International Journal of Education Humanities and Social Science

ISSN: 2582-0745
Vol. 8, No. 06; 2025

DEVELOPING STEM-BASED TOPICS IN TEACHING PROBABILITY AND STATISTICS AT THAI NGUYEN UNIVERSITY OF MEDICINE AND PHARMACY

Do Thi Hong Nga
Thai Nguyen University of Medicine and Pharmacy

https://doi.org/10.54922/IJEHSS.2025.1147

Received:12 November 2025/Published: 23 December 2025

ABSTRACT
This paper presents the development and implementation of STEM-based topics in teaching Probability and Statistics (P&S) at Thai Nguyen University of Medicine and Pharmacy, aiming to enhance medical students' ability to apply knowledge, think critically, and conduct scientific research. Based on the theoretical framework of STEM education and the role of P&S in medical training, the authors propose a five-step process for designing STEM topics: identifying real-world problems, defining objectives, constructing guiding questions, planning lessons, and organizing activities. The paper illustrates this approach through the topic “Building a Logistic Regression Model using R to Predict Diabetes Risk,” which enables students to connect statistical theory with real medical data. Experimental results show that students actively participated, improving their collaboration, creativity, and problem-solving skills. The study recommends that universities, instructors, and educational authorities invest in facilities, provide STEM training, and scale up integrated teaching models to meet the requirements of modern medical education in the industry 4.0 era.

Keywords: STEM education, Probability and Statistics, Medical University, integrated teaching, competency development.

1. INTRODUCTION
In the era of the Fourth Industrial Revolution, medical workforce training is no longer limited to the acquisition of disciplinary knowledge but increasingly emphasizes the development of higher-order competencies such as critical thinking, problem-solving, creativity, and interdisciplinary collaboration (Bybee, 2013). Within this context, Probability and Statistics constitutes a fundamental course in medical education, enabling students to analyze medical data, design research studies, and evaluate treatment outcomes.

However, empirical evidence indicates that many medical students experience considerable difficulties in learning this subject due to its abstract nature and the predominance of theory-oriented teaching approaches that lack meaningful connections to clinical practice (Tran & Le, 2022; Nguyen & Tran, 2019).

The integration of STEM education into the teaching of Probability and Statistics has been identified as a promising pedagogical approach to address these challenges. This approach aims to foster comprehensive competency development, particularly within practice-oriented professional training environments such as medical education (Pham, 2020; Nguyen V.C., 2021).

ISSN: 2582-0745
Vol. 8, No. 06; 2025

2. THEORETICAL BACKGROUND

2.1. STEM Education in Medical Training

STEM education is an integrated instructional approach that combines Science, Technology, Engineering, and Mathematics to promote learners' critical thinking, problem-solving, and creative capacities (Bybee, 2013; National Research Council, 2011). According to the STEM competency framework proposed by the Vietnamese Ministry of Education and Training, STEM education emphasizes learner-centered activities, real-world problem contexts, project-based learning, and product-oriented outcomes to enhance learners' autonomy and practical competence (MOET, 2018).

In medical education, STEM integration can be implemented through the use of medical datasets, clinical case analyses, epidemiological study design, and treatment simulations. Such integration facilitates the connection between theoretical knowledge of Probability and Statistics and practical healthcare applications (Nguyen Duc Quang, 2022; Le Thi Phuong, 2023).

2.2. Characteristics of Probability and Statistics in Medical Curricula

Probability and Statistics plays a pivotal role in medical research, encompassing clinical trial design, drug efficacy analysis, diagnostic test evaluation, and epidemiological modeling. Nevertheless, many students perceive this course as abstract, unengaging, and disconnected from clinical realities (Tran & Le, 2022).

The GAISE College Report (ASA, 2016) recommends that statistics education should begin with real-world problems, employ authentic data, and prioritize the development of statistical thinking rather than procedural computation. Consequently, integrating STEM-based approaches into Probability and Statistics instruction aligns well with contemporary educational reforms.

3. DESIGN OF STEM-BASED TOPICS IN PROBABILITY AND STATISTICS INSTRUCTION

3.1. Design Process

Based on the five-step STEM topic design framework proposed by Nguyen Van Cuong (Nguyen V.C., 2021), including:

1. Identifying relevant real-world problems in medicine;
2. Linking these problems to mathematical content (Probability and Statistics);
3. Designing problem-oriented learning activities;
4. Organizing student-centered learning tasks;
5. Assessing outcomes through products or presentations;

this study proposes a five-step process for designing and implementing STEM topics in Probability and Statistics education.

Step 1. Identifying the STEM Topic

This step involves defining the topic title, real-world context, underlying idea, and STEM-related knowledge components (Table 1).

ISSN: 2582-0745
Vol. 8, No. 06; 2025

Table 1. STEM Knowledge Components in the Topic

Science (S)	Technology (T)	Engineering (E)	Mathematics (M)
Relevant medical knowledge	Statistical software and computational tools	Product development process	Probability and statistics concepts, data analysis methods

Step 2. Defining Learning Objectives
Objectives are specified in terms of knowledge, skills, and attitudes.
Step 3. Developing Guiding Questions
Guiding questions are designed to orient students' inquiry and problem-solving processes.
Step 4. Developing the Teaching Plan
This includes identifying target learners, instructional duration, learning environment, required preparation, and learning activities.
Step 5. Implementing Learning Activities
Learning activities are organized according to the designed plan.

3.2. Proposed STEM Topics

Topic 1: Interpreting Diagnostic Test Results
- **Problem:** How reliable is a positive diagnostic test result?
- **Content:** Conditional probability and Bayes' theorem.
- **Product:** Probability simulation based on sensitivity and specificity (GAISE, 2016).

Topic 2: Comparing Treatment Effectiveness
- **Problem:** Which treatment is more effective?
- **Content:** Hypothesis testing, p-values, chi-square tests.
- **Product:** Data analysis report with comparative visualizations (Nguyen & Tran, 2019).

Topic 3: Disease Risk Prediction
- **Problem:** Does smoking increase lung cancer risk?
- **Content:** Correlation analysis, logistic regression.
- **Product:** Epidemiological modeling and research proposal design (Le Thi Phuong, 2023).

Topic 4: Prognostic Modeling and Disease Assessment
- **Problem:** Developing a logistic regression model using R to predict diabetes.
- **Content:** Logistic regression, probabilistic estimation.
- **Product:** Predictive modeling and research outline.

3.3. Example: STEM Topic "Developing a Logistic Regression Model Using R to Predict Diabetes"

Step 1. Identifying the Real-World Problem
Diabetes is a rapidly increasing chronic disease with serious health consequences and a substantial burden on healthcare systems. Diagnostic capacity at primary healthcare levels remains limited

ISSN: 2582-0745
Vol. 8, No. 06; 2025

due to the lack of laboratory facilities. Applying statistical models through software such as R can support early disease prediction, cost reduction, and improved healthcare access in remote areas.

Table 2. STEM Knowledge Components of the Topic

Science (S)	Technology (T)	Engineering (E)	Mathematics (M)
Basic knowledge of diabetes	R software	Modeling workflow	Sampling theory, data description, correlation, regression

Step 2. Learning Objectives

- **Knowledge:** Diabetes and logistic regression models.
- **Skills:**
  - Applying statistical knowledge to predictive modeling;
  - Using R to analyze data and interpret results;
  - Planning, presenting, critiquing, and self-evaluating individual and group work.
- **Attitudes:**
  - Proactive and responsible learning;
  - Appreciation of statistical applications in medicine;
  - Collaborative and ethical professional behavior.

Step 3. Guiding Questions

- What is diabetes and its clinical manifestations?
- What is logistic regression and how is it applied?
- What is the modeling process?
- Which R packages are required?

Step 4. Teaching Plan

- **Participants:** Undergraduate medical students.
- **Duration:** Two in-class sessions (50 minutes each) and one week of self-study.
- **Learning Environment:** Computer laboratory.
- **Preparation:**
  - Instructors: STEM-based lesson plans, learning materials, R software.
  - Students: Review relevant Probability and Statistics content and R basics.

To guide students, lecturers may use a learning worksheet with the following guiding contents:

LEARNING WORKSHEET

Topic: *"Building a Logistic Regression Model Using R Software for Diabetes Prediction"*

Student's full name: ……………………………………

Student ID: ……………………………………………..

Group: …………………………………………………..

Group number: …………………………………………

Date: ……………………………………………………

1. Identifying the practical problem

What characteristics of diabetes make early prediction necessary?

………………………………………………………………………………………………

ISSN: 2582-0745
Vol. 8, No. 06; 2025

Why is it necessary to build a predictive model instead of relying solely on laboratory tests? .... **2. Statistical knowledge to be applied** List three statistical concepts that you think are essential for building a Logistic Regression model: .... **3. Plan for building the Logistic Regression model** Describe the steps your group will take to build the Logistic Regression model: .... **4. Practice with R software** Which R packages did you use to build the Logistic Regression model? .... Write basic R code to build a Logistic Regression model (with diabetes as the dependent variable): .... What is the statistical significance of the Glucose variable according to the model output? .... **5. Group discussion and peer review** Is the Logistic Regression model built by your group appropriate? Which criteria were used to evaluate it? .... Your comments on another group's model (please specify the group name): ....

Step 5. Implementation of Learning Activities

Activities include topic introduction, knowledge consolidation, planning, implementation, product presentation, peer review, and reflection.

Activity 1: Exploring and defining the topic requirements	
Objectives	Students analyze the requirements of the topic *"Building a Logistic Regression Model Using R for Diabetes Prediction."* Identify the necessary knowledge.
Content	Understanding the practical problem of diabetes. Identifying foundational knowledge: - Medicine: Clinical features of diabetes. - Probability and Statistics: Logistic regression, data types, binary models. - Technology: R software and required packages. - Engineering: The modeling process.
Organization	The lecturer introduces the topic and a sample model. Students discuss and identify requirements and necessary knowledge.
Expected outcomes	A summary table of required knowledge.
Activity 2: Identifying core knowledge	

Objectives	Reinforce and present key knowledge on: Logistic Regression models (concepts, applications, equations); Clinical signs of diabetes; Using R software and analytical packages (caret, pROC, DynNom, epiDisplay, etc.).
Content	New knowledge: Logistic Regression, R packages, medical background. Review knowledge: Variable classification, data coding, sensitivity/specificity, basic R commands.
Organization	Lecturer provides materials, research guidance, and assigns tasks. Students conduct research, group discussions, and present results.
Expected outcomes	Presentation of Logistic Regression models and explanation of applications. List and description of necessary R packages. Description of clinical manifestations of diabetes and the role of input variables.
Activity 3: Developing and presenting the plan	
Objectives	Help students recognize the application of theory to practice and scientific research. Develop a complete plan and modeling procedure using Logistic Regression.
Content	Developing and presenting a plan based on predefined criteria. Discussion, peer review, and plan refinement. Task allocation and timeline planning.
Organization	Lecturer sets requirements; students form groups, develop plans, and report. Lecturer provides support, feedback, and guidance when necessary.
Expected outcomes	A complete project plan. A five-step modeling procedure: - Step 1: Data familiarization – data reading, visualization, processing, and coding (UsingR, DescTools, ggplot2). - Step 2: Model selection – identifying relevant variables using the BMA package and bic.glm() function. - Step 3: Model training – splitting data into training/testing sets (caret package), training using train(). - Step 4: Model validation – calculating AUC, sensitivity, and specificity using the pROC package. - Step 5: Model visualization using nomograms – deploying models via DynNom and shiny or creating nomograms using the rms package.
Activity 4: Implementation and problem solving	
Objectives	Write and run R code according to the plan. Interpret the output results.
Content	Implement the plan using R software according to the steps defined in Activity 3: - Step 1: Data familiarization using read.csv(), summary(), ifelse(). - Step 2: Model selection using bic.glm() to identify the optimal model. - Step 3: Model training using caret to split data and train the model. - Step 4: Model validation using pROC and confusionMatrix() to evaluate performance.

ISSN: 2582-0745
Vol. 8, No. 06; 2025

	- Step 5: Model visualization using DynNom for web-based display and nomogram() for graphical representation.
Organization	Students work in groups using computers. Lecturer observes and provides technical and academic support as needed.
Expected outcomes	Correctly executed R code and corresponding outputs.
Activity 5: Product presentation	
Objectives	Present group products. Critically review and comment on other groups' products. Reflect and refine the final product.
Content	Group presentations of Logistic Regression models. Product evaluation based on predefined criteria. Discussion, feedback, self-assessment. Sharing challenges and lessons learned.
Organization	Lecturer assigns tasks, provides guidance, evaluates, and summarizes. Students present, discuss, and share experiences.
Expected outcomes	A complete Logistic Regression model. Group presentation slides. Online report (illustrative example: https://rpubs.com/DoNga/888781).

- **Evaluation of learning outcomes**

The following criteria are used to evaluate student performance, presented in table format:

Table 3. Criteria for Evaluating Student Learning Outcomes

Assessment Category	Score	Assessed Unit	Specific Criteria
Knowledge	15 points	Individual	- Presentation of basic knowledge (Logistic Regression, etc.)- Answering questions and participating in discussions
Skills	70 points	Group	- Correct problem and objective identification- Detailed planning- Application of statistical, computational, and technical knowledge- Report writing and presentation
Attitude	15 points	Individual	- Effective teamwork- Seriousness and active participation

Notes:

- *Knowledge and attitude* are assessed individually.
- *Skills* are assessed at the group level; the group score applies equally to all members.

ISSN: 2582-0745
Vol. 8, No. 06; 2025

4. PEDAGOGICAL EXPERIMENT

A pedagogical experiment was conducted with two cohorts (K53C: 69 students; K53E: 74 students) at Thai Nguyen University of Medicine and Pharmacy. Results indicated high levels of student engagement and collaboration. More than 70% of students actively contributed to discussions on model design and software utilization. Group-based task allocation and collective reflection were evident.

Preliminary findings suggest that STEM-based topics facilitate knowledge integration, enhance collaboration, creativity, critical thinking, communication skills, and significantly increase students' learning motivation.

5. CONCLUSIONS AND RECOMMENDATIONS

The implementation of STEM-based topics in Probability and Statistics instruction at Thai Nguyen University of Medicine and Pharmacy demonstrated clear effectiveness in improving learning outcomes and practical competence. Students developed a deeper understanding of statistical concepts while enhancing data analysis skills, critical thinking, problem-solving abilities, and teamwork.

- To effectively scale up this model, the following recommendations are proposed:
- Institutional investment in infrastructure, data analysis software, and STEM-focused professional development for instructors;
- Instructor-driven design of clinically relevant, project-based STEM topics;
- Active student engagement in STEM learning activities and technological skill development;
- Policy-level support for integrated STEM curriculum frameworks and dissemination of effective instructional models.

REFERENCES

1. Bybee, R. W. (2013). *The Case for STEM Education: Challenges and Opportunities*. NSTA Press.
2. Le, T. P. (2023). Designing integrated teaching topics in medical student training. *Journal of Medical Science and Technology*, *23*(3), 70–76.
3. Ministry of Education and Training. (2018). *STEM education framework in Vietnam*. MOET Publishing.
4. National Research Council (2011). *Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*. The National Academies Press.
5. Nguyen, D. Q. (2022). Integrating STEM education in university mathematics teaching. In *Proceedings of the National Conference on Innovation in Teaching Methods* (pp. xx–xx). Education Publishing House.
6. Nguyen, H. T. & Tran, D. H. (2019). "Enhancing statistical reasoning for medical students through contextual learning", *Vietnam Journal of Education*, Vol. 6, No. 2, pp. 35–40.
7. Nguyen, V. C. (2021). STEM-oriented teaching: An integrated approach. *Vietnam Journal of Education*, *(478)*, 14–18.
8. Pham, Q. T. (2020). STEM education: Trends and prospects in higher education reform. *Journal of Educational Sciences*, *(176)*, 9–15.

ISSN: 2582-0745
Vol. 8, No. 06; 2025

9. Tran, M. H., & Le, T. T. (2022). Current status and solutions for teaching probability and statistics in medical universities. *Journal of Practical Medicine, (100)*, 42–48.
10. GAISE College Report ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016*. American Statistical Association.