

Prediction of Motion Acceleration under Newton's Second Law: A Multivariate Linear Regression Approach

Dusadee Khamboonrueang¹, Jantakarn Sangnoi¹, Thawatchai Promrat^{2*}

¹ Major of Physics and General Sciences, Department of Science, Faculty of Science and Technology, Nakhon Sawan Rajabhat University, Nakhon Sawan 60000, Thailand

² Major of Information Technology, Department of Science, Faculty of Science and Technology Phranakorn Si Ayutthaya Rajabhat University 13000, Thailand

*Corresponding author e-mail: thawatchai@aru.ac.th

Received: 9 March 2026 / Revised: 8 August 2026 / Accepted: 10 October 2026

Abstract

One of the most significant lessons in basic physics is Newton's Second Law of Motion, which teaches the relationship between the acceleration of an object, force, and mass. However, we found that the problem is that some schools do not have enough experimental equipment for students, causing teaching to focus only on theory and the memorization of content. This research aims to predict the relationship between acceleration caused by force and mass using machine learning with a multivariate linear regression algorithm. We have prepared three datasets from laboratory rooms, from image analyses obtained from laboratory rooms with the Tracker program, and from all of those. After that, we will train and compare models for each dataset using linear regression algorithms. The experimental R-squared values are 99% for dataset one, 99% for dataset two, and 97% for dataset three. Therefore, the model provides relatively stable generalization under each individual experimental condition. It can be used as a tool to help predict physics results in situations where real experiments are not feasible due to time, budget, or equipment constraints.

Keywords: Movement, Multivariate linear regression algorithm, Machine learning

1. Introduction

Teaching physics students at the high school and university levels presents several significant challenges, particularly in developing students' conceptual understanding of fundamental physical principles and their ability to connect theoretical concepts with real experimental observations. Laboratory-based learning plays an important role in physics education, students can understand physical phenomena, collect and interpret experimental data, evaluate theoretical models and develop scientific reasoning skills. These experiences are particularly important in mechanics where students must establish meaningful relationships among physical quantities such as displacement, velocity, acceleration, force, and mass. One fundamental topic that students often find difficult to understand is Newton's Second Law of Motion, which describes the relationship between force, mass, and acceleration. Although the theoretical relationship ($F = ma$) is straightforward, understanding how this relationship is manifested in an actual physical system requires students to connect mathematical representations with experimental observations. Differences between ideal theoretical predictions and experimental measurements may arise from friction, pulley inertia, air resistance, measurement uncertainty, and other non-ideal characteristics of laboratory apparatus. Consequently, experimental activities remain important for helping students distinguish between idealized physical models and the behavior of real systems.

However, opportunities for hands-on experimentation may be limited by insufficient laboratory equipment, restricted instructional time, equipment costs, large class sizes, or limited access to laboratory facilities. In some educational institutions, the number of experimental setups may be insufficient for all students to conduct experiments individually or in small groups. Therefore, students have fewer opportunities for direct observation and repeated experimentation. These limitations can make it difficult for students to investigate physical relationships systematically and compare theoretical predictions with actual experimental measurements. Currently, technology is used to support laboratory activities and computational tools are becoming increasingly

important as complementary approaches for providing students with additional opportunities to investigate physical phenomena and analyze experimental data. Among these approaches, video analysis has become an effective and accessible method for teaching mechanics. Video-based analysis allows motion to be recorded and subsequently analyzed frame by frame, enabling students to connect real-world motion with quantitative representations such as position–time, velocity–time, and acceleration–time data (Khamboonruang, 2021; Poonyawatpornkul & Luksameevanish, 2019; Wee & Leong, 2015). Wattanakasiwich and Poonyawatpornkul (2012), for example, demonstrated the application of high-speed video analysis to mechanics experiments and discussed its usefulness for obtaining experimental motion data as well as identifying potential sources of measurement error. Similarly, Brown and Cox (2009) described innovative applications of video analysis in physics instruction, demonstrating its potential for connecting experimental observations with quantitative physical analysis. The Tracker Video Analysis and Modeling Tool has been particularly useful in this context because it provides an open-source environment for extracting kinematic information from recorded motion. Tracker allows users to determine quantities such as position, velocity, and acceleration and to compare experimental measurements with mathematical models. Wee et al. (2012) demonstrated the use of Tracker as a pedagogical tool for projectile motion and showed that combining video analysis with computer modeling can support active and inquiry-based learning by allowing students to compare their conceptual models with observations from real-life motion. Claessens (2017) further demonstrated that Tracker can be used not only with experimental videos but also with virtual physics simulations to perform kinematic analysis and fit theoretical models to experimentally generated data. More recent studies have continued to demonstrate the educational value of Tracker-based motion analysis. Renika et al. (2024) applied Tracker video analysis to accelerated motion on an inclined plane and free-fall experiments. Their results showed that Tracker could be used to determine kinematic parameters directly from recorded motion and reveal deviations in acceleration associated with non-ideal effects such as air resistance. Such findings are particularly relevant to Newtonian mechanics because they demonstrate that experimentally observed acceleration may deviate from ideal theoretical values even when the underlying physical relationship is well established.

Despite these advantages, video-analysis approaches still require an experimental or simulated motion record as input. In a typical Tracker workflow, a physical experiment must first be conducted and recorded before the resulting video can be analyzed. Therefore, although Tracker facilitates the extraction and analysis of quantitative information from experiments, it does not completely address situations in which laboratory equipment, cameras, experimental setups, or sufficient instructional time are unavailable. Furthermore, repeated experiments under different combinations of force and mass may require considerable laboratory time and resources, particularly when many students need access to the same experimental equipment. These limitations create an opportunity for data-driven predictive modeling to complement conventional laboratory experiments and video-analysis approaches. A predictive model trained on actual experimental data may provide estimates of physical quantities under different experimental conditions without requiring the physical experiment to be repeated for every combination of variables, while still preserving patterns learned from real experimental observations. The increasing integration of computational methods and machine learning into physics education provides new possibilities for using experimental datasets not only for retrospective analysis but also for predictive modeling. Recent developments have demonstrated that machine-learning methods can be integrated with low-cost experimental systems to enable students to collect physical measurements, construct computational models, and investigate relationships contained in experimental data. For example, Liu et al. (2025) developed an affordable experimental platform integrating Internet-of-Things technology and machine-learning methods for applied physics education. Their work demonstrated how experimental data acquisition, preprocessing, model training, and validation can be incorporated into a unified educational workflow, illustrating the potential of data-driven approaches for connecting experimental physics with computational modeling.

From this perspective, regression-based predictive models are particularly appropriate when the objective is to estimate a continuous physical quantity from experimentally measured variables (Montgomery et al., 2012; Qu, 2024). Rather than replacing established physical laws, such models can learn relationships contained in real experimental datasets, including systematic effects associated with the specific apparatus and measurement process. This characteristic is important in Newton's Second Law experiments because the ideal relationship

$a = F/m$ does not explicitly account for friction, pulley inertia, air resistance, and measurement uncertainties that may influence experimentally measured acceleration.

Therefore, this research proposes the use of Multiple Linear Regression (MLR) in machine learning algorithms as a data-driven predictive approach for modeling acceleration from experimental variables associated with Newton's Second Law (Farizal et al., 2021; Kasemsan & Maksuwan, 2023; Tabyang et al., 2022). Three datasets were considered: data obtained directly from laboratory experiments, data derived from video analysis using Tracker, and a combined dataset incorporating both sources. The models were trained to predict acceleration based on relevant experimental variables such as force and mass and, where applicable, measurements obtained through video analysis. In practical terms, the proposed model is intended to function as a lightweight prediction tool that can complement conventional laboratory experiments and Tracker-based activities. Given specified experimental conditions such as force and mass, a trained model can estimate acceleration based on patterns learned from real experimental observations. Consequently, the prediction reflects not only the ideal theoretical relationship but also, to the extent represented in the training data, the non-ideal characteristics of the experimental apparatus.

This approach is particularly relevant for educational settings with limited resources because, once trained, the model does not require students to have continuous access to the original laboratory apparatus or video-analysis environment. The predicted data can instead be used as experimentally grounded examples through which students can compare theoretical acceleration calculated from $a = F/m$ with acceleration predicted from observations of an actual experimental system. The proposed model should therefore be regarded as a low-cost complement to physical experimentation and video analysis rather than a replacement for hands-on laboratory learning. It is important to clarify that the contribution of this study is not to rediscover or verify Newton's Second Law, since the relationship among force, mass, and acceleration is already well established. Instead, the study investigates whether a predictive model trained on experimental and video-derived data can represent the discrepancy between an idealized theoretical model and the behavior of a real experimental apparatus. As demonstrated in Section 3.1, direct application of $a = F/m$ to the force and mass values used in the laboratory trials resulted in substantial deviations from the measured acceleration, with a mean absolute percentage error of approximately 25–29% and a negative coefficient of determination ($R^2 < 0$). These discrepancies are likely associated with non-ideal effects such as friction, pulley inertia, and other experimental losses.

In contrast, the trained MLR models achieved R^2 values of approximately 97–99% against the measured data. Therefore, the principal contribution of this study lies in demonstrating how a relatively simple and interpretable data-driven model can capture experimentally observed behavior that is not represented by the idealized equation alone. The resulting model may subsequently be reused as an equipment-independent predictive tool to support Newton's Second Law instruction in educational settings where direct experimentation is temporarily or practically constrained.

2. Materials and Methods

2.1 Steps in developing an acceleration forecasting model using a multivariate regression algorithm

The process of developing a linear regression model for predicting acceleration is illustrated in Figure 1. The process begins with motion data collection in the laboratory using experimental equipment. The recorded videos are then analyzed using Tracker software to extract numerical motion data. Subsequently, the collected data are examined to understand their structure, characteristics, and quality. At this stage, correlation analysis.

The obtained data is then prepared and organized into three datasets. Data set 1 consists of data obtained from experiments in which the force is kept constant, combined with the corresponding motion data extracted using Tracker software. Data set 2 consists of data obtained from experiments in which the mass is kept constant, together with the corresponding Tracker data. Data set 3 is formed by combining Data sets 1 and 2. Each data set is subsequently used to train a linear regression model. The performance of the developed models is evaluated using statistical metrics such as the coefficient of determination (R^2), mean absolute error (MAE), Mean Squared Error (MSE), and Absolute Percent Error (MAPE) (Settakhumpoo & Benjaoran, 2022). Finally, the evaluation results are compared to select the model with the best predictive performance.

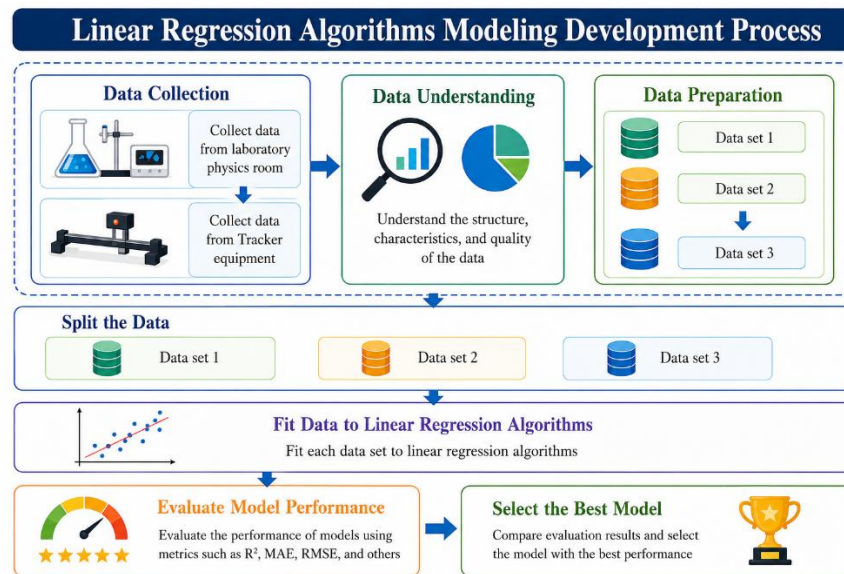


Figure 1. Process for establishing models to estimate motion.

The relationship between the predictor variables and the predicted acceleration can be expressed using the multivariate linear regression model shown in Equation (1) (Boonkong & Hormdee, 2018) :

$$v_n = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \quad (1)$$

This resulted in three models in total. The next step is to measure their performance and select the most appropriate model to predict the acceleration of the experimental motion according to Newton's Second Law of Motion.

2.2 Data collection for developing motion acceleration forecasting models

Data were collected to develop a model for predicting motion acceleration. This research used two data collection methods to obtain diverse data for developing a linear regression model.

Data collection method 1: Data were collected using a Newton's Second Law of Motion experiment set in the laboratory. The equipment used in conducting the research consisted of a Newton's Second Law of Motion experiment set, the equipment for the Newton's Second Law of Motion experiment, and a mobile phone camera to record video during the experiment using a VIVO V20 mobile phone with a rear camera resolution of 64 million pixels, as shown in Figure 2.

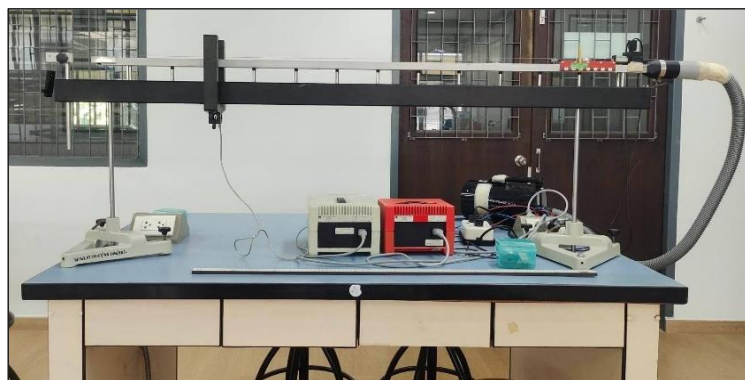


Figure 2. Set up an experiment of Newton's second law.

$$a = 2S/t^2 \quad (2)$$

(2) The second experiment is to find the relationship between acceleration (a) and mass (m) when the force is constant. The purpose was to study how the acceleration changed when the net force was constant and the mass of the object changed. In this experiment, we start by setting the constant tensile force F and using a slide with a mass of 196.7 grams. The switch was pressed, and the slide was allowed to move 1 meter. The time was recorded and the time was recorded 5 times to find the average. Then, add the mass of the slide uniformly for the next four values (total 5 data sets). Each set was repeated. The acceleration (a) was calculated from the same equation and the reciprocal of the mass of $1/m$ (the reciprocal of the mass) was calculated. Then, take the values of $1/m$ and a were used to create a graph to show the linear relationship according to the equation $F = ma$.

The screenshot shows a video analysis software interface. The main window displays a video of a physics experiment setup. A horizontal beam is supported by a stand, and a small cart is on it. A red digital scale is visible on the table. The software interface includes a menu bar (File, Edit, Video, Track, Coordinates System, View, Help), a toolbar with various icons, and a status bar at the bottom. On the right side, there are two plots: 'x (m)' vs 't (s)' and 'y (m)' vs 't (s)'. The 'x (m)' plot shows a red curve representing the horizontal position over time. The 'y (m)' plot shows a red curve representing the vertical position over time. Below the plots, there is a table with columns 'x (m)' and 'y (m)'. The table contains data points for the tracked object. A small window titled 'Track' is open in the top left corner, showing a list of tracks: 'New', 'aria A', and 'aria B'. The 'aria A' track is selected. A status bar at the bottom indicates '100%' zoom and 'World View'.

Figure 3 shows that when the experimental video was analyzed with the Tracor program, 50 data points and 50 data points from actual laboratory experiments were obtained, totaling 100 data points. Three attributes were included: force and mass, average time, and acceleration. To make the composition of the three datasets used for model training fully reproducible, Table 1 summarizes, for each dataset, the data source, the independent and

dependent variables, the experimental condition under which the data were generated, and the total number of samples.

Table 1. Summary of the three datasets used for model development.

Dataset	Data source	Independent variable (s)	Dependent variable	Experimental condition	Total samples (n)
Dataset 1	Laboratory experiment (Table 2) + Tracker video analysis (Table 3)	Force (F); mass (m) held constant	Acceleration (a)	Mass held constant; force varied across 5 levels, each averaged over 5 repeated trials in the laboratory, matched with Tracker-derived measurements of the same trials	50 (25 Lab + 25 Tracker)
Dataset 2	Laboratory experiment (Table 4) + Tracker video analysis (Table 5)	Mass (m); force (F) held constant	Acceleration (a)	Force held constant; mass varied across 5 levels, each averaged over 5 repeated trials in the laboratory, matched with Tracker-derived measurements of the same trials	50 (25 Lab + 25 Tracker)
Dataset 3	Pooled combination of Dataset 1 and Dataset 2	Force (F); mass (m)	Acceleration (a)	Pooled data from both the force-varied and mass-varied conditions above	100 (50 Lab+ 50 Tracker)

2.3 Development of a motion acceleration prediction model

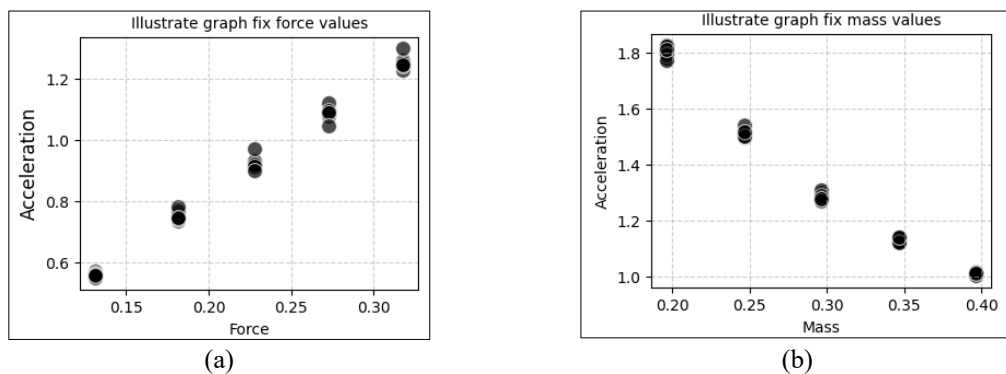


Figure 4. Understanding the fixed force and mass data.

Figure 4 shows the experimental data, which is in a linear form. Figure (a) shows the experimental data with force as a constant value, and Figure (b) shows the mass set as a constant value. The graph shows that the obtained data is in a linear form.

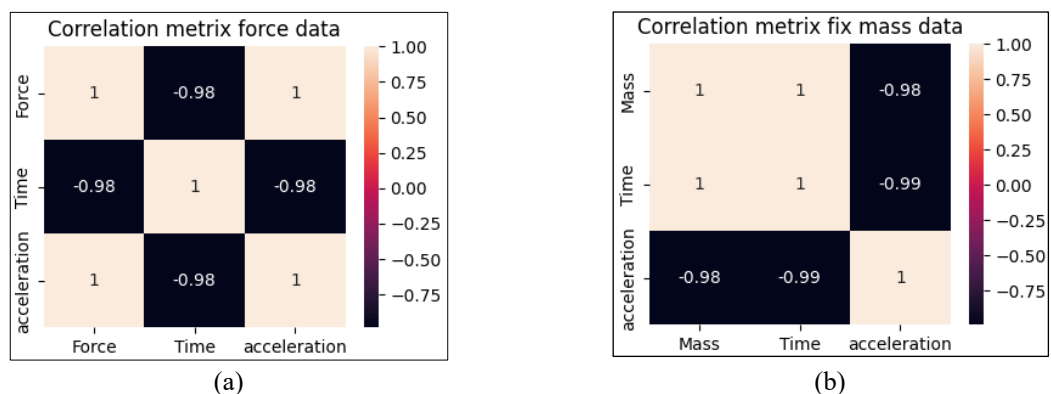


Figure 5. Data correlation of the force and mass data.

The data from Figure 5 have been used to consider the relationship between data variables by finding the correlation coefficient of two variables using the Pearson Correlation Coefficient equation, as shown in Equation 3.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (3)$$

When considering the relationship between each attribute, it can be seen that the values obtained were negative, meaning that the values obtained were opposite, for example, the acceleration value was large when the mass value was small, and the acceleration value was small when the mass value was large, which is consistent with the data in Figure 4.

2.4 Evaluation of model performance

The performance of the predictive models was evaluated using the Mean Squared Error (MSE) and the coefficient of determination (R^2), as defined in Equations (4) and (5), respectively. MSE was used to quantify the prediction error, while (R^2) was used to assess how well the model explained the variation in the observed acceleration. In addition, the agreement between the experimental results and the idealized theoretical baseline was assessed using the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), following the error-evaluation approach described by Prapaporn et al. (2019). MAE quantifies the average absolute deviation in the original measurement units, whereas MAPE expresses the deviation relative to the theoretical values as a percentage. For both metrics, lower values indicate closer agreement with the idealized theoretical baseline.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

MSE	Mean squared error
n	Number of data points
y_i	Observed values
$\hat{y}_i$	Predicted values

$$R^2 = 1 - \frac{RSS}{TSS} \quad (5)$$

R^2	Coefficient of Determination
RSS	Residual Sum of Squares
TSS	Total Sum of Squares

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (6)$$

MAE	Mean Absolute Error
Y_i	Actual value
$\hat{Y}_i$	Predicted value
n	Number of observations
$ Y_i - \hat{Y}_i $	Absolute error

$$MAPE = \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100\% \quad (7)$$

MAPE Mean Absolute Percentage Error
 Y_i Actual dependent variable values
 $\hat{Y}_i$ Predicted dependent variable results given by the model
 n the number of observations

2.5 Software, regression framework, and verification of model assumptions

This research utilized Jupyter Notebook, running Python 3.12, along with the scikit-learn library, to develop the training data and implement the linear regression algorithm.

Table 2. Multiple linear regression model training and validation configuration.

No.	Component	Method/Configuration	Description
1	Model	Multiple Linear Regression	The regression model was implemented using the Linear Regression estimator provided by Scikit-learn.
2	Data Splitting	Random holdout	90% training set and 10% testing set
3	Model Validation	5-Fold Cross-Validation	Divides the data into five subsets and uses each subset for validation
4	Loss Function	Mean Squared Error (MSE)	Measures and compares training and validation errors
5	Randomization Control	shuffle=True, random_state=42	Shuffles the data and ensures reproducibility

As presented in Table 2, the multiple linear regression model was implemented using the Linear Regression estimator provided by Scikit-learn. The dataset was randomly divided into training and testing sets using a 90:10 holdout ratio. Five-fold cross-validation was employed to evaluate the model across different data subsets, with each subset used once for validation while the remaining subsets were used for training. Mean Squared Error (MSE) was used to quantify and compare the training and validation errors. During learning-curve construction, the data were shuffled, and a fixed random state of 42 was specified to ensure the reproducibility of the results.

3. Results and Discussion

The experimental results are presented in four stages. First, the acceleration data obtained from the laboratory experiments and Tracker software are analyzed. Second, the three datasets are prepared. Third, multiple linear regression (MLR) models are developed, and their predictive performance is evaluated. Finally, the trained models are compared with the idealized theoretical relationship ($a = F/m$).

3.1 Experimental and Tracker-based acceleration results

Two experimental conditions were investigated. In the first condition, the mass was kept constant while the applied force was varied. In the second condition, the force was kept constant while the mass was varied. For each condition, acceleration was determined from both the laboratory measurements and video analysis using Tracker software.

Tables 3 and 4 present representative results for the constant-mass condition. As the applied force increased, the measured acceleration generally increased. The acceleration values obtained using Tracker showed a similar trend to those obtained directly from the laboratory experiment.

Table 3. Sample experimental data for the constant-mass condition.

Force (N)	Time (s)	Acceleration (m/s ²)
0.131	1.911	0.547
0.181	1.650	0.734
0.227	1.481	0.911
0.272	1.361	1.078
0.317	1.266	1.247

Note: Values shown are from trial (repetition) 1 of 5 at each of the 5 force levels; the model was trained on all 5 repetitions (25 rows total; see Table 1).

Table 4. Sample Tracker data for the constant-mass condition.

Force (N)	Time (s)	Acceleration (m/s ²)
0.131	1.872	0.571
0.181	1.616	0.766
0.228	1.464	0.933
0.273	1.352	1.094
0.318	1.248	1.284

Note: values shown are from trial (repetition) 1 of 5 at each of the 5 force levels, obtained from Tracker video analysis of the same runs summarized in Table 1; the model was trained on all 5 repetitions (25 rows total; see Table 1).

Tables 5 and 6 present representative results for the constant-force condition. In this case, increasing the mass generally resulted in a decrease in acceleration, which is consistent with the expected inverse relationship between mass and acceleration.

Table 5. Sample experimental data for the constant-force condition.

Mass (kg)	Time (s)	Acceleration (m/s ²)
0.197	1.049	1.816
0.247	1.152	1.508
0.297	1.248	1.285
0.347	1.333	1.125
0.397	1.412	1.003

Note: Values shown are from trial (repetition) 1 of 5 at each of the 5 mass levels; the model was trained on all 5 repetitions (25 rows total; see Table 1).

Table 6. Sample Tracker data for the constant-force condition.

Mass (kg)	Time (s)	Acceleration (m/s ²)
0.197	1.060	1.780
0.247	1.156	1.497
0.297	1.236	1.309
0.347	1.336	1.121
0.397	1.412	1.003

Note: Values shown are from trial (repetition) 1 of 5 at each of the 5 mass levels, obtained from Tracker video analysis of the same runs summarized in Table 3; the model was trained on all 5 repetitions (25 rows total; see Table 1).

Overall, the laboratory measurements and Tracker-derived values exhibited similar trends under both experimental conditions. These results indicate that the Tracker-based measurements can provide motion information that is consistent with the laboratory observations and can therefore be incorporated into the datasets used for model development.

3.2 Predictive performance of the MLR models

Three datasets were used to develop MLR models. Dataset 1 represented the force-varied, constant-mass experiments; Dataset 2 represented the mass-varied, constant-force experiments; and Dataset 3 combined the observations from Datasets 1 and 2.

The predictive performance of the three models was evaluated using the mean squared error (MSE) and coefficient of determination (R^2). The results are summarized in Table 7.

Table 7. Comparison performance model.

Model	Data set	Mean Square Error (MSE)	R squared (R^2)
1	Data set 1	0.0003%	0.99%
2	Data set 2	0.0006%	0.99%
3	Data set 3	0.0063%	0.97%

The models trained using Datasets 1 and 2 achieved the highest predictive performance, with (R^2) values of approximately 0.99. The model developed from the combined Dataset 3 also demonstrated strong predictive performance, achieving an (R^2) of approximately 0.97. The corresponding MSE values were approximately 0.0003, 0.0006, and 0.0063 for Models 1, 2, and 3, respectively. Thus, Models 1 and 2 provided the smallest prediction errors, whereas the pooled-data model showed a modest reduction in performance. This difference is reasonable because Dataset 3 combines observations obtained under two different experimental conditions, increasing the variability that must be represented by a single regression model. These findings demonstrate that the MLR models can effectively capture the relationships among force, mass, time, and acceleration observed in the experimental system.

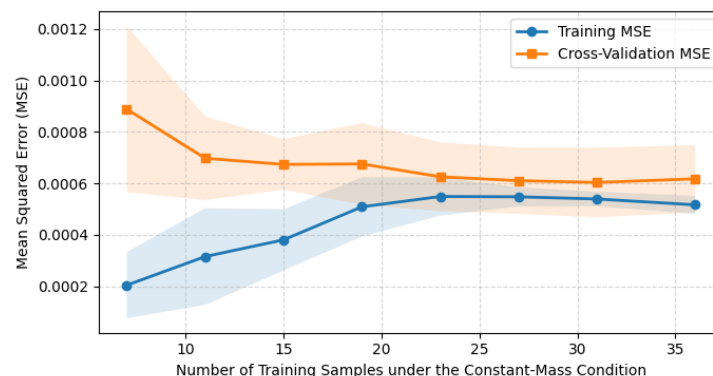


Figure 6. MLR learning curve under constant mass.

Figure 6. shows the learning curve of the MLR model under the constant-mass and varying-force condition. As the number of training samples increases, the training MSE gradually increases, whereas the cross-validation MSE generally decreases. The gap between the two curves becomes smaller as the training set grows, with both curves approaching MSE values of approximately 0.0005–0.0006 at larger training sizes. In addition, the variability indicated by the shaded regions generally decreases as more training data is used. These results suggest that the model achieves relatively stable performance with a small generalization gap at larger training sizes, indicating limited evidence of overfitting under this experimental condition.

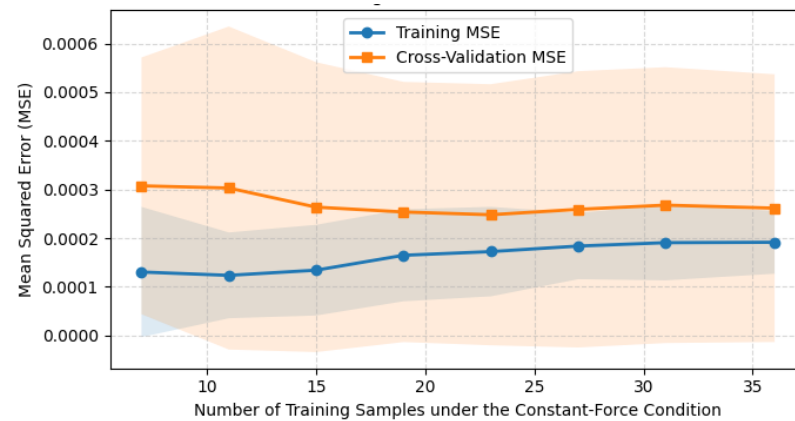


Figure 7. MLR learning curve under constant force.

Figure 7. shows the learning curve of the MLR model under the constant-force and varying-mass condition. As the number of training samples increases, the training MSE gradually increases from approximately 0.00013 to 0.00019, while the cross-validation MSE remains relatively stable, ranging from approximately 0.00025 to 0.00031. The gap between the training and cross-validation errors generally decreases as the training set grows, suggesting improved generalization with increasing sample size. Although some variability remains across the cross-validation folds, particularly at smaller training sizes, the overall learning curves become relatively stable at larger training sizes. These results suggest limited evidence of overfitting under the constant-force and varying-mass condition.

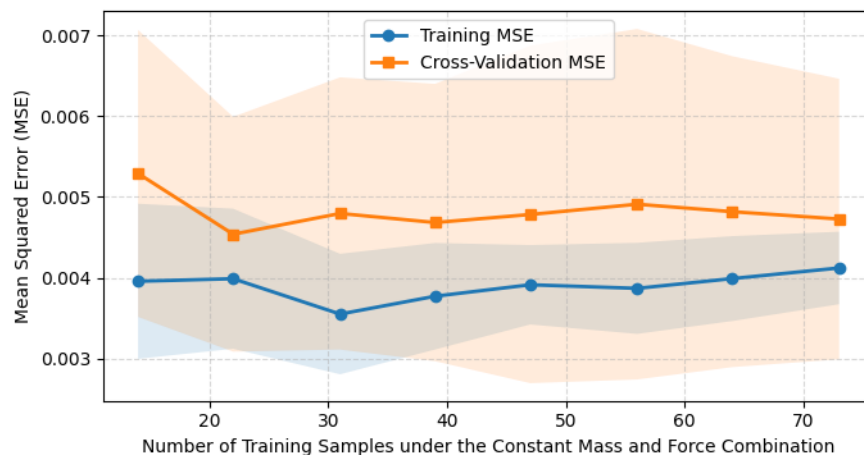


Figure 8. MLR curve under constant mass and force combination.

Figure 8. shows the learning curve of the MLR model using the combined dataset from the constant-mass and constant-force conditions. Across the training sizes, the training MSE remains relatively stable, ranging from approximately 0.0036 to 0.0041, while the cross-validation MSE ranges from approximately 0.0045 to 0.0053. Although the gap between the two curves generally remains small and relatively stable, the cross-validation error is consistently higher than the training error. In addition, the relatively wide variability bands, particularly for the cross-validation MSE, indicate variability in model performance across the folds. Compared with the models trained separately under the constant-mass and constant-force conditions, the combined model exhibits higher error values, suggesting that combining data from the two experimental conditions increases the complexity and variability of the relationship captured by the MLR model.

The learning curves for the constant-mass and constant-force conditions show decreasing gaps between the training and cross-validation errors as the training size increases, suggesting that the MLR model provides relatively stable generalization under each individual experimental condition. In contrast, the model trained on

the combined dataset exhibits consistently higher training and cross-validation errors, together with greater variability across the cross-validation folds. These results suggest that combining the two experimental conditions introduces additional variability and a more complex relationship that may not be fully captured by a single linear regression model.

3.3 Comparison with the idealized theoretical baseline ($a = F/m$)

To assess whether the trained models capture experimental effects not represented in the idealized theoretical relationship, we compared the measured acceleration ($a = 2s/t^2$, obtained directly from the timed laboratory trials) with the theoretical acceleration calculated directly from the idealized relationship $a = F/m$ using the same force and mass values, with no model fitting involved. This comparison was carried out on the 25 laboratory trials underlying Dataset 1 (force varied, mass constant), the 25 trials underlying Dataset 2 (mass varied, force constant), and the pooled set of 50 trials corresponding to Dataset 3. The results are summarized in Table 8.

Table 8. Comparison of the idealized theoretical baseline ($a = F/m$) with measured acceleration and trained MLR model performance.

Experimental condition	n	MAE of theoretical baseline (m/s ²)	MAPE of theoretical baseline (%)	R ² , theoretical baseline	R ² , MLR model	Maximum theoretical error (m/s ² , %)
Constant mass, varying force	25	0.237	25.1	−0.045	0.99	0.369 (29.6%)
Constant force, varying mass	25	0.404	28.8	−1.247	0.99	0.678 (37.3%)
Pooled laboratory data	50	0.320	26.9	−0.041	0.97	0.678 (37.3%)

As shown in Table 8, direct application of the idealized relationship ($a = F/m$) resulted in MAEs of 0.237 and 0.404 m/s² for the constant-mass and constant-force conditions, respectively, with corresponding MAPEs of 25.1% and 28.8%. For the pooled laboratory data, the MAE and MAPE were 0.320 m/s² and 26.9%, respectively. The maximum absolute deviations from the theoretical baseline were 0.369 m/s² (29.6%) for the constant-mass condition and 0.678 m/s² (37.3%) for the constant-force condition. The pooled dataset had the same maximum deviation of 0.678 m/s² (37.3%).

The theoretical baseline yielded negative R² values (−0.045, −1.247, and −0.041), indicating that, for these experimental data, predictions obtained directly from the idealized relationship performed worse than using the mean measured acceleration as a reference predictor. This result indicates that direct application of the idealized relationship did not adequately reproduce the measured acceleration under the present experimental conditions. In contrast, the trained MLR models achieved substantially higher R² values of 0.99, 0.99, and 0.97 for the corresponding datasets (Table 6).

The discrepancy between the idealized theoretical predictions and the measured acceleration is likely attributable to non-ideal characteristics of the experimental apparatus, including friction, pulley rotational inertia, and other mechanical losses that are not explicitly represented in the simplified ($a = F/m$) calculation. The results therefore suggest that the trained MLR models captured empirical relationships within the range of experimental measurements, including systematic effects associated with the actual apparatus conditions. The comparison should not be interpreted as indicating that the MLR models supersede Newton's Second Law; rather, it demonstrates the ability of the data-driven models to approximate the observed behavior of a non-ideal experimental system.

4. Conclusions

The results demonstrate three significant findings. First, the acceleration values obtained from Tracker video analysis followed trends like those observed in the laboratory measurements, supporting its use as an additional source of motion data.

Second, the MLR models achieved high predictive performance, particularly for Datasets 1 and 2, where the experimental conditions were individually controlled. The slightly lower performance of the model trained on

Dataset 3 suggests that combining different experimental conditions increases data variability and makes the prediction task more challenging.

Third, the comparison with the idealized theoretical baseline demonstrates the practical value of the data-driven approach. Although Newton's Second Law provides the fundamental physical relationship among force, mass, and acceleration, the simplified expression ($a = F/m$) does not account for several non-ideal characteristics of the actual laboratory apparatus. The MLR models were able to learn these systematic experimental effects from the measured data, resulting in predictions that more closely matched the observed acceleration values.

These results suggest that integrating laboratory experiments, Tracker-based motion analysis, and MLR modeling provides a useful framework for predicting acceleration under specific experimental conditions. Overall, all developed models demonstrated relatively stable predictive performance within their respective experimental conditions. Therefore, these models may serve as supplementary tools for predicting physics outcomes under similar conditions, particularly when real experiments are constrained by time, budget, or access to laboratory equipment.

Acknowledgements

Thank you to the Department of Physics and General Science, Faculty of Science and Technology, for providing the experimental kit for Newton's Second Law of Motion and its personnel.

Conflict of Interest

The authors declare that there are no conflicts of interest related to the work reported in this paper.

ORCID

Dusadee Khamboonruang
<https://orcid.org/0009-0001-2694-5876>
Thawachai Promrat
<https://orcid.org/0000-0003-2456-969>

Ethical Approval

This study involved only theoretical physics and artificial intelligence and did not involve human participants, human biological materials, medical records, or personal data. Therefore, ethical approval for human-subject research was not required.

References

- Boonkong, A., & Hormdee, D. (2018). Implementing a dual-touch 4-wire analog resistive touchscreen via regression analysis. *Engineering and Applied Science Research*, 45(1), 39–46.
- Brown, D., & Cox, A. J. (2009). Innovative uses of video analysis. *The Physics Teacher*, 47(3), 145–150. <https://doi.org/10.1119/1.3081296>
- Claessens, T. (2017). Analyzing virtual physics simulations with Tracker. *The Physics Teacher*, 55(9), 558–560. <https://doi.org/10.1119/1.5011834>
- Farizal, F., Dachyar, M., Taurina, Z., & Qaradhawi, Y. (2021). Disclosing fast moving consumer goods demand forecasting predictor using multi linear regression. *Engineering and Applied Science Research*, 48(5), 627–636.
- Kasemsan, W., & Makuwan, A. (2023). Soil moisture prediction via a multiple linear regression model for stainless steel tube sensor. *Journal of Vocational Education in Agriculture*, 7(1), 101–117.
- Khamboonruang, D. (2021). Determination of gravitational acceleration using video analysis techniques with Tracker program for the Atwood mechanical experiment. In *Proceedings of the 4th National Academic Conference "Rajabhat Krung Kao" (ARUCON 2021)* (pp. 387–394). Ayutthaya Rajabhat University.
- Liu, Y., Lei, Q., He, X., Xue, Y., He, K., Yang, H., Wang, Y., Zhang, X., Yang, L., Zhou, Y., Hu, R., & Xie, Y. (2025). Building an affordable self-driving lab: Practical machine learning experiments for physics education using Internet-of-Things. *APL Machine Learning*, 3(4), 046105. <https://doi.org/10.1063/5.0283529>

- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
- Poonyawatpornkul, J., & Luksameevanish, V. (2019). Development of scientific understanding using high speed video analysis technique: Case study on projectile motion. *Journal of Education, Prince of Songkla University, Pattani Campus*, 30(1), 71–84.
- Prapaporn, W. (2019). An analysis on material quantity relationship of reinforced concrete buildings between multiple linear regression and artificial neural network. *Naresuan University Engineering Journal*, 14(1), 84–102.
- Qu, K. (2024). Research on linear regression algorithm. *MATEC Web of Conferences*, 395, 01046. <https://doi.org/10.1051/mateconf/202439501046>
- Renika, J., Prima, E. C., & Amprasto, A. (2024). Kinematics analysis on accelerated motion using Tracker video analysis for educational purposes. *Momentum: Physics Education Journal*, 8(1), 23–31. <https://doi.org/10.21067/mpej.v8i1.8883>
- Settakhumpoo, J., & Benjaoran, V. (2022). The comparative accuracy evaluation of water evaporation using multiple linear regression and simplified Penman's equation. *Engineering and Technology Horizons*, 39(4), 138–148.
- Tabyang, W., Wonglakorn, N., Sakdinakonr, R., & Suksiripattanapong, C. (2022). Prediction of unconfined compressive strength of soft clay improved with cement and fly ash geopolymer using multiple-linear regression method. *Journal of Engineering and Innovation*, 15(4), 43–49.
- Wattanakasiwich, P., & Poonyawatpornkul, J. (2012). High-speed video analysis in mechanics. *Srinakharinwirot Science Journal*, 28(2), 212–232.
- Wee, L. K., Chew, C., Goh, G. H., Tan, S., & Lee, T. L. (2012). Using Tracker as a pedagogical tool for understanding projectile motion. *Physics Education*, 47(4), 448–455. <https://doi.org/10.1088/0031-9120/47/4/448>
- Wee, L. K., & Leong, T. K. (2015). Video analysis and modeling performance task to promote becoming like scientists in classrooms. *American Journal of Educational Research*, 3(2), 197–207. <https://doi.org/10.12691/education-3-2-13>