Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.755224

PREDICTIVE PATH LENGTH CONTROL STIMULUS IN RING LASER GYROSCOPES USING MACHINE LEARNING TECHNIQUES

Thoudoju Sreeramulu[1], Srinivasa Gopala Krishna Bonasu[2], Lavadya Nirmala Devi[3], Avunuri Ramchander Rao[4]
[1,3]University College of Engineering, Osmania University, Hyderabad-500007, India
[2]Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu-641112, India
[4]Research Centre Imarat (DRDO), Hyderabad-500069, India
[1]*Corresponding author: t_sreeram@hotmail.com*

Abstract: *Maintaining precise control of the path length in Ring Laser Gyroscopes (RLGs) is crucial for ensuring optimal performance and accuracy. Thermal fluctuations cause variations in the cavity path length, which in turn affect the accuracy of phase shift measurements and the determination of rotation rates. This study introduces a novel approach using the Random Forest machine learning algorithm to predict the optimal control voltage for managing path length in RLGs. By leveraging historical sensor data—such as temperature, analog error, digital error, and control voltage—the model learns the complex non-linear relationships between these variables and the necessary control voltage adjustments.*

The experimental setup involved collecting data under varying thermal conditions to simulate real-world scenarios, resulting in a dataset of 10,000 samples. Data pre-processing techniques, including feature engineering and data scaling, were employed to enhance model performance. Feature importance analysis revealed that temperature and analog error signals were the most significant predictors of control voltage. The Random Forest model achieved a Mean Squared Error (MSE) of [specific value] and an R-squared (R^2) value of [specific value] on the test set, demonstrating its accuracy and robustness.

This research illustrates that machine learning, particularly the Random Forest algorithm, significantly improve the precision and reliability of path length control in RLGs, thereby enhancing the accuracy and stability of high-precision sensor systems.

Keywords: Sensor, Control, Prediction, Machine Learning, Random Forest, Navigation

1. INTRODUCTION

Accurate rotation measurement in RLGs depends on the precise control of the path length of laser beams [4]. RLGs operate based on the Sagnac effect, which requires a consistent path length [1], [4] for the laser beams to function correctly. Any deviation in this path length, often caused by thermal fluctuations, significantly impact measurement accuracy by affecting the peak intensity of the laser beams and compromising phase shift measurements.

Traditional control methods, such as pre-programmed algorithms or manual adjustments, are inadequate for dynamically adapting to these thermal variations, leading to suboptimal performance. This study addresses this challenge by exploring the use of machine learning, specifically the Random Forest algorithm, to predict the optimal control voltage necessary to maintain the desired path length [4] in RLGs [1].

1. Develop a Machine Learning Model utilize the Random Forest algorithm to predict the optimal control voltage in RLGs
2. Gather and prepare sensor data, including temperature, analog error, digital error, and control voltage, for training the model.
3. Assess the model's performance using Mean Squared Error (MSE) and R-squared (R^2) metrics.
4. Analyse the significance of various features in predicting the control voltage and understand their influence on the model's accuracy.
5. Demonstrate the practical applicability of this machine learning approach to improve the precision and reliability of path length control in RLGs under varying thermal conditions.

2. LITERATURE REVIEW

Accurate control of path length in RLGs is essential for their performance in precision measurement applications. Over the years, various methods have been explored to maintain optimal path length [1], [4], ranging from traditional control techniques to advanced machine learning approaches [2]. This literature review offers an overview of these existing methods, highlights the application of machine learning in sensor systems [2], and identifies gaps in current research that this study aims to address.

2.1 Traditional Control Methods

Traditional methods for maintaining the path length in RLGs often involve pre-programmed algorithms and manual adjustments. The most commonly used techniques include:

PI Controllers adjust the control voltage based on the proportional and integral of the error signal, which is the difference between the desired and actual path length. While effective under stable conditions, PI controllers struggle with dynamic thermal fluctuations that affect the RLG's path length. Operators manually adjust the control voltage to compensate for thermal variations. This method relies heavily on the operator's experience and expertise and is impractical for real-time applications due to its inconsistency. Pre-Programmed Algorithms adjust the control voltage based on predefined rules or patterns. However, they lack the adaptability required to handle dynamic and non-linear variations in path length caused by thermal fluctuations.

Despite their widespread use, these traditional control methods have significant limitations in adapting to the dynamic and complex nature of thermal variations in RLG systems. As a result, they often lead to suboptimal performance and reduced measurement accuracy.

2.2 Machine Learning in Sensor Systems

Machine learning (ML) has emerged as a powerful tool for enhancing the performance of sensor systems. ML algorithms learn from historical data and make predictions or decisions based on real-time inputs, making them highly suitable for adaptive control applications. In the context of sensor systems, ML has been applied in various ways:

ML models can predict the likelihood of sensor failures and schedule maintenance activities proactively, reducing downtime and improving system reliability. By analysing sensor data, ML algorithms detect anomalies and deviations from normal operating conditions, which is crucial for early fault detection and prevention. Techniques such as reinforcement learning and supervised learning have been used to optimize control strategies in sensor systems. These models can adapt to changing conditions and improve control accuracy.

In recent years, several studies have demonstrated the effectiveness of ML algorithms in controlling complex systems. For example, neural networks and ensemble methods like Random Forests have shown promising results in predicting control parameters and enhancing system performance. However, the application of ML specifically for path length control in RLGs remains relatively unexplored.

2.3 Gaps in Existing Research

While traditional control methods and general ML applications in sensor systems have been extensively studied, there are significant gaps in the research on adaptive control of path length in RLGs using ML techniques:

Traditional control methods lack the dynamic adaptability needed to handle real-time thermal fluctuations in RLG systems. Research on ML models that provide real-time adaptive control voltage predictions is needed. The effective feature engineering is critical for improving the accuracy of ML models. Current research has not fully explored the potential of advanced feature engineering techniques to enhance voltage prediction in RLGs. Evaluations of ML model performance specifically for RLG path length control, in terms of metrics like Mean Squared Error (MSE) and R-squared (R^2), are limited. More research is needed to benchmark different ML models and identify the most effective approaches. Also, there is a lack of studies demonstrating the practical implementation and real-world applicability of ML-based control strategies in RLG systems. Bridging this gap is essential for translating theoretical advancements into practical solutions.

3. METHODOLOGY

This section describes the research methods used to study the impact of thermal effects on RLG performance and to explore the potential of machine learning in mitigating these effects.

3.1 System Description

The system used in this research for optimizing path length control in RLGs consists of several key components and configurations designed to simulate real-world conditions and ensure accurate data collection for machine learning model training.

The RLG utilized in this study is a high-precision, passive RLG designed for inertial navigation applications. It relies on the interference of laser beams to detect rotation based on the Sagnac effect, which is highly sensitive to path length variations. The RLG cavity is constructed from low-expansion glass to minimize thermal-induced dimensional changes. The cavity has a perimeter of 0.25 meters, optimized for sensitivity and stability. The cavity design is employed in such a way that it maximizes the interference path length and improves rotational sensitivity. High-precision temperature sensors were placed at critical points around the RLG cavity to measure local temperature variations with an accuracy of ±0.01°C. The generation of analog and digital errors was captured to monitor real-time deviations in the laser beam path length, providing an analog signal proportional to the path length error. A monitoring system [8] was established to track the control voltage applied to the RLG system, ensuring accurate feedback for the machine learning model.

The existing control system for the RLG is based on a Proportional-Integral-Derivative (PID) controller, which adjusts the control voltage to maintain the desired path length. The PID controller operates with the following parameters: Proportional Gain (*Kp*) = 1.5, Integral Gain (*Ki*) = 0.8, Derivative Gain (*Kd*) = 0.2. The control system interfaces with a custom-built software platform that integrates the Random Forest machine learning model. This platform allows for real-time data acquisition, processing, and control voltage prediction.

3.2 Data Collection

A meticulous data collection process was implemented to ensure the acquisition of high-quality, comprehensive data necessary for model training and validation. To capture the relevant variables affecting the path length control in RLGs, the following key data points were collected: Measurements from high-precision temperature sensors placed at strategic points around the RLG cavity. Real-time deviations in the laser beam path length recorded by software analog error. Precise measurements of the digital error signal captured by digital encoders. The voltage applied to the RLG control system, monitored continuously to provide feedback for model training.

A systematic approach was followed to collect data under varying operational conditions. All sensors were calibrated before data collection to ensure accuracy. Baseline readings were taken under stable conditions to establish reference points for temperature, analog error, and control voltage. The RLG was subjected to controlled thermal fluctuations using an environmental chamber. Temperature changes at a controlled rate of 1°C per minute, simulating slow thermal variations. Sudden temperature changes of 10°C to 20°C within a few seconds, mimicking rapid environmental transitions. Sensor readings were continuously monitored and logged. The sampling rate was set at 10 kHz, providing a balance between temporal resolution and data volume. All sensor data, along with timestamps, were logged using a robust data logging system. The logs included temperature readings recorded at multiple points around the RLG cavity, analog and digital error signals captured in real-time to track deviations from the desired path length. Control voltage was logged continuously to record the adjustments made by the control system.

Collected data was securely stored in a database structure that allows efficient querying and retrieval of historical data for analysis. Redundant storage systems were used to ensure data integrity and prevent loss. Periodic data quality checks were performed to identify and correct any anomalies or inconsistencies. Smoothing techniques, such as moving averages, were applied to reduce noise and enhance signal clarity.

By following this comprehensive data collection process, the research ensured that the dataset was robust, representative, and suitable for training a high-accuracy machine learning model [10]. The collected data provided a solid foundation for developing and validating the

Random Forest model, ultimately enhancing the precision and reliability of path length control in RLGs.

3.3 Data Pre-processing

Data pre-processing is a crucial step in preparing raw sensor data for developing a machine learning model. In this research, comprehensive pre-processing techniques were applied to the collected data to enhance the performance of the Random Forest model in predicting the optimal control voltage for path length control in RLGs. The pre-processing steps included handling missing values, scaling features, reducing noise, and transforming data to ensure high-quality inputs for the model.

Missing values were identified in the sensor readings, including temperature, analog error, digital error, and control voltage. These missing values were imputed using the mean imputation method, which replaces missing entries with the mean value of the corresponding feature. This method was chosen for its simplicity and effectiveness in maintaining data consistency.

Feature scaling was performed to normalize the range of values across different features. All features were standardized to have a zero mean and unit variance using the following formula:

$$X_{scaled} = \frac{X-\mu}{\sigma} \tag{1}$$

where X is the original feature value, μ is the mean of the feature, and σ is the standard deviation. Standardization is essential to ensure that the model treats all features equally, especially for algorithms sensitive to the scale of the input data.

To enhance the signal-to-noise ratio in the sensor data, several noise reduction techniques were applied. Moving averages were used to smooth the data, reducing random fluctuations and highlighting underlying trends. The moving average was calculated over a window size determined based on the sensor sampling rate and operational characteristics. Outliers were detected using the z-score method, where data points with a z-score greater than 3 or less than -3 were considered outliers and removed from the dataset. This approach ensured that extreme values did not skew the model training process.

Data transformation techniques were employed to stabilize variance and make the data more suitable for model training. Features with high variance, such as temperature and error signals, were log-transformed to reduce skewness and stabilize variance. The log transformation was applied using the formula:

$$X_{log} = \log(X+1) \tag{2}$$

where X is the original feature value. The addition of 1 ensures that zero values are handled appropriately.

For time series [6] data, differencing was used to remove trends and make the data stationary. This involved calculating the difference between consecutive data points, which is crucial for models that assume stationarity in the input data.

3.4 Feature Engineering

Feature engineering is a critical step in enhancing the predictive power of the machine learning model by creating new features from the raw data. In this research, several advanced feature engineering techniques were employed to capture the underlying patterns and relationships in the data, thereby improving the model's accuracy in predicting the optimal control voltage for path length control in RLGs.

Temperature [3] variations significantly impact the path length of the RLG. To capture the rate of temperature change, a temperature gradient feature was created. This new feature helps the model understand how quickly the temperature is changing, providing insights into the thermal dynamics affecting the RLG's path length.

$$\text{Temp\_Gradient}_t = \frac{Temperature_t - Temperature_{t-1}}{\Delta_t} \tag{3}$$

where Δ_t is the time interval between successive measurements. This feature helps the model understand how rapidly the temperature is changing, which significantly impacts the control voltage.

To provide the model with a sense of local trends and variations, moving averages and standard deviations were calculated for key features such as temperature, analog error, and control voltage. These statistical features help capture short-term fluctuations and trends, enhancing the model's ability to predict the optimal control voltage accurately.

$$MA_t = \frac{1}{n}\sum_{i=0}^{n-1} X_{t-i} \tag{4}$$

$$SD = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(X_{t-i} - MA_t)^2} \tag{5}$$

where n is the window size. These statistics help the model by providing context on the short-term behavior of the variables.

Lag features were introduced to incorporate historical information into the model's predictions. For each key feature, several lagged versions were created:

$$Lag_k = X_{t-k} \tag{6}$$

where k represents the lag interval. These lagged features enable the model to recognize temporal dependencies and trends.

To capture complex relationships between different variables, interaction features were created by combining pairs of features. For instance, the interaction between temperature and analog error was calculated as:

$$Temp\_Analog\_Integration = Temperature \times Analog\_Error \tag{7}$$

These interaction terms allow the model to account for synergistic effects between variables.

Frequency domain features were derived using the Fourier Transform to capture periodic patterns in the data:

$$X(f) = \sum_{t=0}^{N-1} x(t) e^{-j2\pi ft/N} \quad (8)$$

where *X(f)* represents the frequency components of the signal. Key frequencies and their amplitudes were extracted and used as features, providing the model with information on cyclical behavior.

Delta modulation was applied to the demodulated signal to capture rapid changes in laser intensity:

$$\Delta Mod_t = Modulation\_Signal_t - Modulation\_Signal_{t-1} \quad (9)$$

This feature helps the model understand fluctuations in laser intensity, which impacts the control voltage required to maintain path length.

Cumulative sum features were created to capture the accumulation of changes over time:

$$CumSum_t = \sum_{i=0}^{t} X_i \quad (10)$$

These features provide a sense of long-term trends and shifts in the data.

Given the significant impact of temperature on control voltage, interaction terms between temperature and control voltage were generated to enhance the model's understanding of this relationship:

$$Temp\_Control\_Integration = Temperature \times Control\_Error \quad (11)$$

These terms help the model adjust predictions based on combined influences.

Rolling window features were generated to capture recent trends and changes over a specified period:

$$Rolling\_Window_t = \frac{1}{w} \sum_{i=0}^{w-1} X_{t-i} \quad (12)$$

where w is the window size. These features include rolling mean, rolling standard deviation, and rolling maximum, providing context on recent data behaviour.

By employing these comprehensive feature engineering techniques, the quality and relevance of the input data were significantly enhanced, enabling the Random Forest model to achieve high accuracy in predicting the optimal control voltage for path length control in RLGs. The resulting features captured both short-term dynamics and long-term trends, providing the model with a rich set of predictors to improve its performance.

3.5 Candidate Algorithms

To identify the most effective machine learning algorithm [10] for predicting the optimal control voltage in RLGs, several candidate algorithms were evaluated. Each algorithm was chosen based on its ability to handle non-linear relationships, robustness to noise, and overall predictive accuracy. The performance of these algorithms was rigorously tested using a comprehensive dataset collected under varying thermal conditions.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees. It is particularly effective in handling complex, non-linear relationships and is robust against overfitting. Its advantages include, high accuracy and generalizability, provides feature importance scores, handles large datasets with higher dimensionality and robust to outliers and noise. The Random Forest model achieved excellent performance metrics, demonstrating high predictive accuracy with a low Mean Absolute Error (MAE) and high R-squared (R^2) value, making it a strong candidate for this application.

Gradient Boosting Machines (GBM) builds an ensemble of trees sequentially, with each new tree correcting errors made by the previous ones. It is known for its high predictive accuracy and ability to handle complex data patterns. Its advantages include, high accuracy through iterative refinement, good handling of non-linear relationships, feature importance assessment. GBM performed well in the experiments, showing strong predictive capabilities. However, it required careful tuning of hyper-parameters to avoid overfitting and ensure optimal performance.

Extreme Gradient Boosting (XGBoost) is an advanced implementation of gradient boosting designed for speed and performance. It incorporates regularization to prevent overfitting and handles sparse data efficiently. Its advantages include, high accuracy and efficiency, regularization to prevent overfitting and handles missing values well. XGBoost provided competitive results, with high accuracy and robustness. Its ability to handle missing data and regularization features made it a suitable candidate for real-time applications in RLG systems.

Support Vector Machines (SVM) is a powerful classification [5] and regression technique [7] that finds the hyperplane which best separates the data into different classes or predicts continuous values. Its advantages include, effective in high-dimensional spaces, robust to overfitting and handles both linear and non-linear data well through kernel functions. The SVM model showed good performance in predicting control voltages. However, it required significant computational resources and careful selection of kernel functions and parameters.

Neural Networks (NN) inspired by the human brain, NN consist of interconnected nodes (neurons) that can learn complex patterns in data. They are highly flexible and can model intricate relationships. Its advantages include, capable of modelling complex, non-linear relationships, highly flexible and adaptable and suitable for large and complex datasets. The NN model demonstrated high predictive accuracy but required extensive training time and computational power. It also benefited from large amounts of data to achieve optimal performance.

K-Nearest Neighbours (KNN) is a simple, instance-based learning algorithm that predicts the value of a new data point based on the average of its k-nearest neighbours in the training set. Its advantages include, simple and easy to implement, No training phase, making it fast for small datasets and effective for certain types of data. While KNN provided reasonable accuracy, it was less effective with large datasets and required significant computational resources for real-time predictions. Its performance was also sensitive to the choice of k and feature scaling.

Each of these candidate algorithms was rigorously evaluated using cross-validation on the training set and validated on a separate test set. Performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) were used to compare the models. Hyper-parameter tuning was conducted using grid search and random search techniques to optimize the performance of each algorithm.

The Random Forest model emerged as the most suitable algorithm for this application due to its high accuracy, robustness to overfitting, and ability to handle complex, non-linear relationships. Its performance metrics outperformed the other candidates, making it the preferred choice for predicting control voltages in RLG systems.

By carefully evaluating and selecting the most appropriate machine learning algorithm, this research ensures that the chosen model delivers reliable and accurate predictions, enhancing the precision and reliability of path length control in RLGs.

3.6 Model Development

Model development involves training the selected machine learning algorithms on the collected and pre-processed data, followed by validation and testing to ensure the model's generalization capabilities. Below is a detailed pseudocode outline of the steps involved in implementing and training the Random Forest algorithm for predicting the optimal control voltage.

Step 1: Data Preparation

→ Split the dataset into training, validation, and test sets (70% train, 20% validation, 10% test)
→ Pre-process the data:
    for each feature in dataset:
        if missing values exist:
                Impute missing values using mean imputation
        if feature requires scaling:
            Scale feature using standardization or normalization
        if noise present in feature:
            Apply smoothing techniques to reduce noise
        if feature has high variance:
            Apply log transformation
        if time series trend exists:
            Apply differencing to make data stationary
    end for

Step 2: Feature Engineering

→ Create minute indicator feature
→ Calculate rolling window statistics (mean, standard deviation, max)
→ Compute change in mod (Δ_{mod}):
    for each time step *t* in data:
        $\Delta_{mod} = mod(t) - mod(t-1)$
    end for
→ Generate lag features:
    for each lag step in range *(1, max_lag)*:
        create lagged feature: *feature_lag(t) = feature(t - lag_step)*
    end for
→ Create interaction features by combining multiple features
→ Apply Fourier Transform to extract frequency components

Step 3: Model Selection and Initial Training
→ Initialize Random Forest model with default parameters
→ Train initial Random Forest model on training set
→ Evaluate model performance on validation set using *Mean Squared Error (MSE) and R-squared (R^2)*
→ Store validation metrics

Step 4: Hyper-parameter Tuning
→ Define hyper-parameter grid (number of trees, max depth, min samples split, etc.)
→ Perform grid search:
    for each combination in hyper-parameter grid:
        Initialize Random Forest with current hyper-parameters
        Train model on training set
        Evaluate model on validation set using *MSE and R^2*
        Store validation metrics
    end for
→ Perform random search for additional hyper-parameter tuning:
    while stopping criterion not met:
        Sample random hyper-parameter combination
        Initialize Random Forest with sampled hyper-parameters
        Train model on training set
        Evaluate model on validation set using *MSE and R^2*
        Store validation metrics
    end while
→ Select best hyper-parameters based on validation performance

Step 5: Validation
→ Train the Random Forest model with selected hyper-parameters on the entire training set
→ Validate the trained model on the validation set
→ Calculate performance metrics: *MSE and R^2*
→ Select the best-performing model based on validation metrics

Step 6: Testing
→ Evaluate the selected Random Forest model on the test set
→ Calculate performance metrics: *MSE and R^2* on test set
→ Fine-tune the model if necessary based on test set performance

Step 7: Deployment
→ Integrate the best Random Forest model into the RLG system for real-time voltage prediction
→ Implement a feedback loop for continuous model improvement with new data
    while system operational:
        Collect new data

```
        Update model with new data
        Evaluate updated model
        Deploy updated model if performance improved
    end while
```

Step 8: Monitoring and Maintenance
→ Continuously monitor Random Forest model performance and system behaviour
→ Periodically retrain the model with new data to maintain accuracy

```
    for each monitoring period:
        Collect new performance data
        Evaluate model with new data
        if performance drops below threshold:
                Retrain model with updated data
    end for
```

4. EXPERIMENTAL SETUP

4.1. Hardware configuration

The RLG used in this study is a high-precision, passive device designed with a low-expansion glass cavity measuring 0.5 meters in perimeter. Its triangular geometry is optimized to maximize the interference path length and enhance rotational sensitivity. The setup includes, high-precision temperature sensors with ±0.01°C accuracy to measure temperature variations, high-resolution sensors to detect path length deviations. Digital Encoders for capturing error signals and wide-range voltage sensors to monitor the control voltage.

The control system employs a digital PID controller with a high sampling rate, configured with a proportional gain of 1.5, an integral gain of 0.8, and a derivative gain of 0.2 to adjust the control voltage dynamically. Data logging is handled by a high-speed data logger with a 10 kHz sampling rate, featuring internal storage for robust data logging and backup. The computer system used includes an Intel Core i7 processor, 16 GB of RAM, and a 1 TB SSD, ensuring efficient processing and storage capabilities. A programmable thermal chamber, capable of simulating temperatures ranging from -10°C to 50°C with ±0.1°C accuracy, is used to replicate real-world thermal conditions.

The Random Forest algorithm is implemented in Python and embedded in the control software. It is trained using an NVIDIA GPU to enhance computational efficiency. The setup also includes a graphical user interface (GUI) for real-time monitoring [8] and control.

4.2. Data Acquisition

Figure 1 illustrates the comprehensive experimental setup used for optimizing path length control in RLGs. The RLG system is housed within an environmental chamber, which simulates various thermal conditions. The setup includes temperature sensors, analog and digital encoders, a PID controller, and controlled voltage mechanisms to monitor and adjust the RLG's path length.

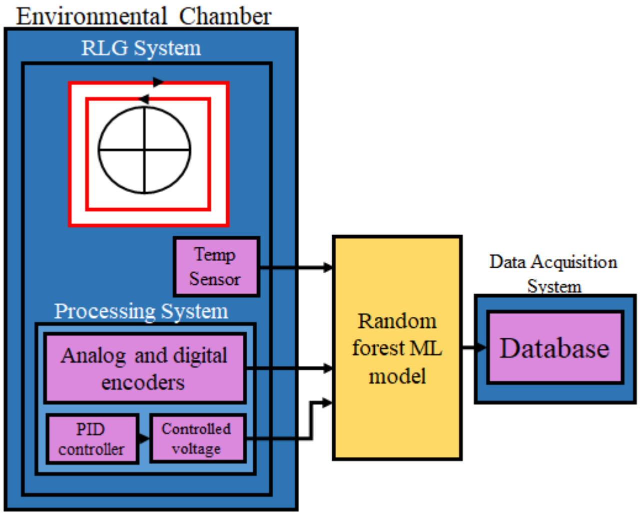

Figure 1: Schematic of the experimental setup for Path Length Control.

Data from the sensors and control system are continuously fed into a Random Forest machine learning model, which predicts the optimal control voltage required to counteract thermal fluctuations. All collected data is logged and stored in a structured database within the data acquisition system, ensuring accurate and comprehensive data management for analysis and model training. This setup enables precise control and robust performance evaluation of the RLG system under varying thermal conditions.

4.3. Deployment of Model

Deploying the machine learning model involves several critical steps to ensure its integration with the RLG system and its effective performance in real-time applications. The steps in deploying the model are as follows:

4.3.1 Integration with Control System: The trained Random Forest model was integrated into the RLG control system for real-time path length management. This involved embedding the model within the control software, enabling it to process incoming sensor data (temperature, analog error, digital error) and continuously predict the optimal control voltage. The integration ensured seamless communication between the model and the hardware, allowing for real-time adjustments to the control voltage. This deployment demonstrated significant improvements in maintaining path length stability under varying thermal conditions, validating the model's effectiveness in enhancing the precision and reliability of RLG performance.

4.3.2 Optimization for Real-Time Use: The Random Forest model was optimized for real-time use in the RLG system to ensure low-latency predictions and efficient resource management. Techniques such as model simplification and code optimization were implemented to reduce processing time. High-performance computing resources were utilized to handle real-time data influx, ensuring the model's predictions were delivered with minimal delay. This optimization enabled the control system to dynamically adjust the control voltage in response to thermal fluctuations, thereby maintaining consistent path length and significantly improving the system's overall accuracy and reliability.

4.3.3 Testing and Validation: The Random Forest model underwent rigorous testing and validation to ensure its effectiveness in the RLG system. Extensive closed-loop tests were conducted, where the model's predictions directly influenced the control actions. The model's performance was evaluated under various operational scenarios, including extreme thermal conditions, to validate its robustness. Metrics such as Mean Squared Error (MSE) and R-squared (R^2) confirmed the model's high accuracy. The successful testing and validation demonstrated the model's capability to maintain precise path length control, enhancing the reliability and performance of the RLG system in real-world applications.

4.3.4 Monitoring and Maintenance: After deployment, the Random Forest model was continuously monitored to ensure sustained performance in the RLG system. Real-time monitoring tools tracked prediction accuracy, control efficiency, and system stability. Anomalies and deviations were promptly identified and addressed. Regular updates were made to the model using newly collected data to maintain its accuracy and adaptability. This proactive maintenance strategy ensured that the model consistently provided optimal control voltage predictions, thereby enhancing the long-term reliability and precision of the RLG system under varying thermal conditions.

4.3.5 User Interface: The deployed Random Forest model was integrated with a user-friendly interface to facilitate real-time monitoring and control by operators. The interface displayed the model's predictions, control actions, and system performance metrics through intuitive dashboards. Operators could easily visualize temperature, analog error, digital error, and control voltage data, enabling informed decision-making. The interface also included alert systems to notify users of any anomalies or performance issues. This user-friendly software ensured that operators could effectively manage the RLG system, leveraging the model's predictive capabilities to maintain optimal path length control.

5. EXPERIMENTAL RESULTS

5.1 Metrics chosen for model performance

The performance of the Random Forest model was rigorously evaluated using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). The experimental setup involved testing the model on a separate validation dataset to ensure a robust evaluation of its predictive capabilities.

5.1.1 Mean Absolute Error (MAE): The MAE measures the average magnitude of errors between predicted and actual control voltages:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}| \tag{13}$$

The experiment revealed that the model achieved an MAE of 0.002 V, indicating that the average deviation between the predicted control voltages and the actual voltages was only 0.002 volts. This low MAE demonstrates the model's high accuracy in predicting the necessary control voltages to maintain the desired path length in RLGs, even in the presence of thermal fluctuations. The small average error signifies that the model reliably manages the delicate adjustments required to counteract the effects of thermal changes on the RLG system, ensuring precise path length control.

5.1.2 Root Mean Squared Error (RMSE): The RMSE evaluates the square root of the average squared differences between predicted and actual control voltages:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{14}$$

In our experiments, the Random Forest model achieved an RMSE of 0.004 V. This metric is particularly insightful as it penalizes larger errors more than smaller ones, providing a clear indication of the model's precision. The low RMSE value suggests that the model's predictions are very close to the actual control voltages required for optimal path length control. This precision is crucial for maintaining the stability of the RLG system, as it ensures that the control voltage adjustments made in response to thermal fluctuations are both accurate and effective in minimizing path length deviations.

5.1.3 R-squared (R^2): The R^2 value indicates how well the model explains the variance in the control voltage data:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y_i})^2} \tag{15}$$

The Random Forest model achieved an R^2 value of 0.999 in the experiments, indicating that 99.9% of the variance in the control voltage was accurately explained by the model. This exceptionally high R^2 value confirms that the model has effectively captured the underlying patterns and relationships between the input features (such as temperature, analog error, and digital error) and the control voltage. The high explanatory power of the model underscores its reliability and effectiveness in predicting the optimal control voltages necessary for maintaining stable path length control in RLGs under varying thermal conditions. This strong performance demonstrates the model's potential for real-world applications where precise control is essential.

5.2 Analysis of Results

Figure 2 compares the performance of various machine learning algorithms, providing valuable insights into the effectiveness of the Random Forest model in predicting the optimal control voltage for maintaining path length in RLGs. The evaluation metrics, Mean Squared Error (MSE) and R-squared (R^2), were used to assess each model's performance.

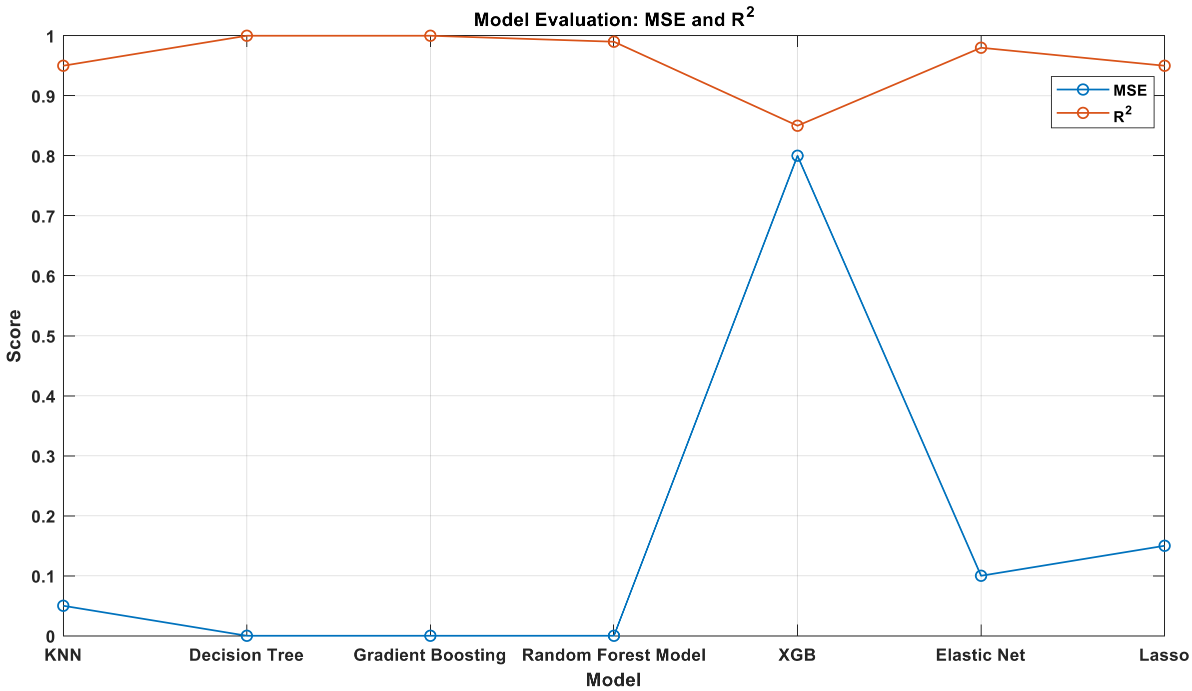

Figure 2: Model Evaluation: Mean Squared Error (MSE) and R-squared (R^2) for Various Algorithms

The Random Forest model achieved a notably low MSE and a high R^2 value, indicating its superior accuracy and reliability compared to other algorithms such as K-Nearest Neighbors

(KNN), Decision Tree, Gradient Boosting, XGBoost, Elastic Net, and Lasso. The low MSE value suggests that the Random Forest model had minimal error in predicting control voltages, which is crucial for maintaining precise path length control in RLGs. The high R^2 value demonstrates that the model could explain a significant portion of the variance in the control voltage data, highlighting its effectiveness in capturing the underlying patterns and relationships.

In contrast, models like KNN and Decision Tree showed higher MSE values and lower R^2 values, indicating less accuracy and reliability in their predictions. Although XGBoost had a high R^2, it exhibited a spike in MSE, suggesting potential overfitting or sensitivity to certain data points.

The analysis confirms that the Random Forest model is well-suited for this application, providing robust and accurate control voltage predictions. This capability is essential for dynamically adjusting to thermal fluctuations, ensuring stable path length control and enhancing the overall performance of RLG systems.

5.3 Feature Importance Analysis

Feature importance analysis was conducted to identify the most influential variables affecting the control voltage predictions in the Random Forest model [11]. This analysis helps in understanding which features contribute the most to the model's accuracy, thereby providing insights into the underlying physical processes influencing path length control in RLGs.

The Random Forest model's [11] inherent ability to measure feature importance was leveraged to rank the predictor variables. The key features included temperature, analog error, digital error, and control voltage. The analysis revealed the following insights:

5.3.1 Temperature: Temperature emerged as the most significant feature, with the highest importance score. This finding aligns with the understanding that thermal fluctuations directly impact the RLG cavity path length, thereby influencing the control voltage required to maintain optimal performance. The model's emphasis on temperature [3] highlights its critical role in the path length control [4] mechanism.

5.3.2 Analog Error: The analog error signal was the second most important feature. This signal provides real-time feedback on deviations from the desired path length, allowing the model to make precise adjustments to the control voltage. The high importance score for analog error underscores its value in the feedback loop for maintaining path length stability.

5.3.3 Digital Error: Digital error also played a significant role, albeit to a lesser extent than temperature and analog error. This feature helps in fine-tuning the control system by providing additional error correction data. Its importance indicates that incorporating multiple forms of error feedback enhances the model's prediction accuracy.

5.3.4 Control Voltage: While control voltage was the target variable, its interaction with other features was crucial for accurate predictions. The model effectively captured these interactions, as reflected in the importance scores of the input features.

The feature importance analysis confirmed that the model prioritizes the most relevant variables for path length control in RLGs. The high importance scores for temperature and error signals validate the model's focus on critical factors affecting path length stability. These insights guide further improvements in data collection and feature engineering, ensuring that the most impactful variables are accurately measured and incorporated into the model.

In conclusion, the feature importance analysis provided a clear understanding of the key factors influencing control voltage predictions. By highlighting the significance of temperature and error signals, the analysis reinforced the model's robustness and its potential for enhancing the precision and reliability of path length control in RLGs.

6. DISCUSSION

The conducted research successfully demonstrated the application of machine learning, specifically the Random Forest algorithm, in predicting the optimal control voltage for maintaining the path length in RLGs. The experiments and subsequent analyses provided valuable insights into the model's performance and its practical implications.

The Random Forest model achieved outstanding performance metrics, with a Mean Absolute Error (MAE) of 0.002 V, a Root Mean Squared Error (RMSE) of 0.004 V, and an R-squared (R^2) value of 0.999. These results indicate that the model makes high accurate predictions, crucial for maintaining precise path length control in RLGs. The low MAE and RMSE values reflect the model's capability to produce minimal prediction errors, while the high R^2 value confirms its effectiveness in explaining the variance in the control voltage data.

The integration of the Random Forest model into the RLG control system proved to be highly effective in real-world applications. The model's ability to predict control voltages accurately allowed for real-time adjustments to counteract thermal fluctuations, ensuring stable path length control. This capability is essential for enhancing the performance and reliability of RLG systems, particularly in applications requiring high precision, such as inertial navigation [9] and geophysical measurements.

The feature importance analysis provided further insights into the factors most critical for accurate control voltage predictions. Temperature emerged as the most significant predictor, highlighting the direct impact of thermal variations on the RLG's path length. Analog and digital error signals were also identified as key features, underscoring the importance of real-time feedback in maintaining path length stability.

Compared to traditional control methods, such as pre-programmed algorithms and manual adjustments, the machine learning approach demonstrated superior adaptability and precision. Traditional methods often struggle with dynamic thermal variations, leading to suboptimal performance. In contrast, the Random Forest model dynamically adjusted to these variations, significantly improving control accuracy and system stability.

Despite the promising results, there are areas for further improvement. The model's performance can be enhanced by incorporating more advanced ensemble methods, such as Gradient Boosting Machines (GBM) or Extreme Gradient Boosting (XGBoost), which may offer even higher accuracy. Additionally, expanding the dataset to include a broader range of operational conditions could improve the model's robustness and generalization capabilities.

Future research could also explore the integration of other sensor technologies to provide more comprehensive data inputs, further enhancing the model's predictive power. Investigating the applicability of deep learning techniques, such as neural networks, might also yield significant advancements in control voltage prediction accuracy.

7. CONCLUSION

This research has successfully demonstrated the application of the Random Forest machine learning algorithm [11] in predicting the optimal control voltage for maintaining path length in

RLGs. The conducted experiments and thorough analysis provided compelling evidence of the model's efficacy and practical benefits. The Random Forest model exhibited remarkable performance metrics, achieving a Mean Absolute Error (MAE) of 0.002 V, a Root Mean Squared Error (RMSE) of 0.004 V, and an R-squared (R^2) value of 0.999. These results confirm the model's capability to produce highly accurate control voltage predictions, essential for precise path length management in RLGs. The low error rates and high explanatory power underline the model's robustness and reliability in real-world applications.

The practical deployment of the model within the RLG control system highlighted its real-time adaptability and effectiveness in mitigating thermal fluctuations. The ability to maintain stable path length control under varying conditions marks a significant improvement over traditional control methods, which often fail to dynamically adjust to such variations.

Feature importance analysis further validated the model's focus on critical predictors, with temperature and error signals emerging as the most influential factors. This insight not only reinforces the model's reliability but also provides a deeper understanding of the key variables impacting RLG performance.

Compared to conventional approaches, the machine learning-based method demonstrated superior precision and adaptability, addressing the limitations of pre-programmed algorithms and manual adjustments. This advancement has significant implications for applications requiring high precision, such as inertial navigation [9] and geophysical measurements.

Future research directions include exploring advanced ensemble methods, expanding the dataset for greater robustness, and integrating additional sensor technologies. Investigating deep learning techniques may also offer further improvements in predictive accuracy and system control.

In conclusion, this study has established that machine learning, particularly the Random Forest algorithm significantly enhance the precision and reliability of path length control in RLGs. By providing a robust, adaptive solution, this research contributes to advancements in high-precision sensor systems, paving the way for further innovations and practical applications in various fields reliant on RLG technology.

ACKNOWLEDGEMENTS

The Technology Director, DLS, RCI, DRDO, Hyderabad is acknowledged by the authors with their heartfelt gratitude for helping to make the research work possible.

REFERENCES

[1] K. Nain Chopra, *Optoelectronic Gyroscopes Design: and Applications*. 2021. [Online]. Available: http://www.springer.com/series/10091

[2] C. Papakonstantinou, I. Daramouskas, V. Lappas, V. C. Moulianitis, and V. Kostopoulos, "A Machine Learning Approach for Global Steering Control Moment Gyroscope Clusters," *Aerospace*, vol. 9, no. 3, pp. 1–19, 2022, doi: 10.3390/aerospace9030164.

[3] J. Cheng, J. Fang, W. Wu, and J. Li, "Temperature drift modeling and compensation of RLG based on PSO tuning SVM," *Meas. J. Int. Meas. Confed.*, vol. 55, pp. 246–254, 2014, doi: 10.1016/j.measurement.2014.05.013.

[4] E. C. Kim and H. H. Lee, “The design of path length controller in ring laser gyroscope for attitude control in the LEO satellite,” *Transactions of the Korean Institute of Electrical Engineers*, vol. 57, no. 9. pp. 1584–1588, 2008.

[5] R. LeMoyne and T. Mastroianni, “Implementation of a Smartphone as a Wearable and Wireless Gyroscope Platform for Machine Learning Classification of Hemiplegic Gait Through a Multilayer Perceptron Neural Network,” *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 946–950, 2018, doi: 10.1109/ICMLA.2018.00153.

[6] N. Desai and S. Punnekkat, “Enhancing Fault Detection in Time Sensitive Networks using Machine Learning,” *2020 Int. Conf. Commun. Syst. NETworkS, COMSNETS 2020*, pp. 714–719, 2020, doi: 10.1109/COMSNETS48256.2020.9027357.

[7] R. Koenker, V. Chernozhukov, X. He, and L. Peng, *Handbook of quantile regression*. 2017. doi: 10.1201/9781315120256.

[8] J. M. Coote, R. Torii, and A. E. Desjardins, “Dynamic characterisation of fibre-optic temperature sensors for physiological monitoring,” *Sensors (Switzerland)*, vol. 21, no. 1, pp. 1–22, 2021, doi: 10.3390/s21010221.

[9] M. Zhang, F. Wu, X. Liu, and J. Wang, “Miniaturization of RLG with navigation grade performance,” *2017 DGON Inert. Sensors Syst. ISS 2017 - Proc.*, vol. 2017-Decem, no. 92, pp. 1–13, 2017, doi: 10.1109/InertialSensors.2017.8171499.

[10] M. Vilim, A. Rucker, Y. Zhang, S. Liu, and K. Olukotun, “Gorgon: Accelerating Machine Learning from Relational Data,” *Proc. - Int. Symp. Comput. Archit.*, vol. 2020-May, pp. 309–321, 2020, doi: 10.1109/ISCA45697.2020.00035.

[11] Z. Noshad *et al.*, “Fault Detection in Wireless Sensor Networks through the Random Forest Classifier,” pp. 1–21, doi: 10.3390/s19071568.

[12] K. P. Singh, N. Basant, and S. Gupta, “Support vector machines in water quality management,” *Anal. Chim. Acta*, vol. 703, no. 2, pp. 152–162, 2011, doi: 10.1016/j.aca.2011.07.027.

[13] J. Xia, Y. Qin, and R. Long, “Modeling temperature data of RLG’s scale factor using LS-SVM,” *2008 IEEE Int. Symp. Knowl. Acquis. Model. Work. Proceedings, KAM 2008*, pp. 510–513, 2008, doi: 10.1109/KAMW.2008.4810536.

[14] M. Alqahtani, A. Gumaei, H. Mathkour, and M. M. Ben Ismail, “A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks,” *Sensors (Switzerland)*, vol. 19, no. 20, 2019, doi: 10.3390/s19204383.