mathematics and Computing**

Email ID: **archi@ma.iitr.ac.in**

Enrollment number: **22323008**

* **Introduction:**

My project involves implementing a research paper titled "**DronLomaly: Runtime Detection of Anomalous Drone Behaviors via Log Analysis and Deep Learning**." and **varying Sequence Length to optimize the Sequence Length**. The central objective is to enhance the safety and performance of drones, which play a pivotal role in area surveillance and cinematography tasks. While the drone technology landscape has seen exploration in various aspects, there exists a noticeable gap in the attention given to log anomaly detection.  
Existing approaches, especially those employing drone fuzzing methods, have demonstrated effectiveness but are predominantly tailored for offline analysis. This intrinsic limitation poses a significant risk of overlooking real-time anomalies during actual drone operations. My project or the approach in the research paper seeks to rectify this critical issue by introducing an innovative approach grounded in Long Short-Term Memory (LSTM) and deep learning. The goal is to enable the online detection of flight log anomalies that could potentially lead to physical instabilities.  
Three distinct sets of flight logs from DJI, ArduPilot, and PX4 will be used to execute this project.

The overarching aim is to mitigate false alarms, thereby significantly improving anomaly detection accuracy.

* **Motivation and literature review:**

The motivation behind choosing this particular project stems from a practical and firsthand experience at the Inter IIT Sports Meet. Witnessing a drone malfunction during an event where drones were utilized for capturing videos highlighted the challenges and potential risks associated with drone operations. The incident served as a catalyst for recognizing the need for a more sophisticated and proactive approach to drone management.

Engaging with the event organizers and understanding the various issues drones can face, such as sensor malfunctions and configuration errors, emphasized the significance of addressing these challenges promptly. The motivation, therefore, is rooted in the desire to enhance the reliability and safety of drone operations, especially in contexts where drones play crucial roles, such as filming or surveillance. The goal is not merely to detect problems but to develop a real-time system that can identify issues, allowing for immediate intervention and ensuring drones' uninterrupted and secure functioning.

* **How is it related to the field of AI?**

The project leverages AI/ML, particularly Long Short-Term Memory (LSTM), a type of recurrent neural network (RNN), to enable online detection of anomalies in drone flight logs. The chosen datasets for training and testing comprise flight logs from various sources, including DJI, ArduPilot, and PX4, which consists of both closed and open-source control programs.

The AI/ML model, implemented using the LSTM architecture, is designed to adapt to new environmental conditions dynamically. It achieves this by incrementally learning new standard patterns during runtime, allowing it to effectively address the evolving nature of drone operations and mitigate false alarms. The model's evaluation involves key metrics such as recall, precision, and F-measure, providing a comprehensive assessment of its effectiveness in detecting anomalies like high vibration levels,sensor calibration errors, actuator faults, Magnetic field interference, Gps Signal interference, Communication issues.

This application focuses on predicting the next state of a drone based on sequential feature vectors with a primary emphasis on anomaly detection in drone flight logs.

* **LSTM Model for Sequential Prediction:**
  + Collects flight logs containing essential parameters like timestamp, flight status, and state units indicating the drone's physical conditions.
  + Implementation of “DronLomaly,”, a LSTM-based deep learning model designed to predict the next state based on sequential flight log data. LSTM is chosen for its ability to learn long-term dependencies over sequences and the short term state helps in the output of current time step.
* **Anomaly Detection Mechanism:**
  + Applying a sliding window approach to create input sequences for the LSTM model, each containing recent log entries.
  + An anomaly is reported if the actual state significantly deviates from the predicted state(made by the LSTM model). This is a crucial step for anomaly detection. The corresponding state unit is flagged and correlation analysis is performed on the flagged states.
* **Normalization and Model Training:**
  + Normalizing(making the mean as 0 and standard deviation as 1) the input sequences by using the average and standard deviation of the corresponding parameter position from the training data. This step ensures consistency in data.**(Z -score normalization)**
  + **Dropout** is a regularization technique that is used to prevent overfitting. In this, a certain percentage of units are dropped off during training. Dropout is applied to the data of sequences. Dropout is applied to the hidden layers of the bidirectional LSTM layer.
  + Training the LSTM model by minimizing the error between predicted and observed feature vectors using Mean Squared Error (MSE) loss. Iteratively adjust model weights for a specified number of epochs. Hyperparameter tuning helps in further refinement of the model.
* **Correlation Analysis for Refining Anomaly Detection:**
  + Incorporation of correlation analysis to refine anomaly detection. Utilization Pearson correlation analysis on flagged and other state units to differentiate between actual anomalies and new stable states.
  + A correlation threshold is decided and if the correlation change is larger than the threshold, an anomaly is reported.
* **Real-world Application and Adaptive Learning:**
  + Implementing the developed model to detect physical instabilities during drone flight missions. This real-world application focuses on ensuring safety and stability in activities such as aerial surveillance, package delivery, and search and rescue operations.
  + Integrating an adaptive mechanism into the model to address false positives and enhance its performance. This adaptive learning approach proves crucial in accommodating changing environmental conditions or flight scenarios, ensuring the model's accuracy in predicting flight states and promptly detecting anomalies during drone operations.
* **Model Persistence and Integration:**
  + Save the trained prediction model, which is represented as a multidimensional weight vector.
  + Implementing the LSTM model using Docker for easy integration into the DroneLogGuard system.

**To implement the LSTM part of the paper,**

1. **Data Collection:** Aggregating flight data from DJI, ArduPilot, and PX4 where DJI is a closed source control program whereas ArduPilot and PX4 are open source control programs . Each log entry should encompass vital details, including timestamps, flight status, and state units reflecting the drone's physical conditions.
2. **DronLomaly Implementation:** Develop and implement DronLomaly, an LSTM-based deep learning model trained on the collected flight logs. LSTM is chosen for its ability to predict the next possible state based on a sequence of past states, effectively learning long-term dependencies over sequences.The short term state also helps in the output of a current time stamp.
3. **Sliding Window Approach:** Utilizing a sliding window approach to create input sequences for the LSTM model. Each sequence shall contain a set of recent log entries. The model's output is the predicted next flight state, and any significant deviation between the actual and predicted states triggers an anomaly report.
4. **Z-Score Normalization:** Normalizing the values in each vector of the input sequences by using the average and standard deviation of the corresponding parameter position from the training data. This step ensures consistency and improved model performance.
5. **Regularization:** Dropout technique is incorporated for regularization to prevent overfitting. A dropout of 0.1 is mentioned in the paper for the dataset that means 10% of the units will be dropped out during training.
6. **Training the LSTM Model:** Training the LSTM model using normalized input sequences, with the objective of minimizing the error between predicted and observed feature vectors through the Mean Squared Error (MSE) loss function. The activation function employed is the Rectified Linear Unit (ReLU) activation function. The model weights are iteratively adjusted over a specified number of epochs, facilitating a robust learning process.

* **Brownie points for secondary goals. (Limitations)**

There are a few limitations of this approach. I am mentioning some of them-  
1) The optimal sliding window size for anomaly detection and the impact of using different thresholds for reporting anomalies were not investigated in detail.

2) The approach focused on detecting 7 types of fault data in the logs but didn’t consider other types of faults.

3)The paper doesn’t investigate the impact of varying configuration parameter values like Sequence Length, LSTM Model Architecture, Epoch, Deviation Threshold, Correlation Threshold etc.

In my future work, I would like to overcome these limitations.  
For the 1st limitation, we can **experiment with different window sizes and check on the effectiveness (recall, precision and F-measure).**For the 3rd limitation, we can do **parameter sensitivity analysis,** that means by changing 1 parameter at a time, we can check the effectiveness and then record the findings.

* **Varying Sequence length to find the optimal sequence Length (Sensitivity Analysis)**

To explore the impact of varying sequence length on Recall, Precision, and F-measure, I propose plotting three curves to analyze the trends and identify an optimized sequence length. The chosen approach involves taking interval steps of 15 in sequence length up to a point where the graph exhibits a discernible trend, either increasing or decreasing.

The procedure is as follows:

1. **Define Sequence Length Range:**

- Set the initial sequence length of 3 and establish an interval step of 15.

- Will determine the upper limit for the sequence length based on the observed trends in the graph.

2. **Generate Curves:**

- Plot three curves separately for Recall, Precision, and F-measure against varying sequence length.

- Calculate these metrics for each sequence length configuration.

3. **Identify Trends:**

- Analyze the curves to identify trends in Recall, Precision, and F-measure as sequence length varies.

- Observe whether the metrics exhibit increasing, decreasing, or optimal trends.

4. **Locate Optimal Sequence Length:**

- Identify the point where the graph attains a maximum or optimal value for each metric.

- If a maximum is found, consider narrowing down the sequence length range around that point for detailed analysis.

5. **Detailed Analysis:**

- Enlarge the graph around the identified optimal sequence length.

- Plot the graph at smaller intervals within this range to pinpoint the exact sequence length that maximizes Recall, Precision, or F-measure.

6. **Graphical Representation:**

- Visualize the results through clear and labeled graphs, emphasizing the relationship between sequence length and each performance metric.

7. **Conclusion:**

- Draw conclusions based on the graphical analysis to recommend an optimized sequence length that balances Recall, Precision, and F-measure effectively.

-Consider trade-offs and practical implications in choosing the sequence length that balances Recall, Precision, and F-measure effectively.

* **Goals till mid-term evaluation**

***Week 1:***

**1. Data Preprocessing:**

* Gather resources for learning the necessary tech stack. (Techstack- Python, Pytorch, Keras, Docker, Pandas, Numpy, Matplotlib/Seaborn)
* Collect relevant datasets for training and testing. First, I will do it on one of the dataset then as time permits, I will include the other 2 datasets.
* Normalize data for consistency and effective model training.
* Generate sequences from the normalized data.

***Week 2:***

**2. LSTM Model Development:**

* Design the LSTM architecture for the model.
* Implement the layers and define the loss function.
* Choose an appropriate optimizer for efficient model training.
* Compile the LSTM model for further development.

***Mid-term Evaluation:***

* Have a functional LSTM model capable of processing sequential drone flight log data.
* Successfully preprocess and normalize datasets, ensuring compatibility with the model.
* Demonstrate the ability to generate meaningful sequences from the preprocessed data.
* Implement the core components of the LSTM model, including architecture, output layer, and loss function.
* **Goals till end-term evaluation**

***Week 3:***

**3. Model Training and Evaluation:**

* Set training parameters for the LSTM model.
* Train the model using the prepared datasets.
* Evaluate the model's performance using key metrics such as recall, precision, and F-measure.
* Conduct hyperparameter tuning for optimal model performance.
* Starting the work of Sensitivity Analysis(Varying Sequence Length to find the Optimal Sequence Length)

***Week 4:***

**4. Bug Resolution and Feature Optimization:**

* Completing Sensitivity Analysis
* Identify and resolve any bugs or issues in the developed LSTM model.
* Optimize additional features to enhance the model's accuracy and efficiency.
* Implementing the above procedure to other 2 datasets also.
* ***Deployment using Docker*:** Implement deployment procedures using Docker for easy integration into the DroneLogGuard system.(if time permits)

***End-term Evaluation:***

* Present a fully trained and evaluated LSTM model capable of detecting anomalies in real-time drone flight logs.
* Demonstrate bug-free functionality and optimized features for improved accuracy.
* Successfully deploy the model using Docker, showcasing its readiness for integration into the DroneLogGuard system.(if time permits)

* **References**

1. <https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=8548&context=sis_research>
2. **Datasets-**  
    DJI ( <https://auto.dji.com/ad4che-dataset> )

ArduPilot(<https://ardupilot.org/copter/docs/common-downloading-and-analyzing-data-logs-in-mission-planner.html> )  
PX4 ( <https://docs.px4.io/main/en/log/flight_log_analysis.html> )

* **About you and your experience with ML/DL**

I'm Archi, a sophomore studying Mathematics and Computing at IIT Roorkee. I reside in Ahmedabad, Gujarat and my hobbies include playing badminton, listening to music etc. I Ventured into Machine Learning (ML) and Deep Learning (DL) around five to six months ago, I developed a strong foundation through the DSG Supervised Learning Challenge which focused on classification tasks.I have a good foundation of ML/DL basics. Currently immersed in an engaging project involving Natural Language Processing (NLP) and optimal techniques for text summarization, I'm integrating Text-to-Speech (TTS) technologies to convert summaries into audio output, bridging NLP and audio technologies. Beyond academics, my ML/DL journey aligns with a keen interest in the subject, driven by curiosity to explore and contribute to this dynamic field's possibilities. Excited about future challenges and real-world applications, I strive to solve real-world problems with the help of my knowledge in the domain of mathematics and machine learning. I feel I can complete this project in a month as I have a good foundation both in mathematics and ML/DL. Also, I don’t have any current engagements so I can give my full time to this project.