J-IAMSTEM
Journal of Innovation, Advancement, and Methodology in STEM Education
Vol. 2, No. 2, pp. 80-97, March-April 2025
https://so13.tci-thaijo.org/index.php/j_iamstem

Conceptual Understanding in Fundamental and Mechanical Physics Among Pre-Service Physics Teachers of Surindra Rajabhat University: a Statistic and Machine Learning Analysis

Choojit Sarapak[1], Aktanat Luengsiriwan[1], Prayut Kong-In[1], Amnuay Wattanakornsiri[1], Jutamas Yoomark[1], Nattaphorn Malingam[1], Birabongse Hardthakwong[2], and Thodsaphon Lunnoo[1]
[1]Faculty of Science and Technology, Surindra Rajabhat University, Surin, Thailand
[2]Faculty of Public Health, Kasetsart University, Chalermphakiet Campus, Sakon Nakhon Province, Thailand
*Correspoding author's E-mail address: thodsaphon.l@srru.ac.th

Received: 2 Jan 2025 Revised: 2 Apr 2025 Accepted: 4 Apr 2025

Abstract

A study of conceptual understanding in fundamental and mechanical physics of pre-service physics teachers (PPTs) is a fundamental aspect that can improve the points of misunderstanding among students, including critical thinking (Cr) and computational thinking (Co) skills. In this work, we analyzed 110 physics exam items, encompassing fundamental concepts, significant principles, units of measurement, vectors and their properties, one-dimensional motion, Newton's laws of motion, projectile motion, circular motion, and simple harmonic motion. To evaluate conceptual knowledge and identify correlations with the total physics score, we employed a machine learning (ML) approach involving five algorithms: random forest, logistic regression, support vector machine, naive Bayes, and K-nearest neighbor. The following conclusions could be drawn regarding the conceptual understanding of PPTs. Overall, they still develop knowledge of fundamental concepts, encompassing critical and computational thinking skills. The average combined score for all study subjects was 39.8%, indicating that their understanding was insufficient to be considered competitive for selection as specialized physics teachers in the civil service examination. We highlighted that GPA played a significant role in mathematics in predicting the total score of the physics exam through the implementation of the random forest algorithm in ML. This algorithm achieved an accuracy of approximately 67.0%, surpassing the performance of other algorithms. Thus, the overall GPA, school size, and high school gender also influenced the overall scores. Our study highlighted the interconnectedness between the subject materials of fundamental and mechanical physics, revealing misconceptions that need to be addressed to enhance the performance of PPTs in competitive examinations, and provided the significance of calculation skills in influencing physics exam scores.

Keywords: Conceptual understanding, Fundamentals and Mechanical Physics Exam, Machine Learning.

1. Introduction

Understanding becomes the primary foundation for self-enhancement through diverse approaches, aiming to foster innovative thinking and improve skills (Banda & Nzabahimana, 2021; Kaltakci-Gurel et al., 2016). Comprehension refers to a practical understanding or the rational capacity to utilize acquired knowledge in situations or circumstances beyond the original context of its acquisition (Banda & Nzabahimana, 2021; Kaltakci-Gurel et al., 2016). According to this definition, understanding concepts was described as the capacity of a learner to reason in situations or circumstances that call for the careful application of ideas, descriptions, connections, or representations. These conceptual understanding definitions imply that learners should be able to amalgamate information and knowledge from diverse schemas, applying them in novel contexts to showcase conceptual understanding (Sarapak et al., 2025). These conceptual understanding definitions imply that learners should be able to amalgamate information and expertise from diverse schemas, applying them in novel contexts to showcase conceptual understanding. Moreover, learners' knowledge of physics' conceptual aspects becomes evident through their innovative and skilful amalgamation of unrelated or diverse fragments of information. This allows them to solve problems by employing familiar concepts in fresh and innovative situations (Mills, 2016).

In the context of 21st-century education, each student must cultivate self-awareness (Sagala et al., 2019). Consequently, cultivating a conceptual understanding of physics is paramount in learning. This is because it is one of the most challenging obstacles students and educators encounter (Mills, 2016). For physics learning, students require the capability to understand concepts more thinking and conceptual knowledge than it does memorization (Capriconia & Fufit, 2022). Understanding concepts is of immense importance in science education, particularly in physics. This allows students to concentrate on understanding the content rather than memorizing formulas (Capriconia & Fufit, 2022). Students must understand ideas clearly and have a positive attitude toward learning physics to succeed. This is because it is expected that students will be able to deliver the most significant outcomes in the learning process.

In research related to understanding, a wide range of studies are being conducted. For example, in 2019, Yana and a co-worker (Yana et al., 2019) investigated the knowledge of the concept of physics in wave material (three materials such as sound waves, waves on water, and mechanical waves) using multiple choice questions. The study involved 29 physics major students who were presented with 15 categorized conceptual questions. The findings demonstrated that assessing concept comprehension solely through multiple-choice questions needed to be improved in measuring the depth of clear conceptual understanding. In 2018, Zulfa and co-workers (Zulfa et al., 2019) researched students' comprehension of work and energy concepts within the context of online hybrid learning. The study revealed that students initially held a comprehensive understanding of work and energy as a collective concept, falling within the category of sufficiency. Furthermore, numerous research studies explore the knowledge of physics concepts. Comprehending physics content is crucial, and assessment strategies significantly impact students' understanding of learning materials.

ML techniques have garnered increasing attention in educational research due to their ability to process large datasets, uncover hidden trends, and support personalized learning environments (Abdrakhmanov et al., 2024) . For example, ML has been utilized to predict students' academic performance, identify those at risk of underachieving, and provide early interventions (Liu & Ardakani, 2022; Su et al., 2022). Moreover, ML algorithms can help understand learning behaviors and cognitive patterns by analyzing student responses, enabling educators to tailor their instruction accordingly (Abdrakhmanov et al., 2024; Nimy et al., 2023; Sathe & Adamuthe, 2021; Zhai et al., 2020). Specific models, such as Random Forests, which aggregate multiple decision trees to achieve more accurate classifications, have been employed in physics education to diagnose misconceptions and predict student success. Logistic Regression and Support Vector Machines have proven effective in modeling student data, particularly when the relationships between variables are complex and nonlinear. One key advantage of ML is its ability to reveal hidden patterns in complex educational datasets that might be missed with conventional statistical methods. For instance, ML can identify nuanced relationships between students' academic backgrounds (such as GPA in mathematics or physics and school size) and their conceptual understanding of specific subject areas. In science education, various ML models, including Random Forests, Logistic Regression, and Support Vector Machines, have been successfully utilized to classify student learning levels, evaluate conceptual comprehension, and optimize assessment strategies (Abdrakhmanov et al., 2024; Alamri et al., 2020; Mameno et al., 2021; Zhai et al., 2020) These models can handle high-dimensional data and learn the underlying structures adaptively without explicit programming.

Considering all aforementioned aspects, we could summarize that (i) conceptual understanding and (ii) assessment strategies significantly impact students' understanding. Thus, in this research, a study was designed to test pre-service physics teachers currently studying at the Surindra Rajabhat University (SRRU) using the examination questions (110 questions) to assess the readiness or understanding of fundamental physics content. This aimed to determine whether they were adequately prepared to teach high school physics. The examination would primarily focus on assessing comprehension of the topic of fundamental and mechanical physics, including basic concepts, significant principles, units of measurement, vectors and their properties, one-dimensional motion, Newton's laws of motion (force), projectiles and their motion, circular motion, and Simple Harmonic Motion (SHM). The aforementioned points would enable us to determine the understanding level of SRRU pre-service physics teachers in specific content units. Additionally, it highlights the need for further comprehension development in units where understanding was lacking. Due to the diverse educational backgrounds of most pre-service physics teachers from various schools, we designed a study to establish the correlation between overall GPA, physics course grade, and mathematics course GPA at the high school level. The aim was to determine how these factors influenced the total score of the physics exam. We used statistical principles and ML techniques to analyze and identify relationships. An ML approach with five algorithms was investigated: random forest (RF), logistic regression (LR), support vector machine (SVM), naive Bayes (NB), and K-nearest neighbor (KNN). Both statistical analysis and ML techniques were used to gain a deeper understanding of the relationships in this research. The study aims to answer the following key questions: (i) What is the level of conceptual understanding in fundamental and mechanical physics among pre-service physics teachers at SRRU? (ii) Which academic background factors significantly influence these students' physics test scores? and (iii) Which machine learning model provides the most accurate prediction of students' physics test performance based on their academic profiles? These results may offer profound information about the development of pre-service physics teachers of SRRU in each learning unit of physics.

2. Method

2.1 Participants

The participants of this study were pre-service physics teachers at SRRU. The majority of the students consisted of 21 female students (55.26%) and 17 male students (44.74%). Among them, the largest group was studying in their first year, consisting of 18 students (47.37%). The second largest group comprised 13 students in their third year (34.21%), while the smallest group consisted of 7 students in their second year (18.42%), as illustrated in Table 1.

Table 1 Fundamental information about the pre-service physics teachers who participated in the examination (n = 38)

Fundamental information	Total	Percentage (%)
Gender		
Male	17	44.7
Female	21	55.3
Grade levels		
First year	18	47.4
Second year	7	18.4
Third year	13	34.2

2.2 Conceptual understating of fundamental and mechanical physics

The variable was evaluated through a multiple-choice exam designed to assess students' conceptual understanding of Newton's Laws of Motion within the fundamental and mechanical physics framework. The variable was measured using an exam intended for the civil service exam in physics teachers in Thailand. The exam assessed candidates' knowledge and understanding of physics concepts relevant to civil service positions. Scores on the examined physics test strongly correlated with students' Fundamental Concept Inventory (FCI) scores. The FCI is a widely used test that measures students' understanding of fundamental physics concepts, including significant numbers, units, one-dimensional kinematics, Newton's laws, two-dimensional motion with constant acceleration, projectile motion, impulsive forces, vector sums, cancellation of troops, and circular motion. In this study, we chose the civil service exam of physics teachers to measure students' conceptual understanding of physics. This exam was preferred as it offers a comprehensive assessment of their knowledge of fundamental and mechanical physics, the chosen unit of study to evaluate pre-service physics teachers. The test exam comprised 110 questions, each with multiple choices (a), (b), (c), and (d) answers. The primary objective of these questions was to assess whether pre-service physics teachers could apply fundamental and mechanical physics for computational and critical thinking.

2.3 Dataset description and pre-processing

In this study, we used a dataset from SRRU, a pre-service physics teacher. We considered demographic and academic background features. As summarized in Table 2, the feature selection and description were presented.

Data pre-processing phase: During the data pre-processing step, the raw dataset was transformed to ensure appropriate forms. This involved eliminating irrelevant features, converting data types as needed, and handling any missing instances. Pre-processing was crucial before using the data in the analysis phase. The analysis step involved several stages, including feature selection, data cleaning to handle errors and outliers, and data normalization to scale the features appropriately for analysis. These processes collectively contributed to preparing the data for meaningful analysis and accurate modeling.

Data cleaning: A dataset might contain missing values due to human typing errors, transmission issues, or duplicate data entries. Addressing these issues was essential to achieving accurate results. This could be done by eliminating duplicate rows and appropriately handling missing values by filling them with the mean or removing them from the dataset. However, in our specific situation, the dataset had no missing tuples, meaning there were no missing values to address.

Table 2 Description and feature selection of data for using ML.

No.	Variable	Type	Description
1	Gender	Binary	Gender (Men = 1, Women = 0)
2	Academic year level	Binary	First year = 0 and other = 1
3	GPA3.5, GPA3, and GPA2.5	Binary	The overall GPA of pre-service physics teachers at the high school level before their undergraduate studies. (GPA > 3.5, 3, and 2.5 = 1 and GPA < 3.5 = 0)
4	MGPA3.5, MGPA3, and MGPA2.5	Binary	The average GPA in mathematics of pre-service physics teachers at the high school level before their undergraduate studies. (GPA > 3.5, 3, and 2.5 = 1 and GPA < 3.5 = 0)
5	PGPA3.5, PGPA3, and PGPA2.5	Binary	The average GPA in physics attained by pre-service physics teachers at the high school level before their undergraduate studies. (GPA > 3.5, 3, and 2.5 = 1 and GPA < 3.5 = 0)
6	SofSCh	Binary	The size of the school where students completed their high school education (either large =1 and small = 0)

Data normalization: Data normalization is considered a crucial preparatory step before training a dataset for any classification or prediction system. Normalization scales the data between smaller values, typically within the range of [0, 1] or [-1, 1]. This process improves the results, mainly when the data is equally distributed, but it is also beneficial for randomly distributed data. Some ML algorithms perform better when the values are within the range [0, 1]. In this study, we employed the min-max normalization approach, as shown in Eq. 1. This method ensures that all the features in the dataset are scaled to lie between 0 and 1. For example, the "GPA" feature contains numerical values but does not provide the range. Through min-max normalization, these values are transformed to fit within the range of 0 to 1, making them suitable for the subsequent ML analysis.

$$Z_i = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (1)$$

where zi is the ith normalized value in the dataset, xi, min(x), and max(x) are the ith value in the dataset, the minimum value in the dataset, and the maximum value in the dataset, respectively.

2.4 ML Models

In this study, five widely used supervised machine learning (ML) algorithms were employed to evaluate and predict the performance of pre-service physics teachers based on various academic and demographic factors. These algorithms represent different learning paradigms and offer unique strengths in educational data classification tasks.

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks. Aggregating the results from multiple trees reduces the risk of overfitting and enhances model generalization. One of its key advantages is the ability to rank the importance of input variables, which is particularly useful for interpreting the influence of educational factors on student performance.

Logistic Regression is a simple yet effective linear classifier that predicts the probability of a binary or categorical outcome. It models the relationship between one or more independent variables and a binary dependent variable using the logistic function. Despite its simplicity, Logistic Regression is powerful in educational research due to its interpretability and efficiency in handling linearly separable data.

Support Vector Machine is a supervised learning algorithm that constructs a hyperplane or set of hyperplanes in a high-dimensional space to classify data points. It is especially effective when the data are not linearly separable, utilizing kernel functions (such as radial basis function or polynomial kernel) to map input features into higher-dimensional spaces. In this study, SVM was used to explore potential nonlinear boundaries between passing and non-passing students based on their academic profiles.

Naïve Bayes is a probabilistic classifier grounded in Bayes' theorem, which assumes strong independence between features. It calculates the probability of each class given a set of features and selects the class with the highest posterior probability. This algorithm is computationally efficient and often performs well with high-dimensional datasets, despite its strong assumptions.

K-Nearest Neighbor is a non-parametric, instance-based learning algorithm that classifies a new observation based on the majority vote of its k nearest neighbors in the feature space. KNN does not assume any underlying data distribution and is highly intuitive. However, it can be sensitive to the choice of 'k' and the scale of the data, necessitating feature normalization before implementation.

All ML models were implemented using Python's Scikit-learn library, which provides reliable tools for preprocessing, model training, validation, and evaluation. The performance of each model was assessed using standard classification metrics, including accuracy (the overall correctness of predictions), precision (true positives over predicted positives), recall (true positives over actual positives), and F1-score (the harmonic means of precision and recall). These metrics provided a comprehensive view of each algorithm's ability to classify students into passing and non-passing categories based on their high school academic data.

3. Results and Discussion

This section presented the study's results as follows: (i) measurement of conceptual understanding of fundamental and mechanical physics, (ii) identification of factors influencing the physics test scores, and (iii) ML analysis of the physics test scores of pre-service physics teachers from SRRU.

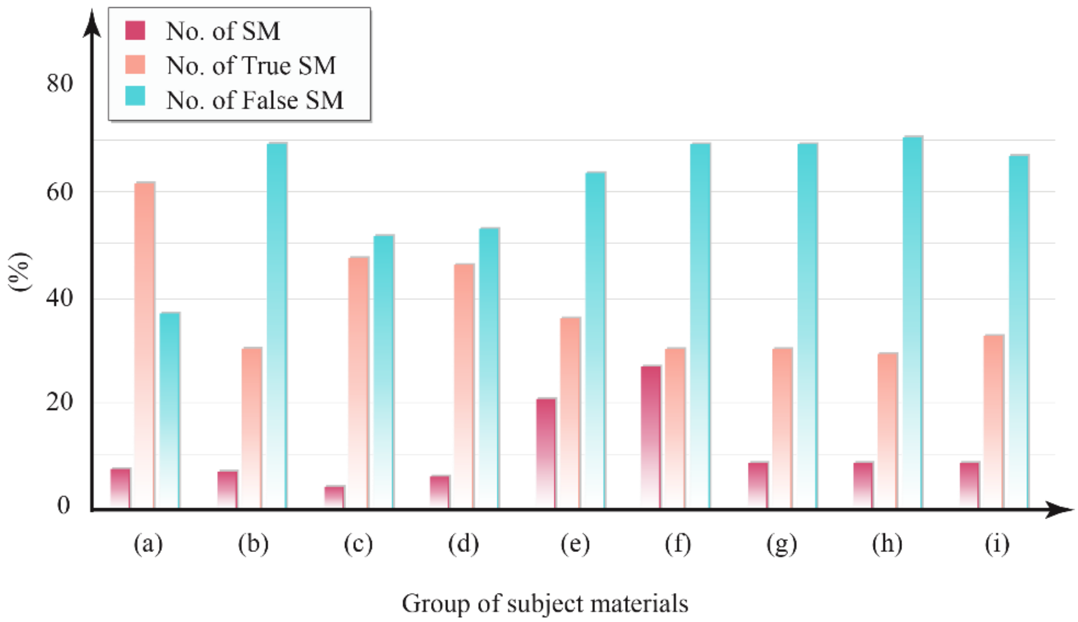

Figure 1 Summarization of subject materials (SM), including (a) fundamental concepts, (b) significant principles, (c) units of measurement, (d) vectors and their properties, (e) one-dimensional motion, (f) Newton's laws of motion (force), (g) projectiles and their motion, (h) circular motion, and (i) SHM. The topics were colored in Indian red, while the percentage of correct answers (true) for each item test was highlighted in salmon, and incorrect answers (false) were highlighted in medium turquoise.

3.1 Measurement of conceptual understanding of fundamental and mechanical physics

The outcomes of the analysis of students' answers on the first knowledge checked are illustrated in Figure 2. The study revealed the inclination of students' answer choices for each topic. Regarding the test, it was observed that the average of the highest percentage of subject materials was categorized as Newton's laws of motion (force), accounting for approximately 27%, closely trailed by one-dimensional motion at around 21%. On the other hand, the remaining items exhibited an average percentage falling within the range of 4.5% to 9.1%. This implied that the proportion between physics exams, Newton's laws of motion (force), and one-dimensional motion could potentially influence the movement of students' scores to some extent. Furthermore, when considering the overall student preferences for physics exams, it became evident that their performance was below average across all exams except for the fundamental concepts. This study's findings indicated that most pre-service physics teachers understood the basic concepts, calculating the correctness of 62.3% (True answer), indicating accuracy, and 37.7% (False answer), indicating inaccuracy. In contrast, they displayed a deficiency in comprehension of other topics. As an illustration, significant principles were also observed, and only a tiny percentage of physics exams were completed accurately. This led to a relatively high error percentage of 69.4%, making it the predominant source of mistakes. Additionally, there were other subject materials in which pre-service physics teachers made mistakes, with analogous percentage values of one-dimensional motion, Newton's laws of motion (force), projectiles and their motion, circular motion, and simple harmonic motion, contributing to approximately 70% of the errors.

Table 3 displays the percentage distribution of responses for each item in the initial knowledge test. Regarding question number 23, 23.7% of students provided the least accurate answers for fundamental content. It was worth noting that this question was associated with critical thinking skills. The following statement on problem number 23

> "Consider the following text: (1) Object A with a mass of 6 kilograms and Object *B with a mass of 8 kilograms have a combined mass on Earth of 14 kilograms. However, their combined mass in a weightless condition is zero kilograms. (2) Mr. Dang walks eastward for 200 meters, then westward for 150 meters. Mr. Dang is now 50 meters away from the starting point. This 50-meter distance represents the magnitude of a scalar quantity.*
>
> *(a) 1 and 2 True*
> *(b) Only 1 True*
> *(c) Only 2 True*
> *(d) 1 and 2 False" (Correct)*

Most students responded to question C in a manner that suggested a lack of comprehension of fundamental physics concepts. Sure, students regarded statements (1) and (2) as accurate despite being incorrect. This reveals a need for more understanding among students regarding concepts like mass in microgravity and the distinction between vector and scalar quantities.

The percentage of students providing the least accurate responses concerning significant material was 15.8% for question 7 regarding critical thinking skills. This item measured the understanding of significant principles. The subsequent statement regarding problem 7 was as follows:

> *Which option correctly represents the quantity of 250,000,000 meters in exponential form?*
>
> *(a) 2.5 x 10^8 m (Correct)*
> *(b) 2.50 x 10^8 m*
> *(c) 2.500 x 10^8 m*
> *(d) 2.500 x 10^8 m."*

The percentage of students providing the least accurate responses concerning the topic of unit material was 26.3% for question number 10, regarding computational thinking skills. The subsequent statement regarding problem 10 was as follows:

> *A car travels at a speed of 54 km/h. What is its speed in meters per second?*
>
> *(a) 10 m/s*
> *(b) 15 m/s (Correct)*
> *(c) 20 m/s*
> *(d) 25 m/s*

This indicates that the majority of students need more comprehension of unit conversions.

Table 3 Analyzed answers of pre-service physics teachers. (i) Various SM, (ii) (ii) item question tests, (iii) thinking process including Cr and Co, and (iv) the percentage summation of students answers with each item tests (% of summarized student answer of each item tests) were presented.

Subject materials	Item tests	Thinking process	% of summarized student answer of each item tests	
			True	False
Fundamental	1	Cr	84.2	15.8
	2	Cr	81.6	18.4
	3	Cr	86.8	13.2
	4	Cr	47.4	52.6
	18	Cr	52.6	47.4
	20	Cr	60.5	39.5
	21	Cr	86.8	13.2
	22	Cr	36.8	63.2
	23	Cr	23.7	76.3
Significant	7	Cr	15.8	84.2
	11	Cr	31.6	68.4
	12	Co	28.9	71.1
	13	Co	34.2	65.8
	14	Co	28.9	71.1
	15	Co	31.6	68.4
	16	Co	47.4	52.6
	17	Co	26.3	73.7
Unit	5	Cr	57.9	42.1
	6	Cr	57.9	42.1
	8	Co	52.6	47.4
	9	Co	44.7	55.3
	10	Co	26.3	73.7
Vector	24	Cr	15.8	84.2
	25	Cr	55.3	44.7
	26	Co	39.5	60.5
	27	Co	18.4	81.6
	28	Co	47.4	52.6
	29	Co	86.8	13.2
	30	Co	63.2	36.8
One-dimension motion	19	Co	23.7	76.3
	31	Cr	18.4	81.6
	32	Co	57.9	42.1
	33	Co	44.7	55.3
	34	Cr	86.8	13.2
	35	Cr	10.5	89.5
	36	Cr	44.7	55.3
	37	Co	89.5	10.5
	38	Cr	26.3	73.7
	39	Cr	10.5	89.5

Subject materials	Item tests	Thinking process	% of summarized student answer of each item tests	
			True	False
	40	Cr	44.7	55.3
	41	Co	5.3	94.7
	42	Co	13.2	86.8
	43	Co	65.8	34.2
	44	Co	39.5	60.5
	45	Co	36.8	63.2
	46	Co	31.6	68.4
	47	Co	13.2	86.8
	48	Co	21.1	78.9
	49	Co	15.8	84.2
	50	Co	55.3	44.7
Newton law of motion (force)	51	Cr	39.5	60.5
	52	Cr	26.3	73.7
	53	Cr	39.5	60.5
	54	Cr	10.5	89.5
	55	Cr	28.9	71.1
	56	Co	26.3	73.7
	57	Co	42.1	57.9
	58	Co	23.7	76.3
	59	Co	52.6	47.4
	60	Co	26.3	73.7
	61	Cr	36.8	63.2
	62	Cr	31.6	68.4
	63	Cr	57.9	42.1
	64	Co	15.8	84.2
	65	Co	34.2	65.8
	66	Co	28.9	71.1
	67	Co	31.6	68.4
	68	Cr	15.8	84.2
	69	Co	7.9	92.1
	70	Co	15.8	84.2
	71	Cr	28.9	71.1
	72	Cr	18.4	81.6
	73	Cr	5.3	94.7
	74	Cr	60.5	39.5
	75	Cr	71.1	28.9
	76	Cr	39.5	60.5
	77	Co	31.6	68.4
	78	Co	31.6	68.4
	79	Co	23.7	76.3
	80	Co	21.1	78.9
Projectile	81	Cr	36.8	63.2
	82	Cr	63.2	36.8
	83	Cr	13.2	86.8

Subject materials	Item tests	Thinking process	% of summarized student answer of each item tests	
			True	False
	84	Cr	10.5	89.5
	85	Co	15.8	84.2
	86	Co	31.6	68.4
	87	Co	44.7	55.3
	88	Co	28.9	71.1
	89	Co	28.9	71.1
	90	Co	34.2	65.8
Circular motion	91	Cr	13.2	86.8
	92	Cr	15.8	84.2
	93	Co	55.3	44.7
	94	Cr	18.4	81.6
	95	Cr	23.7	76.3
	96	Cr	34.2	65.8
	97	Co	52.6	47.4
	98	Co	26.3	73.7
	99	Co	31.6	68.4
	100	Co	23.7	76.3
SHM	101	Cr	21.1	78.9
	102	Cr	21.1	78.9
	103	Cr	34.2	65.8
	104	Co	36.8	63.2
	105	Cr	15.8	84.2
	106	Co	55.3	44.7
	107	Cr	52.6	47.4
	108	Co	44.7	55.3
	109	Co	15.8	84.2
	110	Co	31.6	68.4

For one-dimensional motion and Newton's laws of motion (force) materials, the percentage of the pre-service physics teachers' lowest correct answers were in questions number 41 and 73, which were 5.3% identical values. The question statements are presented in Figure 2. The correct statement is for options D and C, corresponding with one-dimensional and Newton's laws of motion (force) materials.

(a)

41. Mr. Daeng is running at a speed of 10 m/s to the south direction. One second later, her speed is measured at 4 m/s in the same direction. Find the acceleration of Mr. Daeng as he slows down.

(a) 6 m/s^2, South direction

(b) 6 $m/s^{2,}$ North direction

(c) 14 m/s^2, South direction

(d) 14 m/s^2, Nouth direction

(b)

73. The frictional force between two solids in contact. How is the magnitude of the frictional force related to the following Force?

(i) The frictional force varies with the applied force.
(ii) The frictional force varies with the applied traction force.
(iii) The frictional force varies with the reaction force which is perpendicular to the surface.

What is the accurate statement?

(a) Item *(i)* (b) Item *(ii)*

(c) Item *(iii)* (d) Items *(i), (ii),* and *(iii)*

Figure 2 The lowest percentage of examination tests. (a) Test number 41 presented CO of one-dimensional motion in the context of pre-service physics teachers. In contrast, (b), Test number 73 was the context of Newton's laws of motion (force) regarding the critical thinking process.

For projectile, circular, and SHM, the lowest percentage of correct answers were 10.5%, 15.8%, and 15.8%, respectively. These physics examinations were associated with questions 85, 92, and 109. The projectile and SHM material were CO skills, while the circular motion was related to critical thinking skills. For example, the statement of equation number 85 was followed:

Launching an object with an initial velocity of 25 meters per second at an angle of 37 degrees with the horizontal, how many seconds will it take for the object to fall along the same vertical line?

a) 1.5 s
b) 2.0 s
c) 3.0 s (Correct)
d) 4.0 s"

(item 92) Please consider the following text:

(i) Angular velocity equals the angle change rate per unit of time.
(ii) Circular motion with a constant speed is a type of motion where the speed remains constant, but the velocity is not continuous.
(iii) Circular motion with a constant speed is a type of motion where the speed remains steady, and no acceleration is involved.

The correct statement is: (i) and (ii) (Correct)

(a) (i) and (iii)
(b) (ii) and (iii)
(c) (i), (ii), and (iii)"

(item 109) A pendulum clock's pendulum has a length of 0.4 meters and swings at a rate of 0.6 revolutions per second. To change the oscillation frequency to 1.0 revolution per second, the pendulum's length would need to be how many meters?

(a) 0.144 m (Correct)
(b) 0.240 m
(c) 0.310 m
(d) 0.420 m"

The percentage of the pre-service physics teachers' highest correct answers on all subject materials was test number 37, which was 86.5%. The following statement was:

A car is traveling at a speed of 12 meters per second. The driver applies the brakes, causing the vehicle to slow down at 3 meters per second squared. How many seconds will it take for the car to come to a complete stop?
(a) 15 m
(b) 11 m
(c) 4 m (Correct)
(d) 2.5 m."

The examination item, test number 37, assessed the understanding of one-dimensional motion in the context of pre-service physics teachers' computational thinking abilities. The results indicated that pre-service physics teachers achieved the highest percentage of scores in this particular area.

The collective proficiency level of pre-service physics teachers in grasping subject matter concepts was 39.8%. This percentage suggests that students' comprehension of fundamental and mechanical physics concepts needs to be improved.

3.2 Factors influencing the physics test scores of pre-service physics teachers

The exam scores of all students are summarized in Table 4. Overall, the average exam score was 39.8 ± 8.2 points, ranging from the lowest score of 23.00 to the highest score of 75.00. When categorized by educational level, the findings are as follows: Students in the 2nd year had the highest average exam score of 44.0 ± 10.5 points, ranging from the lowest score of 32.00 to the highest score of 75.00. Following that, students at the 3rd year level achieved an average exam score of 41.1 ± 7.0 points, ranging from the lowest score of 30.00 to the highest score of 60.00 points. Lastly, 1st year students obtained an average exam score of 37.0 ± 7.3 points, ranging from the lowest score of 23.00 to the highest score of 54.00.

Table 4 The exam scores of pre-service physics teachers are divided by academic year level (n=38).

Academic year level	Examination Score (Total Score = 110)			
	Mean ± SD	Minimum	Maximum	95^{th} Percentile
First year (n=18)	37.0 ± 7.3	23.0	54.0	54.0
Second year (n=7)	44.0 ± 10.5	32.0	75.0	75.0
Third year (n=13)	41.1 ± 7.0	30.0	60.0	60.0
Total (n=38)	39.8 ± 8.2	23.0	75.0	60.0

The study of all factors affecting students' exam scores was further analyzed, as summarized in Table 4. The findings were as follows: (i) gender showed no significant correlation with exam scores at a confidence level of 95% (p-value = 0.05), (ii) the size of the school where students completed their high school education (either large or small) did not show a significant relationship with exam scores at a confidence level of 95% (p-value = 0.05), (iii) the average GPA performance throughout the high school level did not have a significant relationship with exam scores at a confidence level of 95% (p-value = 0.05), (iv) the average performance in mathematics GPA throughout the high school level, whether greater than or equal to 2.50, was not significantly related to students' exam scores at a confidence level of 95% (p-value = 0.05), and (v) the average performance in physics courses throughout the high school level, within the three groups of 3.50 or greater, 3.00, and 2.50, did not show a significant relationship with students' exam scores at a confidence level of 95% (p-value = 0.05). Based on the study results, the factors examined, including gender, school size, overall GPA performance, mathematics GPA average, and physics GPA average, did not exhibit a statistically significant correlation with students' exam scores at a confidence level of 95%.

However, a study of factors affecting the average performance in mathematics courses throughout the high school level, specifically for students with an average score of 3.50 or greater and 3.00, found a significant relationship with students' exam scores at a confidence level of 95% (p-value = 0.05), as presented in Table 5.

Table 5 The correlation between factors and pre-service physics teachers' exam scores (n = 38).

Factors		Number (percentage)		P-value
		passing criteria	No-passing criteria	
Gender				0.292
	Male	16(94.12)	1(5.88)	
	Female	20(95.24)	1(4.76)	
SofSCh				0.72
	Big and Bigger	6(60.00)	4(40.00)	
	Medium and small	15(71.43)	13(76.47)	
GPA3.5				0.20
	Less than	7(35.00)	13(65.00)	
	More than or equal	10(55.56)	8(44.44)	
MGPA3.5				0.04*+
	Less than	9(33.33)	18(66.67)	
	More than or equal	8(72.73)	3(27.27)	
PGPA3.5				0.06
	Less than	7(31.82)	15(68.18)	
	More than or equal	10(62.50)	6(37.50)	
GPA3				0.20
	Less than	2(25.00)	6(75.00)	
	More than or equal	15(50.00)	15(50.00)	
MGPA3				0.01*+
	Less than	4(22.220	14(77.78)	
	More than or equal	13(65.00)	7(35.00)	
PGPA3				0.27
	Less than	14(70.00)	3((30.00)	
	More than or equal	14(50.00)	14(50.00)	
GPA2.5				0.44
	Less than	1(100.00)	0(0.00)	
	More than or equal	16(43.24)	21(56.76)	
MGPA2.5				0.42
	Less than	2(28.75)	5(71.43)	
	More than or equal	15(48.39)	16(51.61)	
PGPA2.5				0.67
	Less than	2(33.33)	4(66.67)	
	More than or equal	15(46.88)	17(53.13)	

* Significant level of confidence <0. 05(p-value<0.05), +Fisher exact test

3.3 ML analysis of the physics test scores of pre-service physics teachers

The optimal random forest institutional model outperformed all other models, as illustrated in Table 6, achieving an accuracy of 67% and an f1-score of 0.71, corresponding to an ML evaluation metric. The f1-score combines the precision and recall scores of a model. It measures how accurately a model predicts outcomes across the entire dataset. Afterward, the Naive Bayes and k-nearest neighbor models exhibited an accuracy of 58%, with identical values. However, the accuracies of the two models differed significantly at 0.67 and 0.71, respectively. The model's discrimination showed lower performance than the Logistic regression and support vector machine models, achieving an accuracy of 50% with identical values. Therefore, the random forest model emerged as the optimal choice.

Table 6 Correlation accuracy of each algorithm including random forest, logistic regression, support vector machine, naive Bayes, and K-nearest neighbor.

Measures / Algorithm	Accuracy (%)	Precision	Recall	F1-score
Random forest	67.0	0.62	0.83	0.71
Logistic regression	50.0	0.50	0.83	0.62
Support vector machine	50.0	0.50	1.00	0.67
Naive bayes	58.0	0.56	0.83	0.67
K-nearest neighbor	58.0	0.55	1.00	0.71

As depicted in Figure 4, the institutional random forest model's mean decrease in accuracy indicates each variable's relative importance. Larger values signified factors that significantly impacted the goodness of fitting of the random forest models. If a specific variable was removed from the decision tree, the mean drop in accuracy represents the average decrease in classification accuracy across all decision trees that utilized the variable. This metric helps assess the importance of the variable in contributing to the overall accuracy of the decision trees and the model's performance. This analysis revealed that although the ideal model only necessitated six variables, several other factors were crucial in predicting pre-service physics teachers' physics test scores. In Figure 2, all six ideal variables demonstrated either high or relatively high relevance. The most significant drop in importance was also observed after MGPA (cutoff GPA 3 and 3.5). In order of importance, the following crucial parameters for the model fit were GPA3.5, MGPA2.5, SofSch, and Gen. Regarding PGPA, average GPA in physics exerted minimal influence on the random forest model, which variables performed very similarly in terms of their importance, in the model. Figure 4 suggests the need for further comprehensive analysis. Physics test scores for pre-service physics teachers, the average GPA in mathematics, and the overall GPA were identified as the most crucial factors in the model. However, it is essential to note that the score prediction was also significant, with both SofSch and Gen features playing critical roles in the model.

As discussed above, Newton's laws of motion (force) and one-dimensional motion topics would significantly impact the overall combined scores due to having the highest number of data points, accounting for 27.3% and 21.0%, respectively, when calculated as percentages. Compared to other subject materials, they had approximately ten times greater influence. Based on the study results, it's evident that most students still need to understand various topics related to comprehending physics content. Therefore, to enhance students' knowledge and experience through effective teaching, emphasis should be placed on significant principles, units of measurement, vectors and their properties, one-dimensional motion, Newton's laws of motion (force), projectiles and their motion, circular motion, as well as SHM in the context of fundamental and mechanical physics. Furthermore, the research study also explored the correlation between academic performance in high school, including overall grade point average and GPA performance in physics and mathematics subjects, and how teachers perform on their exams. The findings revealed that mathematical computation skills in high school had the most significant impact on exam scores. Following this, the GPA in physics subject, and the overall GPA had a decreasing influence. Additionally, the size of the school also affected the overall scores.

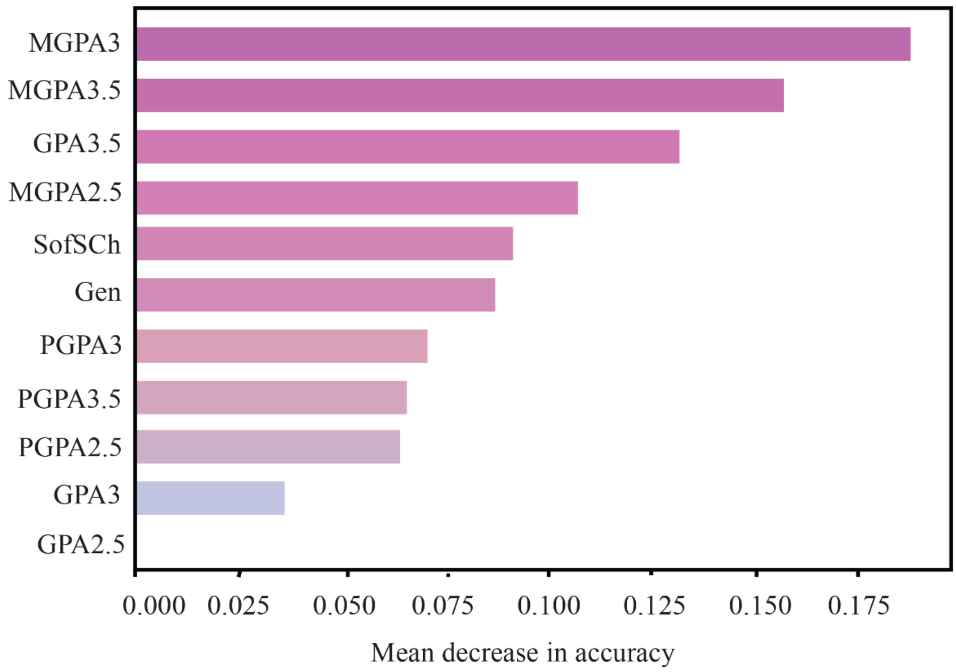

Figure 4 The correlation between various factors and the total score of pre-service physics teachers, as determined by the random forest model.

4. Conclusion

A study was conducted to examine the level of conceptual understanding in fundamental and mechanical physics among pre-service physics teachers. This study involved analyzing physics exams and utilizing a ML approach to assess conceptual knowledge and to identify correlations between variables such as high school grades and performance in physics exams. The conceptual understanding of pre-service physics teachers suggests that they lack a sense of fundamental concepts in both CR and CO skills. For example, significant principles, units of measurement, vectors and their properties, one-dimensional motion, Newton's laws of motion (force), projectiles and their motion, circular motion, and SHM were mistaken for conceptual understanding. The average combined score for all study subjects was 39.8%, less than 50% of the total score. This indicated that their knowledge still needed to be considered competitive for selection as specialized physics teachers in the civil service examination. Furthermore, it was discovered that the correlation between the average grade in high school mathematics and the overall exam scores had the most significant impact, based on the study conducted using ML techniques by random forest algorithm, having an accuracy of about 67.0%. Following this, the overall GPA, school size, and gender also affected the overall scores. Our study highlighted the interconnectedness between the subject materials of fundamental and mechanical physics, revealing misconceptions that need to be addressed to enhance the performance of pre-service physics teachers in competitive examinations, and provided the significance of calculation skills in influencing physics exam scores.

Acknowledgments

The Thailand Science Research and Innovation Fundamental Fund, fiscal year 2024 (Grant No. FRB. 68019), provided financial support for this research via Surindra Rajabhat University (SRRU). We thank the physics students at the faculty of science and technology, SRRU, for all support.

References

Abdrakhmanov, R., Zhaxanova, A., Karatayeva, M., Zholaushievna Niyazova, G., Berkimbayev, K., & Tuimebayev, A. (2024). Development of a framework for predicting students' academic performance in STEM education using machine learning methods. *International Journal of Advanced Computer Science and Applications*, *15*(1), 38–46. https://doi.org/10.14569/IJACSA.2024.0150105

Alamri, L. H., Almuslim, R. S., Alotibi, M. S., Alkadi, D. K., Ullah Khan, I. & Aslam, N. (2020). Predicting student academic performance using support vector machine and random forest. *n Proceedings of the 2020 3rd International Conference on Education Technology Management (ICETM 2020)*, 100–107. https://doi.org/10.1145/3446590.3446607

Banda, H. J. & Nzabahimana, J. (2021). Effect of integrating physics education technology simulations on students' conceptual understanding in physics: A review of literature. *Physical Review Physics Education Research*, *17*(2), 023108. https://doi.org/10.1103/PhysRevPhysEducRes.17.023108

Capriconia, J. & Mufit, F. (2022). Analysis of Concept Understanding and Students' Attitudes towards Learning Physics in Material of Straight Motion. *Jurnal Penelitian Pendidikan IPA*, *8*(3), 1453–1461. https://doi.org/10.29303/JPPIPA.V8I3.1381

Kaltakci-Gurel, D., Eryilmaz, A. & McDermott, L. C. (2016). Identifying pre-service physics teachers' misconceptions and conceptual difficulties about geometrical optics. *European Journal of Physics*, *37*(4), 045705. https://doi.org/10.1088/0143-0807/37/4/045705

Liu, X. & Ardakani, S. P. (2022). A machine learning enabled affective E-learning system model. *Education and Information Technologies*, *27*(7), 9913–9934. https://doi.org/10.1007/S10639-022-11010-X

Mameno, T., Wada, M., Nozaki, K., Takahashi, T., Tsujioka, Y., Akema, S., Hasegawa, D. & Ikebe, K. (2021). Predictive modeling for peri-implantitis by using machine learning techniques. *Scientific Reports*, *11*(1), 11090. https://doi.org/10.1038/s41598-021-90642-4

Mills, S. (2016). Teaching and Learning Medication Calculations: A Grounded Theory of Conceptual Understanding. *International Journal of Nursing Education Scholarship*, *13*(1), 1-9. https://doi.org/10.1515/IJNES-2015-0076

Nimy, E., Mosia, M. & Chibaya, C. (2023). Identifying At-Risk Students for Early Intervention—A Probabilistic Machine Learning Approach. *Applied Sciences*, *13*(6), 3869. https://doi.org/10.3390/APP13063869

Sagala, R., Umam, R., Thahir, A., Saregar, A. & Wardani, I. (2019). The effectiveness of stem-based on gender differences: The impact of physics concept understanding. *European Journal of Educational Research*, *8*(3), 753–761. https://doi.org/10.12973/EU-JER.8.3.753

Sarapak, C., Kong-In, P., Jindasri, P., Kakkaew, V., Wattanakornsiri, A., Yoomark, J., Sumrandee, C., Sreejivungsa, K., Malingam, N., Krongsuk, S., & Lunnoo, T. (2025). Instrument design and validation for enhancing instructional design using the TPACK framework: A study in Surin Province. *Journal of Innovation, Advancement, and Methodology in STEM Education (J-IAMSTEM), 2*(1), 10–24. https://so13.tci-thaijo.org/index.php/j_iamstem

Sathe, M. T. & Adamuthe, A. C. (2021). Comparative study of supervised algorithms for prediction of students' performance. *International Journal of Modern Education and Computer Science*, *13*(1), 1-21. https://doi.org/10.5815/IJMECS.2021.01.01

Su, Y. S., Lin, Y. D & Liu, T. Q. (2022). Applying machine learning technologies to explore students' learning features and performance prediction. *Frontiers in Neuroscience*, *16*, 1018005. https://doi.org/10.3389/FNINS.2022.1018005

Yana, A. U., Antasari, L. & Kurniawan, B. R. (2019). nalisis pemahaman konsep gelombang mekanik melalui aplikasi online Quizizz. *Jurnal Pendidikan Sains Indonesia*, *7*(2), 143–152. https://doi.org/10.24815/JPSI.V7I2.14284

Zhai, X., Yin, Y., Pellegrino, J. W., Haudek, K. C. & Shi, L. (2020). Applying machine learning in science assessment: a systematic review. *Studies in Science Education*, *56*(1), 111–151. https://doi.org/10.1080/03057267.2020.1735757

Zulfa, I., Kusairi, S., Latifah, E. & Jauhariyah, M. N. R. (2019). Analysis of student's conceptual understanding on the work and energy of online hybrid learning. *Journal of Physics: Conference Series*, *1171*(1), 012045. https://doi.org/10.1088/1742-6596/1171/1/012045