MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

NGUYEN THI KIM SON

RESEARCH ON THE APPLICATION OF DEEP LEARNING MODELS FOR PREDICTING LEARNERS' ACADEMIC PERFORMANCE

DOCTORAL DISSERTATION ON INFORMATICS

Hanoi - 2025

MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

NGUYEN THI KIM SON

RESEARCH ON THE APPLICATION OF DEEP LEARNING MODELS FOR PREDICTING LEARNERS' ACADEMIC PERFORMANCE

DOCTORAL DISSERTATION ON INFORMATICS

Major: Information systems

Code: 9 48 01 04

Verification of Graduate University of Science and Technology	Advisor 1	Advisor 2
KT. GIÁM ĐỐC PHÓ GIÁM ĐỐC		
Dr. Nguyen Thi Trung	**Assoc. Prof. Nguyen Huu Quynh**	**Assoc. Prof. Ngo Quoc Tao**

Hanoi - 2025

BỘ GIÁO DỤC VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

NGUYỄN THỊ KIM SƠN

NGHIÊN CỨU ỨNG DỤNG MỘT SỐ MÔ HÌNH SỬ DỤNG HỌC SÂU TRONG DỰ ĐOÁN KẾT QUẢ HỌC TẬP CỦA NGƯỜI HỌC

LUẬN ÁN TIẾN SĨ MÁY TÍNH

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Xác nhận của Học viện Khoa học và Công nghệ	Người hướng dẫn 1	Người hướng dẫn 2
KT. GIÁM ĐỐC PHÓ GIÁM ĐỐC HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM		
TS. Nguyễn Thị Trung	**PGS.TS. Nguyễn Hữu Quỳnh**	**PGS. TS. Ngô Quốc Tạo**

Hanoi - 2025

DECLARATION

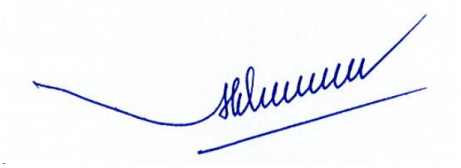
I declare that the results presented in this dissertation are my works, carried out under the guidance of my supervisors and in line with academic rules. The data and results are honestly reported. All sources of information used in the dissertation have been properly cited.

Some research results included in the dissertation were done in collaboration with other researchers, and they have agreed to their inclusion.

This dissertation was completed during my time as a doctoral student at the Vietnam Academy of Science and Technology.

Hanoi,2025

Dissertation Author



Nguyen Thi Kim Son

ACKNOWLEDGMENTS

During the process of working on this dissertation, I was fortunate to receive support, guidance, helpful feedback, and encouragement from many teachers, scientists, colleagues, friends, and family members.

First, I would like to sincerely thank Associate Professor Nguyen Huu Quynh and Associate Professor Ngo Quoc Tao. Their dedicated guidance and support were a great help throughout my research.

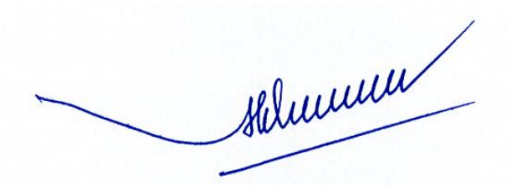
I am also very grateful to the faculty and scientists at the Institute of Information Technology - Vietnam Academy of Science and Technology, at the School of Information and Communications Technologies, Hanoi University of Industry and at the Hanoi Metropolitan University. Their advice and support played an important role in helping me complete this dissertation.

My thanks also go to the Board of Directors and the Graduate Training Department of the Graduate University of Science and Technology, Vietnam Academy of Science and Technology for creating a good environment for me to study and do research.

Finally, I want to thank my colleagues, friends, and family for always being there for me, encouraging me, and sharing this journey with me.

Hanoi,2025

Dissertation Author



Nguyen Thi Kim Son

CONTENTS

INTRODUCTION 1

1. General introduction 1

2. Research objectives 2

3. Research subjects and scope 2

4. Research methodology 4

5. Key contributions of the dissertation 4

6. Layout of the dissertation 5

7. Overview of main content flow 7

8. Significance of the dissertation 7

CHAPTER 1. OVERVIEW OF ACADEMIC PERFORMANCE PREDICTION FROM MACHINE LEARNING AND DEEP LEARNING APPROACHES 9

1.1. Research context and motivation 9

1.1.1. The transformative role and challenges of data and technology in modern education *9*

1.1.2. Approaches to predicting academic performance *9*

1.2. Machine learning and deep learning methods 10

1.2.1. Overview of machine learning *10*

1.2.2. Some deep learning models *12*

1.3. Overview of related research 16

1.3.1. Related works *16*

1.3.2. Research gap *21*

1.4. Datasets 22

1.4.1. HNMU1 dataset *22*

1.4.2. HNMU2 dataset *25*

1.4.3. VNU dataset *29*

1.4.4. International datasets *32*

1.4.5. Issues of privacy and sensitive data handling *32*

1.5. Evaluation metrics for predictive models 33

1.5.1. Some metrics for classification models *33*

1.5.2. Some metrics for regression models *33*

The conclusion of Chapter 1.....35
CHAPTER 2. EARLY PREDICTION OF SEMESTER GRADE POINT AVERAGE USING DEEP LEARNING APPROACHES..... 37
2.1. Problem formulation.....37
2.2. NeutroDL models.....39
2.2.1. The theoretical basis for model selection.....*39*
2.2.2. Proposed model.....*40*
2.2.3. Experiment.....*49*
2.2.4. Results and discussions.....*51*
2.3. NeutroGNT model 55
2.3.1. The theoretical basis for model selection.....*55*
2.3.2. Proposed model.....*56*
2.3.3. Experiments.....*60*
2.3.4. Results and discussions.....*62*
2.4. Appendix to Chapter 2.....65
2.4.1. Overview of Neutrosophy theory.....*65*
2.4.2. Summary of GAN and CGAN.....*68*
2.4.3. The Transformer model for the SGPA prediction task.....*70*
The conclusion of Chapter 2.....72
CHAPTER 3: ENHANCING THE PERFORMANCE OF EARLY GRADUATION CLASSIFICATION MODELS..... 73
3.1. Problem formulation 73
3.1.1. Early prediction of graduation classification problem*.....*73
3.1.2. Learning Analytics with graph data*.....*75
3.2. The LATCGAd model.....75
3.2.1. The theoretical basis for model selection.....*75*
3.2.2. Proposed model.....*78*
3.2.3. Experiments.....*81*
3.2.4. Results and discussion.....*87*
3.3. The AWG-GC model.....91
3.3.1. The theoretical basis for model selection.....*91*
3.3.2. Proposed model.....*93*

3.3.3 Experiments *97*

3.3.4. Results and discussions *110*

3.4. Appendix to Chapter 3 117

3.4.1. Wasserstein GANs (WGAN) *117*

3.4.2. The Transformer model for the task of predicting graduation classification *117*

3.4.3. Graphormer *118*

The conclusion of Chapter 3 120

CONCLUSION AND FUTURE DEVELOPMENT 122

LIST OF PUBLICATIONS 124

REFERENCES 125

SYMBOLS AND ABBREVIATIONS

No.	Abbreviation	Description
1	AI	Artificial Intelligence
2	ANN	Artificial Neural Network
3	CGAN	Conditional Generative Adversarial Network
4	CNN	Convolutional Neural Network
5	DL	Deep Learning
6	DNN	Deep Neural Network,
7	DT	Decision Tree
8	GAN	Generative Adversarial Network
9	GAT	Graph Attention Network
10	GCN	Graph Convolutional Network
11	GNN	Graph Neural Network
26	GPA	Grade Point Average
12	KNN	k Nearest Neighbor
13	LDA	Linear Discriminant Analysis
14	LR	Logistic Regresion
15	LSTM	Long Short-Term Memory
16	ML	Machine Learning
17	MAE	Mean Absolute Error
18	PCA	Principal Component Analysis
19	R^2	Coefficient of Determination
20	RF	Random Forest
21	RMSE	Root Mean Square Error
22	RNN	Recurent Neural Network
27	SGPA	Semester Grade Point Average
23	SVM	Support Vector Machine
24	XGBoost	Extreme Gradient Boosting
25	WGAN	Wasserstein Generative Adversarial Network

LIST OF TABLES

Chapter 1

Table 1. 1. Advantages and disadvantages of deep supervised learning techniques14
Table 1. 2. Results of student performance prediction using machine learning and deep learning techniques 18
Table 1. 3. Letter grade conventions and grade conversion....23
Table 1. 4. List of HNMU1 variables 24
Table 1. 5. Survey variables of the HNMU2 dataset 26
Table 1. 6. List of HNMU2 score variables 27
Table 1. 7. List of VNU score variables 30
Table 1. 8. Dataset description 32
Table 1. 9. Selection of evaluation metrics for the model....35

Chapter 2

Table 2. 1. Layer structure of DNN model....43
Table 2. 2. Layer structure of CNN model....43
Table 2. 3. Layer structure of CNN model....43
Table 2. 4. Layer structure of LSTM model 44
Table 2. 5. Layer structure of Transformer model....44
Table 2. 6. Model parameters 45
Table 2. 7. Description of the training dataset 49
Table 2. 8. Average error for cases 1, 2, 3 with Neutrosophic approach....52
Table 2. 9. Average error comparison for cases 1, 2, 3....54
Table 2. 10. The parameters of the models 60
Table 2. 11. Training dataset description 62
Table 2. 12. Demonstrated errors (averaged over 10 runs - case 1)....63
Table 2. 13. Demonstrated errors (averaged over 10 runs - case 2)....63
Table 2. 14. Demonstrated errors (averaged over 10 runs - case 3)....64

Chapter 3

Table 3. 1. Graduation classification based on final GPA 73
Table 3. 3. Description of the training dataset 81
Table 3. 4. Number of samples before and after creation with CGAN....82
Table 3. 5. Generator model parameters on the HNMU1, HNMU2, and VNU datasets....83
Table 3. 6. Discriminator model parameters....83
Table 3. 7. Prediction results on the HNMU1 dataset....88
Table 3. 8. Prediction results on the HNMU2 dataset....88

Table 3. 9. Prediction results on the VNU dataset.....89
Table 3. 10. Prediction results on the HNMU2 dataset.....110
Table 3. 11. Prediction results obtained on the VNU dataset 111
Table 3. 12. Prediction results obtained on the SATDAP dataset.....112
Table 3. 13. Per-class performance evaluation table of the AWG-GC model on the HNMU2 dataset 113
Table 3. 14. Per-Class Performance Evaluation Table of the AWG-GC Model on the VNU Dataset 113
Table 3. 15. Per-Class Performance Evaluation Table of the AWG-GC Model on the SATDAP Dataset.....114

LIST OF FIGURES

Figure 0. 1. Student performance prediction system *5*

Chapter 1

Figure 1. 1. Classification of machine learning algorithms *11*
Figure 1. 2. The ML models: LR, KNN, RF and SVM *12*
Figure 1. 3. Deep Learning models *13*
Figure 1. 4. Model architecture of DNN, CNN and RNN *14*
Figure 1. 5. Transformer architecture *15*
Figure 1. 6. The structure of HNMU1 dataset *24*
Figure 1. 7. The structure of HNMU2 dataset *29*
Figure 1. 8. The structure of VNU dataset *29*

Chapter 2

Figure 2. 1. Prediction framework for Case 1 ..... *38*
Figure 2. 2. Prediction framework for Case 2 ..... *38*
Figure 2. 3. Prediction framework for Case 3 ..... *38*
Figure 2. 4. The NeutroDL models *41*
Figure 2. 5. Loss function value chart for training and validation of models *46*
Figure 2. 6. The pineline of NeutroDL model *47*
Figure 2. 7. The neutrosophic functions for the concepts a) Good, ... *50*
Figure 2. 8. Graph of prediction (neutrosophic data, Case 1) *52*
Figure 2. 9. Graph of prediction (neutrosophic data, Case 2) *53*
Figure 2. 10. Graph of prediction (neutrosophic data, Case 3) *53*
Figure 2. 11. NeutroGNT model *57*
Figure 2. 12. The pineline of NeutroGNT model *58*
Figure 2. 13. The single-valued trapezoidal neutrosophic functions *67*
Figure 2. 14. CGAN model *69*
Figure 2. 15. The basic Transformer model for the SGPA prediction task. *70*

Chapter 3

Figure 3. 1. The LATCGAd model *78*
Figure 3. 2. The pineline of LATCGAd model *80*
Figure 3. 3. Training the CGAN model (in the LATCGAd model) *85*
Figure 3. 4. FID values *86*
Figure 3. 5. Training the Transformer model (in the LATCGAd model) *87*
Figure 3. 6. Confusion Matrices (in the LATCGAd model) *90*
Figure 3. 7. The AWG-GC model *93*
Figure 3. 8. The pineline of AWG-GC model *95*

Figure 3. 9. Number of samples per class in the SATDAP dataset 98
Figure 3. 10. Autoencoder model training according to AutoGAT 103
Figure 3. 11. Training of the GAT model according to AutoGAT 103
Figure 3. 12. Autoencoder model training according to AWG-GAT.....104
Figure 3. 13. Training of the WGAN model on the HNMU2 dataset according to AWG-GAT. 105
Figure 3. 14. Training of the WGAN model on the VNU dataset 105
Figure 3. 15. Training of the GAT model according to AWG-GAT: 106
Figure 3. 16. Training of the Graphomer model 107
Figure 3. 17. Autoencoder model training according to AWG-GC.....108
Figure 3. 18. Training of the WGAN model on the HNMU2 dataset according to AWG-GC 108
Figure 3. 19. Training of the WGAN model on the VNU dataset 108
Figure 3. 20. Training of the WGAN model on the SATDAP dataset according to AWG-GC 109
Figure 3. 21. Training of the Graphomer model according to AWG_GC.....109
Figure 3. 22. Confusion Matrices (in the AWG-GC model) 114
Figure 3. 23. Transformer model for the task of predicting graduation classification. 118
Figure 3. 24. The Graphormer model.....118

INTRODUCTION

1. General introduction

The rapid development of data science and artificial intelligence (AI) in education has opened new opportunities to enhance teaching and learning in the digital transformation era ([1] - [3]). Among these, predicting students' academic performance has become a key application, enabling the early detection of at-risk learners and timely interventions ([1], [4]), in line with the goals of personalized learning and improving graduation rates ([5] - [7]).

However, most existing studies still rely on traditional machine learning models such as LiR, LR, SVM, DT, KNN, and NB ([8], [9]). While simple and interpretable, these models are limited in capturing nonlinear, sequential, and multifactorial characteristics of educational data ([1], [2], [10], [11]). Deep learning, particularly LSTM and Transformer architectures, offers a promising alternative by effectively modeling sequential behaviors and complex relationships ([6], [13], [14]).

In theory, a practical solution would be to apply pre-trained deep learning models or transfer learning techniques, which have proven effective in domains like computer vision and natural language processing when data is limited. However, in education worldwide, there is still a lack of large, standardized, and publicly available datasets, together with a shortage of domain-specific pre-trained models, which limits the adoption of transfer learning in this field ([15]; [16]). In Vietnam, for instance, the Ministry of Education and Training issued Circular No. 42/2021/TT-BGDĐT dated November 30, 2021, on the Regulations of the Education Database (Ministry of Education and Training of Vietnam, 2021), which provides a framework for building a unified national education database. Nevertheless, its implementation remains fragmented and not yet openly accessible for research, reflecting the broader global challenges.

To address these constraints, this study proposes deep learning–based and hybrid approaches that integrate data augmentation, feature selection, and advanced optimization techniques, combining the representational power of deep models with the interpretability of traditional methods ([11], [18] - [20]).

Based on this rationale, the dissertation investigates deep learning and hybrid models for predicting academic performance, aiming to process sequential data, incorporate diverse contextual factors, and ensure reliable

performance under limited data conditions. This work contributes to advancing Learning Analytics, supporting evidence-based decision-making in higher education, and expanding the role of AI in educational research.

2. Research objectives

General Objective: To research and develop machine learning and deep learning models for analyzing educational data with the goal of early prediction of student's academic performance.

Specific Objectives:

(1) To propose and compare the performance of modern machine learning and deep learning models: k-Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machines (SVM), Logistic Regression (LR), Random Forests (RF),Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Transformers,…for predicting academic performance (e.g., semester GPA, graduation classification), with an emphasis on improving accuracy and generalizability.

(2) To construct hybrid deep learning models, perform appropriate feature selection, and apply data augmentation techniques to address the challenges of small-scale and heterogeneous educational datasets.

The experimental evaluation will be conducted using training datasets collected from both domestic and international universities and colleges.

3. Research subjects and scope

Research Subjects:

Early prediction problems related to student academic performance can be categorized into several specific types, depending on the objectives and scope of the analysis. Specifically:

- Grade prediction problems: including the prediction of semester Grade Point Average (GPA), annual GPA, cumulative GPA, individual course scores, short-term course results, continuous assessment scores, etc.

- Classification prediction problems: including the prediction of academic classifications for individual courses, semesters, stages of study, or final graduation classifications.

These prediction tasks play an important role in academic early warning systems, helping institutions identify students at risk of failing courses,

repeating semesters, or being unable to graduate on time. They also support the recommendation of interventions to improve student performance and provide data-driven evidence for educational administrators to make informed decisions.

In the context of this dissertation, we focus on two core prediction problems:

- The early prediction of semester GPA,
- The early prediction of final graduation classification.

Hereinafter, the term "academic performance" as used in this dissertation refers specifically to "semester GPA" or "graduation classification".

In addition, the dissertation also considers research subjects at the model level, including:

- Traditional machine learning algorithms (KNN, DT, SVM, LR, RF) as baselines.
- Deep learning architectures (DNN, CNN, RNN, LSTM, Transformer, GNN/GCN/GAT) for sequential and relational data.
- Hybrid and advanced models (NeutroDL, NeutroGNT, LATCGAd, AWG-GC) to address small, imbalanced, and uncertain datasets.

Research Scope: Modern machine learning and deep learning models, including hybrid model architectures.

Datasets collected from Hanoi Metropolitan University (HNMU), Vietnam National University (VNU), and selected publicly available international datasets for reference and benchmarking.

The data used in this research includes:

- Student grade records, collected from university academic management systems.
- Survey data on factors related to students, such as personal information, preferences, academic background prior to university, family circumstances, and socio-occupational factors that may influence academic performance, etc.
- Institutional data from higher education institutions, including facilities, curriculum, and faculty-related information, etc.

4. Research methodology

The research adopts a combination of theoretical study, literature review, empirical research, and survey-based investigation.

Theoretical research: Theoretical analysis is conducted to evaluate the advantages and limitations of various machine learning and deep learning models in predicting academic performance. Based on this analysis, appropriate models are selected for application to the available datasets. These models include, but are not limited to: *k-Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machines (SVM), Logistic Regression (LR), Random Forests (RF), Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Transformers, Graph Neural Networks (GNN), Graph Convolutional Networks (GCN), Graph Attention Networks (GAT), Conditional Generative Adversarial Networks (CGAN), Wasserstein GANs and Graphomer.*

The study includes: (i) a literature review to synthesize prior works, highlight trends, strengths, and limitations for model development; (ii) surveys and data collection at Vietnam National University and Hanoi Metropolitan University to build student datasets; (iii) empirical experiments validating machine learning, deep learning, and hybrid models on both local and benchmark datasets; and (iv) technical implementation using Python and MATLAB for model development, evaluation, and comparison.

5. Key contributions of the dissertation

(1) Two novel methods, NeutroDL and NeutroGNT models, are proposed, integrating the neutrosophic process into deep learning models to enhance early SGPA prediction performance.

(2) Two novel hybrid models are proposed: LATCGAd, and AWG-GC for the prediction of graduation classification for students.

(3) Development of 03 multi-attribute datasets from diverse sources and proposal of analytical frameworks tailored to educational data.

From an information systems perspective, where data, software, hardware, people, and processes are integrated to support decision-making, this dissertation makes the following contributions:

- Data sources: Constructed and standardized educational datasets (HNMU, VNU, and survey data), providing a reliable foundation for Educational Data Mining (EDM) and Learning Analytics (LA).
- Data pipeline: Designed a rigorous processing, normalization, and integration pipeline to ensure consistency, quality, and model reliability.
- Prediction & Analytics: Applied advanced deep learning and hybrid models (NeutroDL, NeutroGNT, LATCGAd, AWG-GC) to predict SGPA and graduation classification, leveraging CPU/GPU infrastructures for efficient training and real-time analysis.
- User services: Delivered prediction and analysis results that can be integrated into early-warning systems, reporting tools, and decision-supportservices for students, lecturers, advisors, and administrators - thereby fostering intelligent, adaptive, and student-centered educational management.

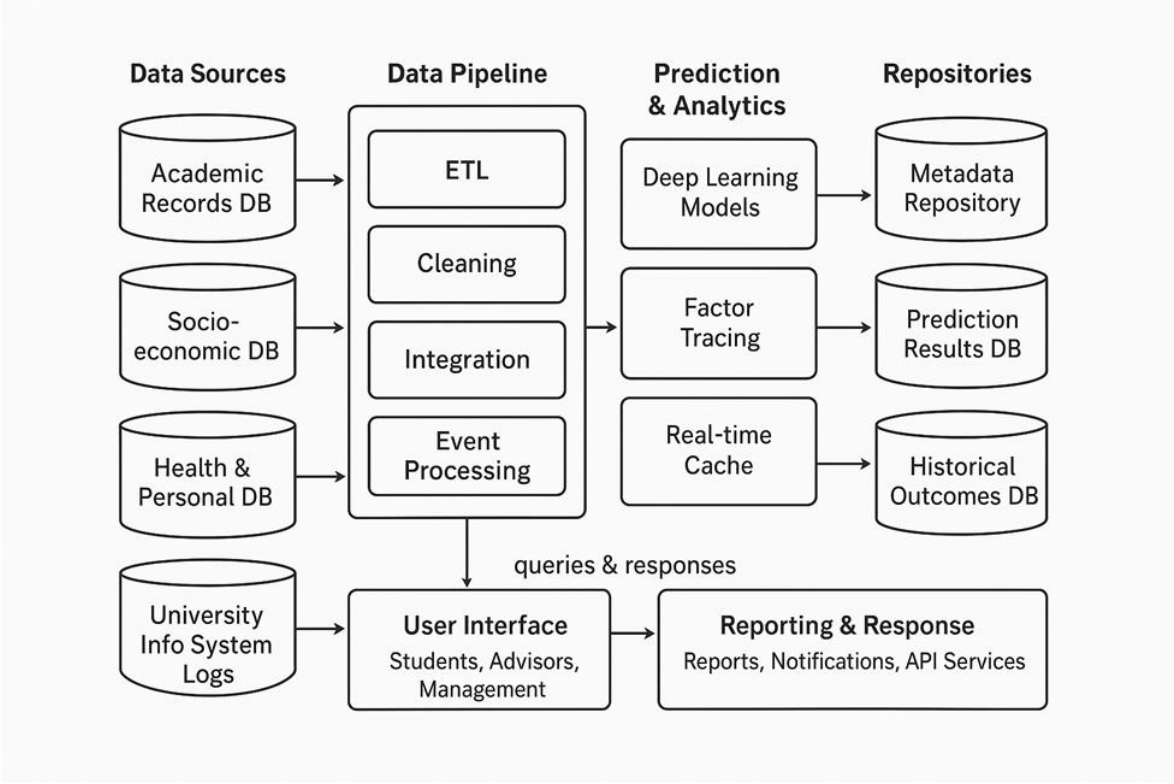


Figure 0. 1. Student performance prediction system

6. Layout of the dissertation

This dissertation is presented with a structure that includes an introduction, three main chapters, a conclusion and future development, a list of publications, and references, as follows:

The **Introduction** outlines the scientific significance and urgency of the topic, as well as the reasons for choosing the research topic. It also presents the objectives, subject, scope, methods, key contributions of the dissertation, and contents of the study.

Chapter 1 provides an overview of educational data analysis, highlighting machine learning and deep learning applications in predicting student's academic performance. It reviews related research to establish the dissertation's motivation and introduces three key datasets (HNMU1, HNMU2, VNU) from Hanoi Metropolitan University and Vietnam National University, which form the experimental basis for the models developed in later chapters.

Chapter 2 focuses on SGPA prediction using deep learning models combined with Neutrosophy theory to manage data uncertainty. Models such as DNN, CNN, RNN, LSTM, and Transformer are implemented in neutrosophic environments (Neutrosophic DLs) to predict next-semester GPA from historical academic data. To further enhance performance, the chapter introduces NeutroGNT, a hybrid model integrating data neutrosophicization, CGAN-based data generation, noise injection, and Transformer, improving prediction accuracy and adaptability in uncertain conditions.

Chapter 3 shifts to predicting students' graduation classification, a more long-term and system-level task. It introduces LATCGAd and AWG-GC, which leverage graph-based models (Graphformer), advanced GANs (CGAN, WGAN), and Autoencoders, along with AdaLN for stability, to handle small and imbalanced datasets. These models expand data and improve predictive performance, offering higher accuracy, robustness, and scalability for educational analytics systems.

In the **Conclusion and Future development**, the dissertation synthesizes the achieved results and draws several conclusions, while also outlining future research directions based on the findings.

List of publications: The dissertation includes a list of 08 papers authored by the researcher, which have been published or accepted for publication in domestic and international journals and conference proceedings.

Finally, a list of references used in the dissertation is provided.

7. Overview of main content flow

Apart from Chapter 1, which provides an overview and introduces the research problem and datasets, Chapters 2 and 3 form a cohesive structure, presenting two complementary approaches to the early prediction of student academic performance based on both academic and non-academic data. In terms of problem nature, Chapter 2 addresses a regression task aimed at predicting semester GPA - a continuous, quantitative indicator that reflects short-term academic progress. In contrast, Chapter 3 focuses on a classification task to predict graduation classification - a discrete, system-level, and longer-term outcome.

These two tasks are inherently linked: the GPA results from multiple semesters form a key part of the input for the graduation classification model. Accurate SGPA predictions in earlier stages thus help improve the performance of classification in later stages.

From a modeling perspective, the deep learning architectures developed in Chapter 2 (such as DNN, LSTM, Transformer), combined with techniques for handling data uncertainty (Neutrosophy) and data augmentation (CGAN), lay a crucial technical and experimental foundation for the extended models in Chapter 3. There, new models like LATCGAd and AWG-GC are developed by building upon and integrating advanced components such as WGAN, Graphformer, and Autoencoder, effectively addressing the classification problem under imbalanced and complex data conditions.

The strong connection between chapter 2 and chapter 3 is reflected not only in the data relationship between the tasks but also in the progression of model development, which is carefully aligned with the characteristics and objectives of each educational prediction task.

8. Significance of the dissertation

The dissertation holds both academic and practical significance in the context of digital transformation in higher education:

Academic Significance:

The research contributes to advancing the field of Educational Data Mining (EDM) by integrating deep learning models into educational information systems. The proposed models for predicting GPA and graduation classification, trained on real-world data with high accuracy, provide a strong scientific foundation for applying artificial intelligence in analyzing student learning behaviors.

Practical Applications:

The findings of the dissertation have high applicability in educational management, particularly in:

Personalized learning: supporting academic advising and customized learning pathways for students;

Early identification of at-risk learners: enabling timely interventions by educational institutions;

Data-driven decision-making: assisting in educational planning, evaluation, and policy development.

System-level Contribution:

The dissertation exemplifies the integration of deep learning technologies with core components of educational information systems (data - hardware - software - people - processes), aiming to build a smart, adaptive, and efficient learning environment in the era of artificial intelligence.

The results of this dissertation have been presented at:

1. FS&IS Seminar, School of Information and Communications Technology, Hanoi University of Industry.

2. VNICT Conference, 2024.

3. MCO Conference, 2025.

CHAPTER 1. OVERVIEW OF ACADEMIC PERFORMANCE PREDICTION FROM MACHINE LEARNING AND DEEP LEARNING APPROACHES

*This chapter outlines the research context and motivation (Section 1.1), emphasizing the importance of early prediction of student performance. Section 1.2 reviews key machine learning and deep learning foundations. Section 1.3 synthesizes related domestic and international studies, highlighting research gaps. Section 1.4 introduces experimental datasets, including three from Vietnamese universities (**[**CT1**]**, **[**CT3**]**, and **[**CT4**]**) and several international datasets for benchmarking. Finally, Section 1.5 presents the evaluation metrics used to assess and compare model performance in later chapters.*

1.1. Research context and motivation

1.1.1. The transformative role and challenges of data and technology in modern education

The Fourth Industrial Revolution, characterized by rapid data growth, has turned data into a strategic asset essential for decision-making and efficiency across sectors, including education ([21]). In this domain, LMS, online platforms, and intelligent technologies generate vast datasets that enable progress tracking, personalized learning, and evidence-based management ([1]). While these technologies provide opportunities to optimize engagement and outcomes ([2]), they also pose challenges in data quality, unstructured information, privacy, and the technical requirements of advanced analytics such as machine learning and deep learning. To address these issues, educational data science has emerged as an interdisciplinary field that integrates computer science, education, psychology, and statistics to collect, process, and analyze data for enhancing learning and teaching ([5]; [6]).

1.1.2. Approaches to predicting academic performance

In recent years, there has been a growing trend of students at higher education institutions receiving academic warnings or being forced to withdraw. Despite decades of efforts to improve student retention, the rates have remained low ([22]). According to a report by [23], the average retention rate from the first to the second year was only 66.5%. Nearly one-fourth of students leave college after their first year ([22]). One of the main causes of poor academic performance is that students often select courses that do not

match their capabilities and lack an effective study plan. This results in students either dropping out or extending their study duration, wasting time and resources for families, institutions, and society ([24]).

Academic success is a key factor in helping students persist in their studies ([25]; [26]), and the risk of dropping out decreases as academic performance improves ([27]). Therefore, an effective way to increase retention is to improve academic performance through early prediction of academic performance. This enables early warnings about risks of failure and supports decision-making in developing optimal study plans for students, advisors, and administrators ([28]; [29]).

Predictive results not only help students choose subjects appropriate to their abilities but also assist instructors and academic managers in identifying students who need additional support, thereby reducing academic warnings and forced withdrawals ([30]). In turn, this saves time and costs while improving the quality of education. As such, predicting student academic performance has become a crucial research topic in the field of LA, attracting increasing attention.

Among these problems, this dissertation focuses on two main tasks: predicting semester GPA scores and early prediction of graduation classification.

1.2. Machine learning and deep learning methods

1.2.1. Overview of machine learning

Machine learning is a field of study focused on developing computer algorithms that improve automatically through experience. It is commonly categorized into four types: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning ([31]). In supervised learning, models learn a mapping function from labeled training data. Unsupervised learning involves data without labels, aiming to discover hidden patterns or structures. Semi-supervised learning combines both labeled and unlabeled data to improve learning accuracy. In reinforcement learning, an agent interacts with its environment to learn actions that maximize cumulative reward. Figure 1.1 illustrates the classification of machine learning systems.

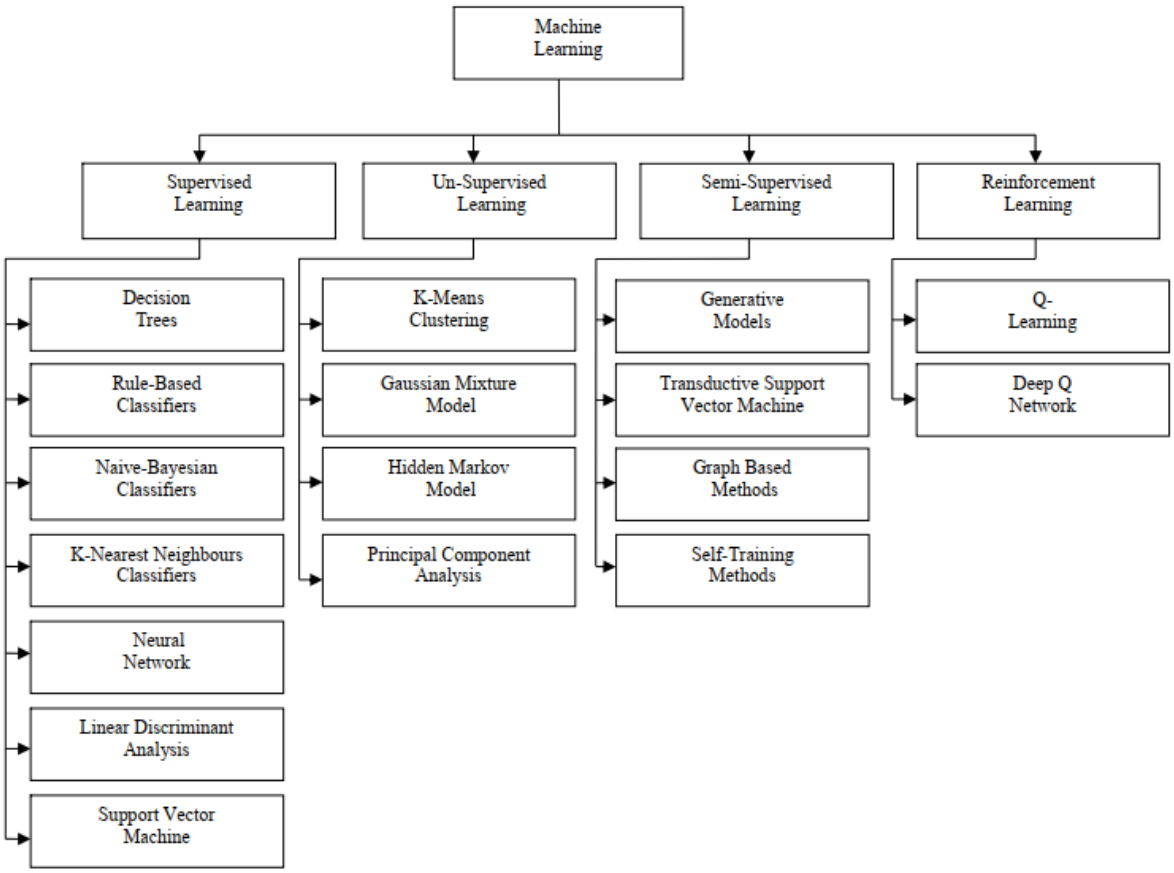


Figure 1. 1. Classification of machine learning algorithms ([31])

LR is a classical statistical method for identifying predictors of binary outcomes. KNN is a non-parametric algorithm that performs classification or regression based on the majority vote or average of the k nearest data points. DT uses a binary tree structure to split data via decision rules but is prone to overfitting without pruning. RF, as an ensemble method, aggregates multiple randomized DTs to improve accuracy and reduce overfitting. SVM classifies data by finding the optimal hyperplane maximizing class separation and

effectively handles non-linear patterns through kernel functions (see Figure 1.2) ([32]).

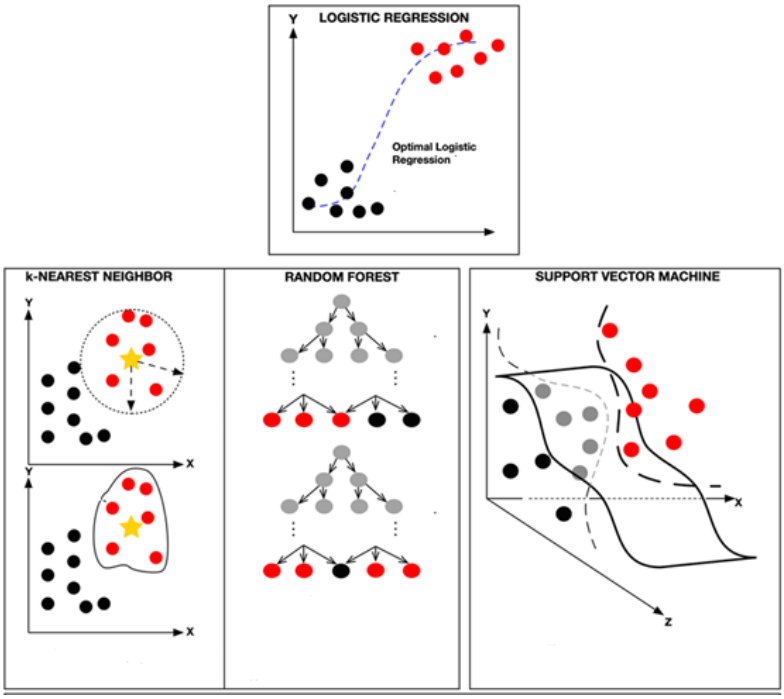


Figure 1. 2. The ML models: LR, KNN, RF and SVM ([32])

1.2.2. Some deep learning models

The foundation of artificial neural networks (ANN) was introduced in 1943 as a mathematical model of an artificial neuron ([13]). In 2006, the concept of deep learning (DL) emerged, extending ANN into multi-layer architectures with significantly enhanced learning capabilities. In recent years, DL has achieved remarkable success in solving complex problems such as anomaly detection, object recognition, disease diagnosis, semantic segmentation, social network analysis, and video recommendation systems ([33]; [34]).

Deep learning models are generally classified into four main categories: deep supervised learning, unsupervised learning, reinforcement learning, and hybrid models. Figure 1.3 illustrates these categories along with representative models for each.

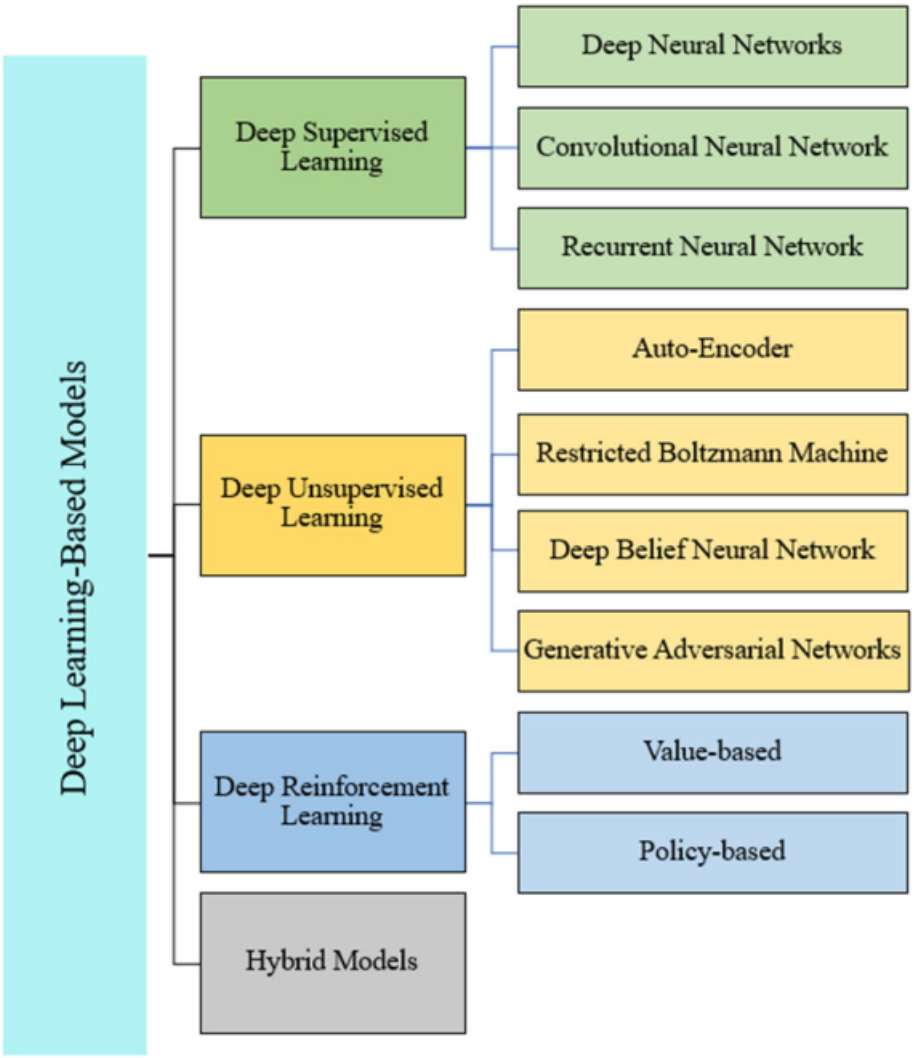


Figure 1. 3. Deep Learning models ([13])

Within the category of deep supervised learning, three prominent models have been identified: Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and models based on RNN(RNN), as illustrated in Figure 1.4. ANN and DNN (with multiple hidden layers) model complex nonlinear relationships, CNN extract spatial features and patterns for image-related tasks using convolution and pooling layers, while RNN (including LSTM) capture temporal dependencies and long-term patterns in sequential or time-series data. Table 1.1 summarizes the key advantages and limitations of the deep learning models: DNN, CNN, and RNN.

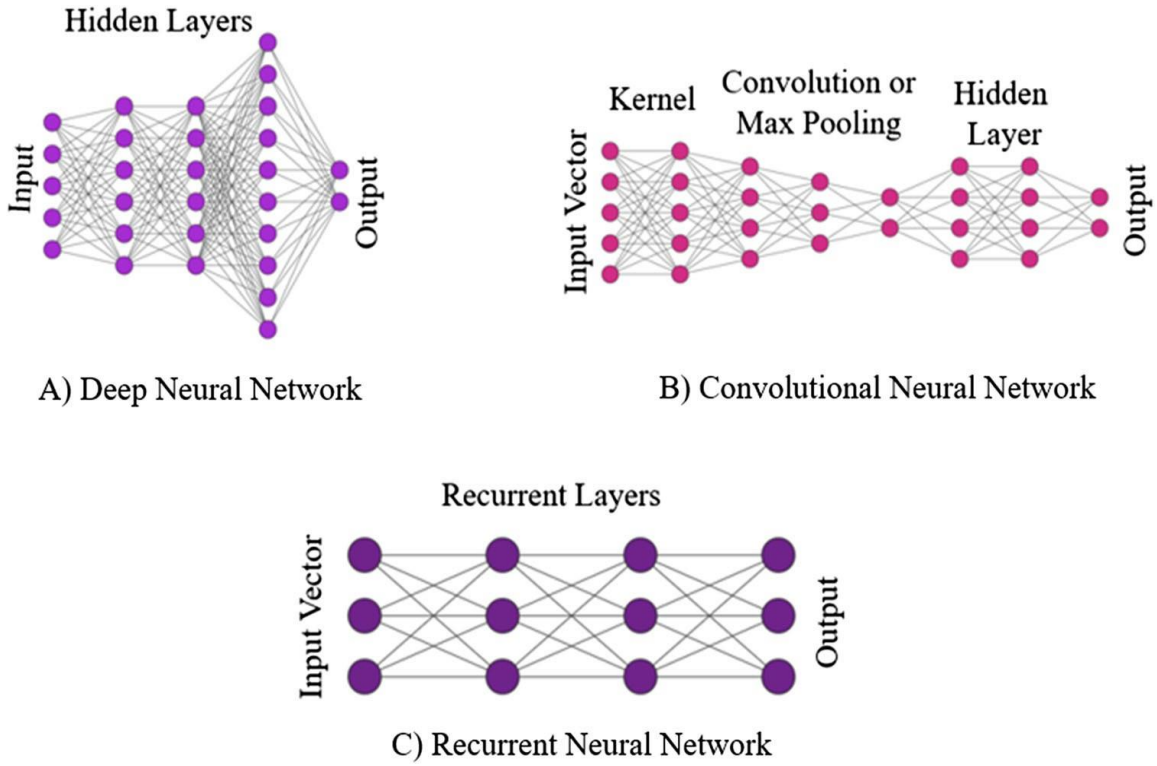


Figure 1. 4. Model architecture of DNN, CNN and RNN ([13])

Table 1. 1. Advantages and disadvantages of deep supervised learning techniques

Learning methodology	**Category**	**Advantage**	**Disadvantage**
Deep supervised learning	Deep neural networks	Tendency to high nonlinear relationships, Easy to develop	Slow learning, Hard for parameter tuning, Insufficient for high-dimensional input space
	Convolutional neural network	Ability to capture spatial correlations, high potentiality at generalization	Difficult parameter tuning, High computational cost
	Recurrent neural network	Sometimes fast converge with minimum parameters, improve the vanishing gradient issues	Difficult parameter tuning, Poor spatial feature representations

Transformer

The architecture of the Transformer model, originally proposed by Vaswani et al. ([35]), is presented in Figure 1.5. It consists of two principal components: the encoder and the decoder. The encoder is composed of a series of identical layers, each containing two sub-components, a multi-head self-attention mechanism and a position-wise feed-forward neural network. To enhance training stability and gradient flow, residual connections and layer normalization are applied following each sub-layer, as illustrated in Figure 1.5.

Unlike conventional convolutional networks, which combine feature aggregation and transformation in a single step, the Transformer architecture separates these processes: self-attention handles aggregation, while the feed-forward layer performs transformation. Similarly, each Transformer decoder layer, stacked like those in the encoder, consists of three sub-layers: self-attention, feed-forward (same as the encoder), and a cross-attention mechanism that attends to the encoder's output.

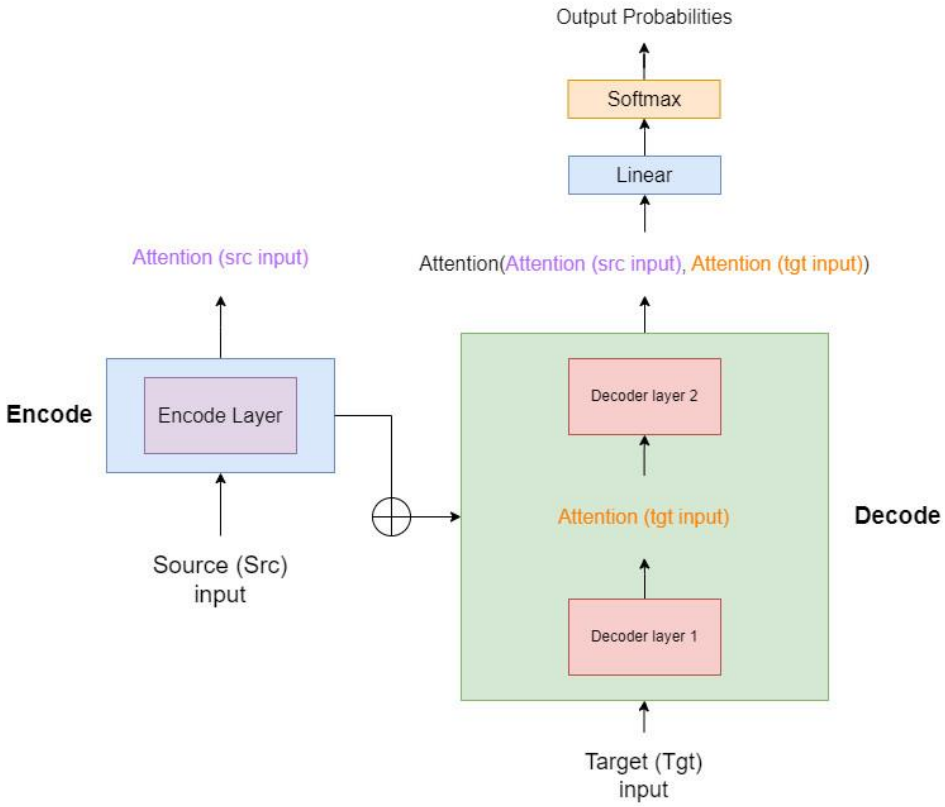


Figure 1. 5. Transformer architecture ([35])

The original Transformer model in [35] was trained for machine translation. The input to the encoder is a sequence of words (i.e., a sentence) in

the source language. Positional encoding is added to the input sequence to capture the relative position of each word. These positional encodings have the same dimensionality as the model input $d = 512$, and can be either learned or fixed.

As an auto-regressive model, the Transformer decoder uses previously generated predictions to produce the next word in the sequence. Consequently, the decoder receives input from both the encoder and the preceding output tokens to generate the next token in the target language. To support residual connections, the output dimension of all sub-layers is kept constant, i.e., $d = 512$

The dimensions of the query, key, and value weight matrices in the multi-head attention mechanism are typically set to $d_q = 64$; $d_k = 64$; $d_v = 64$.

While deep learning models have demonstrated remarkable performance, they also exhibit limitations, particularly in hyperparameter tuning and their sensitivity to data volume. These limitations can hinder their deployment in various real applications. Nonetheless, each DL model possesses characteristics that make it suitable for specific tasks. To address these shortcomings, hybrid DL models have been proposed, which combine individual architectures to overcome application-specific challenges ([36]).

1.3. Overview of related research

1.3.1. Related works

a) Emergence of EDM and Learning Analytics

In recent years, EDM and LA have become prominent research directions in educational science, fueled by the growth of digital technologies and online learning platforms such as LMS and MOOCs. These environments generate large-scale data on learner behaviors, enabling the application of AI, ML, and DL methods to predict, classify, and support learning processes. One of the central problems is the prediction of academic performance, including grades, graduation likelihood, dropout risk, and achievement classification. Supervised machine learning algorithms such as DT, NB, LR, KNN, RF, and SVM have been widely used and proven effective in identifying risk factors and enabling early intervention ([37]).

b) Deep Learning and hybrid approaches

Alongside traditional models, deep learning has gained increasing attention for its ability to capture nonlinear and sequential relationships.

Architectures such as DNN, RNN, LSTM, and GNN have been applied to improve prediction accuracy and analyze learning behavior ([38]). To address challenges such as imbalanced and incomplete datasets, researchers have also explored hybrid methods, including fuzzy logic integration, feature selection using genetic algorithms, and data augmentation techniques like SMOTE ([39]). While these methods improve accuracy, several works highlight concerns about computational complexity and potential overfitting.

c) Academic performance prediction models

Academic performance prediction remains one of the most extensively studied problems in EDM. Traditional algorithms (DT, KNN, NB, LR, rule-based systems) continue to be used in predicting GPA, academic classification, or graduation outcomes ([4]; [40]; [41]). For example, Waheed et al. ([38]) and Wasif et al. ([42]) focused on identifying at-risk students, while Elbadrawy et al. ([43]) applied linear regression and matrix factorization. However, these models often neglect sequential dependencies between courses, reducing their practical relevance. Fei and Yeung ([44]) applied HMMs and RNNs to MOOC datasets, though their findings were limited to online contexts.

d) Learner behavior and personalized interventions

Another important research strand emphasizes the analysis of learner behavior for personalized support. Okubo et al. ([45]) applied RNN to predict academic performance, but the study was restricted to a small cohort, limiting generalization. Corrigan and Smeaton ([46]) and Waheed et al. ([38]) confirmed the potential of RNNs and LSTMs in online environments, though the specific impact of interaction types remained unclear. Other works (e.g., Anggrawan et al. [39]) applied SMOTE and genetic algorithms to tackle imbalance, while Christou et al. ([47]) explored grammatical evolution for feature selection. Despite improvements, challenges such as training cost, scalability, and risk of overfitting persist.

e) Applications of ML and DL in education

Recent studies further confirm the potential of ML and DL models in predicting student success. Algorithms including LR, DT, neural networks, RF, and XGBoost have been employed to analyze exam scores, study habits, and participation data ([48]; [49]; [50]). For instance, Sapkota et al. ([48]) predicted graduation rates using XGBoost and RF, while Halat et al. ([49]) applied ML

to progression analysis in health sciences. Although deep learning models achieve high accuracy in online learning contexts ([51]), limitations remain: dependency on historical data ([43]; [52]), lack of temporal modeling ([44]), neglect of causal relationships ([43]), small sample sizes ([45]), imbalanced datasets ([53]), and inconsistent evaluation metrics across studies ([37]).

f) Context-specific studies and challenges

Several works have investigated education data in specific institutional contexts. An et al. ([54]) used statistical analysis to explore factors influencing early-year student performance, though without predictive modeling. Other research employed recommender-system toolkits (e.g., Mymedialite) to analyze student competency ([55]; [56]), yet noted barriers in adapting generic ML models to educational logic. Uyên and Tâm ([41]) applied Naïve Bayes and LR for academic performance and dismissal risk prediction, while Nghe and Dinh ([57]) designed an AI-based admission advisory system, though still at an experimental stage. More advanced studies have begun integrating deep learning: e.g., a multilayer perceptron (MLP) with 18 features ([58]) and a CNN model with 21 features ([59]) to predict student performance. Despite these advances, reliance on traditional input features (e.g., GPA, gender) limits personalization and reduces adaptability.

Table 1. 2. Results of student performance prediction using machine learning and deep learning techniques

Study	Purpose	Dataset	Method	Results
Thai-Nghe, Horváth, and Schmidt-Thieme, 2011 [40]	To predict student performance in a course	2 datasets from KDD Cup 2010	MF (Matrix Factorization Model)	Tensor-based factorization can be useful for predicting student performance
Fei and Yeung, 2015 [44]	To analyze learning behavior sequences for predicting outcomes	2 datasets from MOOC	Hidden Markov Model (HMM), RNN	Explored learning progression; limitations in traditional education environments
Elbadrawy et al., 2016 [43]	To predict academic performance and	Course data and student achievements	LiR, Matrix Factorization	Effective in centralized environments;

	personalize education			does not consider course order
Iam-On and Boongoen, 2017 [60]	To cluster students for personalized teaching	811 student data from MFLU	Clustering (k-means)	Effective clustering; lacks real-time data, reducing accuracy
Okubo et al., 2017 [45]	To predict student grades	108 students	RNN	Using log data from 6 weeks, accuracy was above 90%
Xu, 2017 [4]	To predict student performance	1,169 data from UCLA	Latent Factor Method	Latent factor method outperforms benchmark approaches
Corrigan and Smeaton, 2017 [46]	To explore how student interactions with virtual learning environments can predict performance	2,879 data from VLE	RF, RNN, simple LSTM	RNN outperforms all other algorithms
Zafar Iqbal et al., 2019 [52]	To suggest improvements in grades using recommendation models	Data on student course scores and activities	Collaborative Filtering (CF), SVD, NMF, RBM	Improved personalized recommendations; not tested across different academic fields
Waheed et al., 2019 [38]	To predict students at risk of underperforming	Behavioral data from online learning platforms	RNN, LSTM	BiLSTM achieved 90.16% accuracy; effective but requires large datasets
Okubo, 2019 [45]	To predict student grades from specific course behavior	108 students from a university course	RNN	Good grade prediction; small sample size, lacks generalizability

Uyên and Tâm (2019) [41]	Predict at-risk students likely to be dismissed	Student academic records from an unspecified university	NB, LR	Identified critical courses and risks, but lacked higher-order feature learning
Anthony Anggrawan, 2020 [39]	To improve predictions through data processing and feature selection	Data on grades + personal information	SMOTE, Genetic Algorithm + SVM	Improved accuracy; risk of overfitting if not controlled properly
Sang et al. (2020) [58]	Predict student academic performance	Student records from Can Tho University	Multi-Layer Perceptron (MLP) with 18 features	Promising results using gender and GPA, but model lacked behavioral and unstructured data
Dien et al. (2021) [59]	Predict academic outcomes using deep learning	Data from a multi-disciplinary Vietnamese university	CNN	Applied CNN successfully, but relied on basic features and lacked personalized recommendation ability
Alturki et al., 2023 [53]	To handle imbalanced data for predicting academic performance	Imbalanced class learning data	RF + Oversampling	Improved accuracy; risk of generating fake samples leading to data bias
Christou et al., 2023 [47]	To select optimal features for an RBF model	Complex learning data	Evolutionary Grammar + RBF kernel	Long training time and high computational cost
Halat et al., 2023 [49]	To predict academic performance at Qatar University	Medical and health science student data	XGBoost	XGBoost provided the most accurate results in predicting academic performance

Sapkota et al., 2025 [48]	To predict graduation and dropout rates of students	Student data from Qatar University	XGBoost, RF, AdaBoost	XGBoost model achieved 92% accuracy, outperforming other models

1.3.2. Research gap

After reviewing the current body of research both in Vietnam and internationally on the application of machine learning and deep learning in educational data science, it is evident that most existing studies still focus on utilizing standalone machine learning models, such as linear regression, DT, RF, or SVM, for tasks like academic performance prediction, student classification, or dropout risk detection. These models are generally considered easy to implement, interpretable, and perform relatively well on medium-sized and low-dimensional datasets.

However, the effectiveness of traditional machine learning models remains limited when applied to more complex educational problems, particularly those involving temporal sequences or strong nonlinear relationships. Moreover, the majority of current studies rely heavily on static data (e.g., semester grades, exam scores), and have yet to fully leverage dynamic, longitudinal information that reflects the learning process over time.

In response to these limitations, recent research has increasingly advocated for the adoption of deep learning models, particularly architectures designed for sequential data processing, such as RNN, LSTM, and Transformer models, to better capture temporal features in educational datasets. DL models offer the advantage of automatic feature learning and complex representation extraction without the need for manual feature engineering, thereby significantly enhancing predictive accuracy.

Furthermore, a promising direction gaining attention is the use of hybrid models, which combine deep learning with traditional machine learning approaches, or integrate multiple deep learning architectures (e.g., CNN, LSTMs, and Transformers enhanced with customized attention mechanisms). These hybrid models have the potential to deliver superior performance by combining the nonlinear learning power of deep learning with the interpretability and robustness of classical algorithms.

Despite these advantages, a major obstacle lies in the lack of high-quality, structured, and temporally rich educational data. Educational datasets are often small-scale, fragmented, heterogeneous, and lack standardization, posing significant challenges for training deep learning models, which typically require large datasets to reach optimal performance. Moreover, sequential data reflecting learning trajectories are rarely collected or shared due to privacy and data protection concerns. This further impedes the development and benchmarking of models on standardized datasets.

In summary, many current studies still rely on standalone machine learning or deep learning models with limited performance. The shift toward deep learning and hybrid approaches opens up promising opportunities to improve the accuracy and generalizability of academic performance prediction. However, realizing this potential will require addressing key data-related challenges, specifically, constructing high-quality sequential datasets, standardizing input features, and developing techniques tailored to small, heterogeneous datasets commonly found in educational contexts.

1.4. Datasets

1.4.1. HNMU1 dataset

This section introduces the dataset constructed from academic records at Hanoi Metropolitan University (HNMU), a public institution governed by the Hanoi People's Committee [CT1].

The raw data is provided by the training departments and the Student Management and Training Office. All data, including student management status (tuition, personal information, etc.), entrance exam scores, foreign language scores, computer science scores, and scores for each course completed by the students, is divided into 8 semesters across 4 academic years. Student academic performance is evaluated at the end of each semester or academic year, based on the results of the modules required by the training program that the student has completed. The average grade of the modules taken by a student in a semester (semester GPA), in an academic year (annual GPA), or throughout the course of study (cumulative GPA) is calculated using the official grade of each module, weighted by the number of credits assigned to that module.

The grades (on a 10-point scale and a 4-point scale) and letter grades for each course are presented in detail, along with the specific number of credits for each course. The letter grade conventions and grade conversion are detailed in Table 1.3.

Table 1. 3. Letter grade conventions and grade conversion

Ranking	Scale 10	Scale 4
A^+	[9.5;10]	4.0
A	[8.5;9.5)	3.7
B^+	[8.0;8.5)	3.5
B	[7.0;8.0)	3.0
C^+	[6.5;7.0)	2.5
C	[5.5;6.5)	2.0
D^+	[5.0;5.5)	1.5
D	[4.0;4.9)	1.0
F	[0.0;4.0)	0.0

Additionally, admission data were collected through surveys conducted via Google Forms.

The dataset includes 2,763 records of Primary Education students from cohorts D2016 to D2020, with 89 features: 4 admission-related features (scores in the National High School Graduation Examination for Mathematic, Literature, English, and the total score), and 85 subject scores (including elective courses, which may vary by student). Each record corresponds to one student.

The data were cleaned to remove irrelevant variables and attributes outside the scope of this study, such as physical education, arts-based subjects, and student financial variables. Attributes with sparse or missing values, mostly electives, were also eliminated. The analysis focused on numerical exam scores, discarding any textual grading elements. Specific features for each student were selected for correlation analysis with the target variable. As a result, the cleaned dataset includes 932 student records (11 Mediums, 430 Goods, 468 Very Goods and 23 Excellents) and 39 selected attributes, including 4 pre-university academic attributes and 35

university course grade attributes and its label (Excellent, Very Good, Good, Medium).

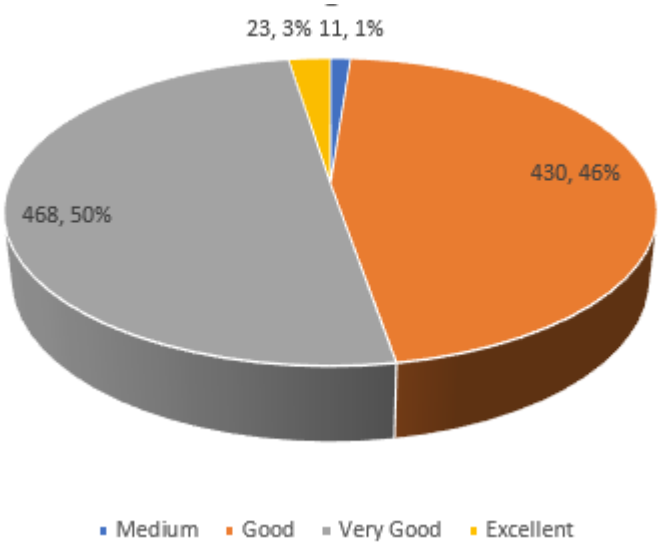


Figure 1. 6. The structure of HNMU1 dataset

The HNMU1 training dataset includes records from 932 students, categorized into 4 graduation classes. It contains 28 variables, including high school graduation exam scores and academic results from the first two years of university. The list of HNMU1 training variables is given in Table 1.4.

Table 1. 4. List of HNMU1 variables

No.	Subject Name	No.	Subject Name	No.	Subject Name	No.	Subject Name
1	Subject 1 (Mathematics)	8	NDSE 1	15	FML 2	22	PVL
2	Subject 2 (Literature)	9	NDSE 2	16	GPA – Semester 2	23	Informatics
3	Subject 3 (English)	10	GPA – Semester 1	17	Research Methodology	24	HCMI
4	Total Score	11	NDSE 3	18	PASM	25	GPA – Semester 4
5	FMT 1	12	Physical Education 1	19	Psychology	26	CGPA (4-point scale)
6	RGCP	13	Physical Education 2	20	Teaching Practicum 1	27	CGPA (10-point scale)
7	Educational Science	14	FML 1	21	GPA – Semester 3	28	

(FMT) Fundamentals of Mathematical Theory; (RGCP) Revolutionary Guidelines of the Communist Party; (NDSE) National Defense and Security Education; (PASM) Public Administration and Sectoral Management; (HCMI) Ho Chi Minh's Ideology; (PVL) Practical Vietnamese Language; (FML) Fundamentals of Marxism-Leninism

Remarks: Most variables have average scores in the mid-to-high range approximately 6.5–7.8). SGPA values are as follows:

- GPA Semester 1: 7.237
- GPA Semester 2: 6.792
- GPA Semester 3: 7.690
- GPA Semester 4: 6.363
- CGPA on a 10-point scale: 7.874

The average scores are generally stable, indicating that most students perform at a good level. Overall, academic performance is consistent and concentrated around the "Good" to "Very Good" range.

Predictive or statistical models should account for left-skewed and low-variance data distributions, especially in subjects where many students achieve near-perfect scores.

Variables with extreme skewness - such as Teaching Practicum 1 -should be treated separately when building predictive models or evaluating academic performance.

1.4.2. HNMU2 dataset

The second dataset, HNMU2, was also collected from Hanoi Metropolitan University in 2023. It comprises 2,613 data records from students in the Mathematics and Physics Education programs [CT3].

To ensure model accuracy, the academic performance prediction task was conducted separately for each major, as different programs follow distinct curricula and graduation requirements. For the HNMU2 dataset, this study selected the Mathematics Education program.

A landmark of this dataset is the inclusion of survey responses from over 2,613 current and former students. Unlike the HNMU1 dataset, which only contains input scores and academic performance during university studies, the HNMU2 dataset was constructed based on extensive surveys and multi-source data collection. It incorporates a diverse set of questionnaire items covering multiple dimensions, including personal characteristics, family background, environmental factors, prior academic achievements, and the influence of the current university environment (e.g., faculty, curriculum, facilities, and related factors).

After completing the data collection and preprocessing steps, the final dataset comprises 551 records of Mathematics Education students, with 88 features, including 36 survey-based attributes and 52 academic performance variables (course grades on a 10-point scale). Survey attributes of the HNMU2 dataset: Personal, environmental, and prior academic performance variables are detailed in Table 1.5.

Table 1. 5. Survey variables of the HNMU2 dataset

Attribute		**Attribute**	
Individuals' information	Gender	Academic Performance and Exam Results Prior to Enrollment	HSGE score for Chemistry
	Parents' educational level		HSGE score for Biology
	Part-time job		High school graduation exam scores
	Funding for tuition fees		Entrance English score
	Study time	Learning Conditions and Support	Methods of admission
	Social media usage time		Ranking choices
	The total number of social media platforms used		Scholarship
	Health condition		Level of adaptation to the environment
Academic Performance and Exam Results Prior to Enrollment	Secondary school graduation exam scores		Learning methods
	High school graduation exam for Mathematics		Level of school support
	HSGE score for Literature		Level of instructor support
	HSGE score for English		Facility conditions
	Groups of subject for admission		Quality of instructors
	HSGE score for History		Suitability of the training program
	HSGE score for Geography		Competitiveness in studies
	HSGE score for Civic Education	Other Factors	Influence of friends
	HSGE score for Physics		Level of interest in the field of study

The HNMU2 dataset includes 62 score-related variables representing students' academic performance over eight university semesters. As presented in Table 1.6, 52 of these variables correspond to individual subject scores, while the

remaining 10 variables represent semester grade point averages (SGPAs) across eight semesters and cumulative grade point averages (CGPAs) on both 4-point and 10-point scales

Table 1. 6. List of HNMU2 score variables

No.	Subject Name	No.	Subject Name	No.	Subject Name	No.	Subject Name
1	Linear Algebra	17	Calculus 3	33	Elementary Algebra	49	TMA
2	Calculus 1	18	GTMM	34	Measure Theory and Integration	50	Research Methodology
3	Analytic Geometry	19	PVL	35	PTMS	51	Elective (HMET)
4	FML 1	20	Elective (Music,Art, ..)	36	Differential Equations	52	Elective (ACTS)
5	Informatics	21	AEG	37	TMMC	53	Elective (DGM)
6	Psychology	22	Arithmetic	38	English for Specific Purposes	54	Semester 7 GPA
7	Semester 1 GPA	23	Teaching Skills 2	39	Functional Analysis	55	Teaching Practicum 3
8	Calculus 2	24	Semester 3 GPA	40	General Law	56	Graduation Thesis
9	Educational Science	25	Ho Chi Minh's Ideology	41	PDE	57	DTA
10	FML 2	26	Complex Functions	42	PASM	58	DTG
11	English	27	Projective Geometry	43	Linear Programming	59	ATSF
12	Elective (Vietnam culture)	28	Number Theory	44	Teaching Practicum 2	60	Semester 8 GPA
13	Teaching Skills 1	29	Teaching Skills 3	45	Semester 6 GPA	61	CGPA (4-point scale)
14	Semester 2 GPA	30	Teaching Practicum 1	46	Numerical Analysis	62	CGPA (10-point scale)
15	RGCPV	31	General Topology	47	Elementary Geometry		
16	General Algebra	32	Semester 4 GPA	48	Semester 5 GPA		

(ATSF) Advanced Topics in Sequences and Functions; (PTMS) Probability Theory and Math. Statistics; (AEG) Affine and Euclidean Geometry; (TMMC) Teaching Methods in Math Content; (PASM) Public Administration & Sector Management;(DTG) Differentiated Teaching – Geometry; (GTMM) General Teaching Methodology for Mathematics; (DTA) Differentiated Teaching – Algebra; (PDE) Partial Differential Equations; (DGM) Differential Geometry, Mechanics; (ACTS) Advanced Calculus, Topo Spaces; (HMET) History of Math., Educational Tools; (TMA) Teaching Mathematics in English; (RGCPV) Revolutionary Guidelines of CPV

Remark: Most variables have average scores ranging between 7.0 and 8.2, indicating that most students performed at a "Good" to "Very Good" academic level. Semester-wise GPA values (10-point scale) are as follows:

- GPA Semester 1: 7.205/GPA Semester 2: 7.320
- GPA Semester 3: 7.195/GPA Semester 4: 7.992
- GPA Semester 5: 7.964/GPA Semester 6: 7.902

- GPA Semester 7: 7.637/GPA Semester 8: 8.599

Overall, the GPA trend shows gradual improvement over time, especially in the final years. This pattern reflects students' better engagement and academic maturity in later stages of their program. The highest average GPA is observed in Semester 8, primarily due to high scores in practicum and thesis-related subjects.

Most subject scores exhibit low to moderate standard deviations (typically around 0.8 - 1.2), suggesting tight clustering of student performance. However, a few subjects such as *"*Measure Theory and Integration*"* (SD = 1.496) and *"Calculus 2"* (SD = 1.412) display more variability.

Distribution characteristics:

- Skewness is moderately negative in most variables (between -0.2 and -0.9), indicating that many students scored toward the higher end of the scale.

- Several subjects show extreme left-skewness, such as: *Teaching Practicum 2*: skewness = -1.444; *Teaching Practicum 3*: skewness = -1.221; *Probability and Statistics*: skewness = -1.614; *Elective: Music/Arts/Islands*: skewness = -2.614. These variables should be treated with caution in predictive models because they lack variance and may bias learning algorithms toward majority scores. They can also distort performance comparisons across students.

- Subjects like "English" (mean = 6.82, skewness = +0.308) exhibit slight positive skew, meaning some students may struggle more compared to other subjects.

Summary:

- Academic performance is consistently strong, with stable GPA across semesters.
- Low variance and left-skewed distributions dominate the dataset.
- Special attention should be given to practicum and thesis components due to their near-perfect scoring patterns.

The HNMU2 dataset suffers from a severe class imbalance in the distribution of student performance categories. Specifically, the Medium class contains only 19 samples, the Good class has 338 samples, the Very Good class has 190 samples, and the Excellent class includes just 4 samples. This imbalance, along with the relatively small overall sample size, poses significant

challenges for training predictive models and achieving high classification accuracy.

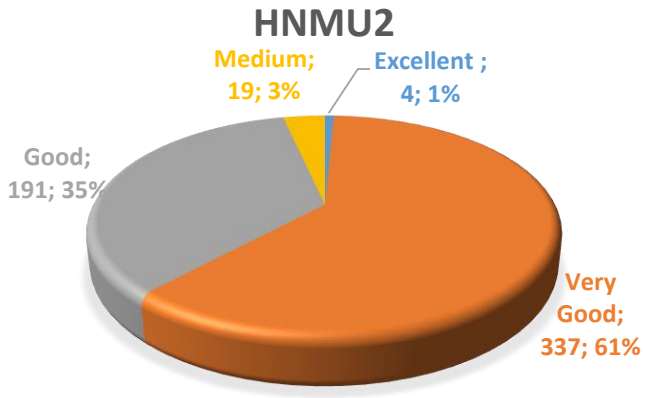


Figure 1. 7. The structure of HNMU2 dataset

1.4.3. VNU dataset

Similar to the HNMU2 dataset, the dissertation selected data from the Literature Education major at Vietnam National University, Hanoi (VNU), for empirical investigation [CT4]. The raw dataset contains 521 samples and 91 attribute fields. These include: 29 features related to individual learner characteristics, 9 features about the learning environment, 10 features on prior academic performance, and 43 features representing students' university-level academic performance. After completing the data collection and preprocessing steps, the final dataset comprises 271 samples, labeled with graduation classifications: 46 "Medium", 187 "Good", and 38 "Excellent" . This distribution is more balanced than that of the HNMU1 and HNMU2 datasets.

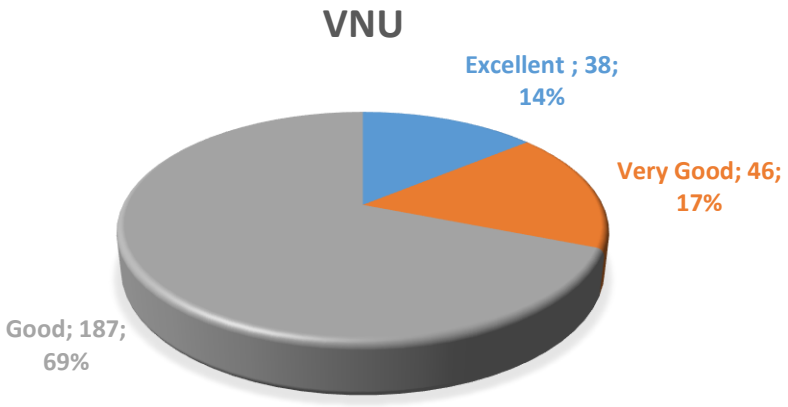


Figure 1. 8. The structure of VNU dataset

This dataset includes the most survey attributes among all datasets considered in this study. Specifically, 48 surveyed attributes cover:

- Personal factors: age, gender, interests, strengths and weaknesses, interpersonal relationships, time spent on social media, part-time work, and study hours;
- Family background: parental age, education level, occupation, family traditions, hometown, and local culture;
- Social factors: social trends, university entrance exam subject combinations, social groups, and community influences affecting students' learning attitudes and performance;
- Educational history and institutional characteristics: academic achievements in lower and upper secondary school, university entrance scores, faculty quality, facilities, and curriculum.

Survey data were collected via Google Forms and matched with official academic records (including 43 performance indicators and graduation classification). Following table shows the list of score variables in this dataset.

Table 1. 7. List of VNU score variables

Semester 1	**Semester 2**	**Semester 3**	**Semester 4**
Vietnamese Cultural Foundations	Vietnamese Grammar	General Psychology and School Psychology	Revolutionary Path of the Communist Party of Vietnam
Introduction to Linguistics	Principles of Literary Theory	Literary Genres	Basic Sino-Nom Studies
Fundamental Principles of Marxism-Leninism I	Fundamental Principles of Marxism-Leninism II	Fundamentals of Informatics	Literary Works
Vietnamese Folk Literature	Statistics for Social Sciences	Ho Chi Minh Thought	The Short Story: Theory and Genre Practice
General Sociology	Vietnamese Literature (10th – mid-18th century)	Chinese Literature	Practical Vietnamese Writing
Introduction to Educational Science	General Vietnamese Linguistics	Vietnamese Literature (late 18th – 19th century)	Russian Literature
Introduction to Applied Statistics in Education	Marxist-Leninist Political Economy	Vietnamese Stylistics	Vietnamese Literature (1900–1945)
Marxist-Leninist Philosophy	Didactic Theory	Sino-Vietnamese Classical Texts	General Pedagogy
GPA Semester 1	Introduction to Educational Technology	ICT Applications in Education	Applied Linguistics

	Educational Psychology	Scientific Socialism	Organization of Educational Activities in Schools
	English B1	History of the Communist Party of Vietnam	Introduction to Educational Measurement and Evaluation
	GPA Semester 2	GPA Semester 3	Organizing Experiential Activities in Teaching Literature
			GPA Semester 4

Remark:

- Mean scores are generally high and stable, mostly in the 7.5 - 8.2 range.
- SGPAs reflect consistent academic achievement:
    - GPA Semester 1: 7.39
    - GPA Semester 2: 7.68
    - GPA Semester 3: 7.75
    - GPA Semester 4: 7.78

This indicates strong academic performance across the student population, suggesting that most students fall in the Good to Very Good category.

- Standard deviations are mostly low to moderate (typically 0.7 – 1.2), showing tight clustering of scores.
- Distribution Characteristics: Skewness is predominantly negative. This suggests a large proportion of students scored near the upper end (8.0–10.0).
- Outliers and Special Cases: Some subjects show extreme skew and low variance, which may reflect highly uniform grading practices or cause bias in modeling or prediction tasks, such as V*ietnamese Folk Literature*: Mean = 8.35, Skewness = -3.125; *Vietnamese Literature (1900–1945)*: Mean = 7.64, Skewness = -3.654.

Predictive models trained on this dataset should account for non-normal distributions, apply appropriate data transformations, and consider excluding or reweighting variables affected by near-ceiling effects in comparative analyses.

In summary, three original datasets HNMU1, HNMU2, and VNU are severely imbalanced, with very few samples in the Medium and Excellent classes (for example, HNMU2 has only 4 Excellent samples, and VNU has none in the Medium class), while the Good and Very Good classes dominate. This imbalance can easily cause the predicted model to bias toward the majority classes and overlook the minority ones.

1.4.4. International datasets

In this dissertation, six international datasets were employed, collected from diverse universities and educational institutions worldwide. These datasets include those from Covenant University in Nigeria ([61]), the University of Malaya in Malaysia ([62]), the SATDAP Program-Capacitação da Administração Pública in Portugal ([63]), and the well-known Portuguese school performance dataset ([64]).

Detailed descriptions and characteristics of these datasets are presented in Table 1.8.

Table 1. 8. Dataset description

Dataset	Name	Institutions	N	k	Web link
1	SATDAP	SATDAP program, Potugal	4424	36	UCI dataset
2	Malaya-Stud	Universiti Malaya, Malaysia	493	16	Mendeley data
3	Portugal-Math	Portuguese schools, Portugal	395	33	UCI dataset
4	Portugal-Lang	Portuguese schools, Portugal	649	33	UCI dataset
5	Covenant-Priv	Covenant University, Nigeria	1841	9	Data in Brief

Data's name Institutions, Sample size (n), the number of features (k), and web-link to data sources.

1.4.5. Issues of privacy and sensitive data handling

- *Privacy & Ethics:* Educational data include sensitive personal information such as grades, learning behaviors, psychological surveys, and family or social factors, requiring strict compliance with legal and ethical standards.

- *Program diversity:* Differences in curricula, assessment methods, credit systems, and frequent program updates make it difficult to standardize data across disciplines.

- *Institutional disparity:* Variations in university scale, data digitization levels, and training policies lead to fragmented and non-uniform datasets.

- *Personalized learning paths:* Students' flexible course selections and pacing create inconsistent time-series data, posing challenges for deep learning models that rely on continuous learning trajectories.

1.5. Evaluation metrics for predictive models

1.5.1. Some metrics for classification models

The evaluation metrics used include: Accuracy (Acc), Precision (P), Recall (R), and F1-Score (F1), calculated using the following formulas:

$$\text{Acc} = \frac{Correct\ predictions}{All\ predictions}; \quad (1.1)$$

$$(Macro - Averaged\ Precision)\ P = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i+FP_i}; \quad (1.2)$$

$$(Macro - Averaged\ Recall)\ R = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i+FN_i}; \quad (1.3)$$

$$(Macro - F1\ score)\ F1 = \frac{2*P*R}{P+R}; \quad (1.4)$$

where N is the number of classes, TP_i (True Positive of the class i), FP_i (False Positive of the class i), and FN_i (False Negative of the class i) are key metrics in classification tasks and all predictions are the total number of data samples.

The greater the values of Accuracy, Precision, and Recall, the better the model performance.

1.5.2. Some metrics for regression models

To evaluate the accuracy of a regression model, the dissertation uses evaluation metrics such as: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and R-square (R^2).

Mean Square Error (MSE)

MSE measures the average of the squared differences between predicted values and actual values. It shows how far the predicted values are from the actual values on average, with larger errors being penalized more due to squaring.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2, \quad (1.5)$$

where: y_i is the actual value of the dependent variable, y_i is the predicted value and n is the sample size.

Root Mean Square Error (RMSE)

RMSE measures the average deviation between predicted and actual values, but keeps the same units as the original data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i\right)^2} \quad (1.6)$$

where y_i is the actual value of the dependent variable, y_i is the predicted value and n is the sample size.

RMSE has the advantage of being measured in the same units as the dependent variable, making it easy to compare between models and across different dependent variables. It also provides the average deviation between predicted and actual values, helping to assess the model's predictive ability. However, RMSE can be affected by noise or outlier values in the data. If the data contains noise or outliers, RMSE can be significantly reduced.

Mean Absolute Error (MAE)

MAE measures how much predictions deviate from actual values on average.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_i - y_i |, \quad (1.7)$$

where y_i is the actual value of the dependent variable, y_i is the predicted value and n is the sample size.

MAE also measures the average error of the model compared to the actual data; however, MAE calculates the average of the absolute values of the errors. The advantage of MAE is that it has the same units as the dependent variable, making it easy to compare between models and across different dependent variables. However, MAE does not assess the magnitude of the errors.

R-square ($\boldsymbol{R^2}$)

The R-square metric measures the extent to which the model explains the dependent variable. R-square is calculated using the formula:

$$R^2 = 1 - (SSE / SST) = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (1.8)$$

where y_i is the actual value of the dependent variable, y_i is the predicted value and n is the sample size; SSE (Sum of Squared Errors): The sum of squared errors of the model; SST (Total Sum of Squares): The total sum of squares of the sample mean, $\bar{y}$ is the mean of the actual values.

R^2 measures the extent to which the model explains the dependent variable. R^2 has the advantage of being simple, easy to understand, and easy to use. It indicates the percentage of variance in the dependent variable explained by the model. R^2 also helps compare the explanatory power of different models, allowing for the selection of the best model. However, R^2 does not provide any information about the model's error. Additionally, R^2 can be affected by adding independent variables to the model, leading to a higher R^2 value while the model still has significant errors.

In summary, depending on the characteristics of the problem and the prediction goals, selecting the appropriate metric helps accurately evaluate the model's effectiveness and practicality.

Table 1. 9. Selection of evaluation metrics for the model

Type of Problem	Priority Metrics	Reason for Selection
Multiclass Classification (Graduation Classification)	F1, Accuracy	Ensure no bias towards the majority class
Regression (Predicting GPA)	MAE, RMSE, R^2	Evaluate both absolute error and model explanatory power

The conclusion of Chapter 1

Chapter 1 identifies two core problems addressed in this dissertation: short-term regression-based GPA prediction and long-term classification-based graduation classification prediction, reflecting distinct aspects of the learning process and academic achievement.

By synthesizing existing machine learning and deep learning models and highlighting research gaps, this chapter sets the direction for subsequent chapters to deeply explore advanced deep learning architectures and develop hybrid models aimed at optimizing predictive performance within the unique educational context characterized by limited, heterogeneous, and uncertain

data. Additionally, the dissertation emphasizes integrating diverse influencing factors, ranging from individual traits and learning environments to social impacts, to enhance the accuracy and comprehensiveness of predictions for the two main tasks: SGPA and graduation classification.

CHAPTER 2. EARLY PREDICTION OF SEMESTER GRADE POINT AVERAGE USING DEEP LEARNING APPROACHES

In modern education, predicting students' semester Grade Point Average (SGPA) is important for tracking learning outcomes, identifying students at risk, and guiding personalized study plans. However, SGPA is not an exact or stable measure. It can change over time under the influence of many factors, such as grading methods, teaching approaches, students' mental conditions, and differences between institutions. Therefore, SGPA should be considered a flexible indicator that reflects both uncertainty and variability. From this view, this chapter presents predictive models that apply deep learning together with uncertainty-based methods to improve accuracy and better represent the complexity of real educational environments.

Two modeling approaches are proposed:

***NeutroDLs**: Embeds neutrosophic logic into standard deep learning models.*

***NeutroGNT**: A hybrid model combining Transformer, CGAN, and neutrosophic representation to handle data imbalance and uncertainty.*

Experiments on seven real datasets show that the models significantly improve prediction accuracy, with NeutroGNT achieving MSE = 0.018 and R^2 = 96.05%.

The content of this chapter is based on the publications [CT5] and [CT6].

2.1. Problem formulation

In this chapter, we consider the semester GPA prediction problem. Higher education institutions commonly structure their academic programs over a period ranging from a minimum of eight semesters to a maximum of twelve, corresponding to an overall duration of approximately three to six years. The GPA is a standardized metric widely employed in universities to assess students' academic performance. It is computed based on individual course grades and consolidated into an overall average, typically measured on a 4.0 scale or a 10.0 scale.

The semester GPA (SGPA) is calculated at the end of each academic semester and serves as an indicator of a student's ongoing academic performance. The SGPA for each semester is typically computed using the following formula:

$$x_{TB} = \frac{x_1 k_1 + \cdots + x_n k_n}{k_1 + \cdots + k_n}, \quad (2.1)$$

where n denotes the number of courses taken in a given semester; x_i and k_i represent the grade and the number of credit hours for the i-th course, respectively, $i \in \{1, ..., n\}$; and x_{TB} denotes the SGPA.

The SGPA prediction problem can be formulated as follows: For a given student, assuming that the GPAs for the first m semesters are known, the task is to predict the GPA of the $(m + 1)$-th semester.

Let X be nonempty set. $X = X_1 \otimes X_2 \otimes ... \otimes X_m$, $x = (x_1, x_2, ..., x_m) \in X$.

In this chapter, the dissertation investigates the following three scenarios:

Case 1: Predict the SGPA of the nth semester if the SGPA of the $n-1$ semester is given. That is, knowing the value of x_{n-1}, predict the value of $x_n, 1 < n \leq m$, see Figure 2.1.

Figure 2. 1. Prediction framework for Case 1

Case 2: Predict the student's nth term SGPA when the SGPA of the previous 2 semesters are given. That is, knowing the values of x_{n-2}, x_{n-1}, predict the value of $x_n, 2 < n \leq m$, see Figure 2.2.

Figure 2. 2. Prediction framework for Case 2

Case 3: Predict the student's nth SGPA when knowing the SGPA of the previous 3 semesters. That is, knowing the value of x_{n-3}, x_{n-2}, x_{n-1}, predict the value of $x_n, 3 < n \leq m$, see Figure 2.3.

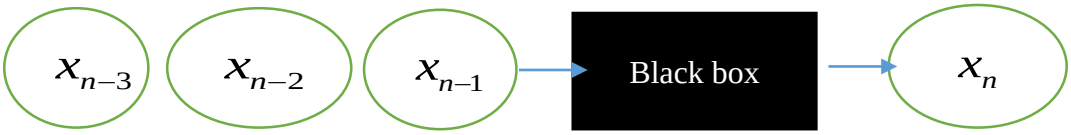


Figure 2. 3. Prediction framework for Case 3

2.2. NeutroDL models

2.2.1. The theoretical basis for model selection

In the educational environment, the assessment of students' academic performance is inherently complex, uncertain, and variable. Several key factors contribute to the uncertainty and indeterminacy of GPA results, as outlined below:

i) Multi-component mordern assessment structure

Student grades are typically derived from various components, including:

- Attendance/Class participation,
- Midterm and final examinations/Assignments, projects, presentations,
- In-class question-and-answer sessions.

Each of these components is influenced by: The context and objectives of the educational system/The instructor's level of enthusiasm and teaching style/Various unstructured or subjective elements *([65]; [66]; [67]).*

ii) Shifts in instructional and assessment methods due to online education

The rapid expansion of online education, especially following the COVID-19 pandemic, has introduced major changes:

- Use of online interaction metrics (e.g., click rates, login frequency, engagement time),
- Assessment based on task completion rather than solely on traditional examinations,
- Introduction of new grading scales and conversion methods, such as: Converting letter grades to numerical scores/GPA calculations based on classifications like excellent, very good, good, average, poor, and very poor.

These developments lead to: Inaccuracies due to conversion rules/Subjectivity in online assessment/Inconsistencies across educational systems.

iii) Personal and psychological factors affecting students

Student SGPA is also affected by non-quantifiable factors such as:

- Individual learning strategies,
- Perceived difficulty of courses or exams,
- Psychological conditions (stress, motivation, confidence, etc.),
- Teaching methods used by instructors.

Given the above, the evaluation of student academic performance - especially through SGPA -should not be treated as a precise or static indicator. Instead, SGPA must be viewed as a value characterized by uncertainty and imprecision, requiring:

- The use of more flexible and robust prediction models,
- Integration of multiple data sources and “soft” factors,
- The application of machine learning or analytical methods capable of handling uncertainty (e.g., fuzzy models, neutrosophic models, etc.).

In efforts to minimize uncertainty in the evaluation process, many studies have attempted to apply fuzzy set theory in educational assessment ([67]; [68]; [69]; [70]; [71]).

As a result, the application of uncertainty theories (such as fuzzy theory and neutrosophic theory) into machine learning and deep learning models has become a crucial research direction, improving the accuracy of predicting students' academic performance ([72]; [73]).

Therefore, this dissertation aims to develop predictive models for SGPA based on educational datasets that are often incomplete, contain uncertainty, and are influenced by various subjective factors.

To effectively address these challenges, the study proposes an integrated modeling framework that combines neutrosophic theory with deep learning techniques. This approach not only improves the accuracy of SGPA prediction but also enables the model to represent and quantify the uncertainty present in the input data.

By explicitly modeling indeterminacy, imprecision, and ambiguity, the proposed framework offers a more flexible and practical evaluation method, well-suited to the complex nature of real-world educational environments.

2.2.2. Proposed model

In this section, the dissertation introduces the integration of neutrosophic theory with several deep learning models to predict the final semester grade and the overall course grade of university students. The overall model is presented in Figure 2.4.

` While several similar studies have been documented in the existing literature, our proposed model distinguishes itself by incorporating the time factor, semester variations, and the use of neutrosophic functions for input data fuzzification.

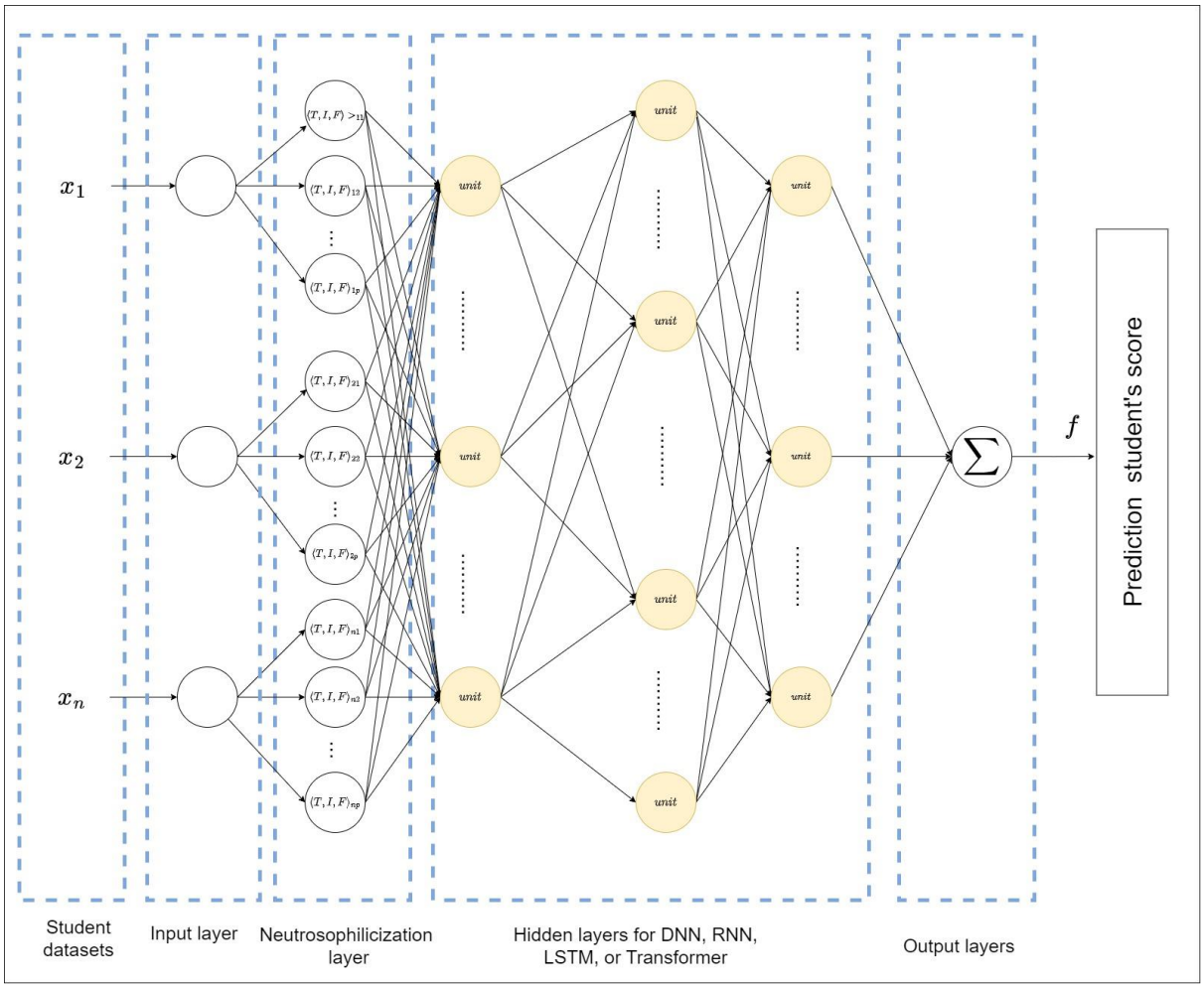


Figure 2. 4. The NeutroDL models [CT6]

Figure 2.4 illustrates the general architecture of the neutrosophic neural networks (DNN, CNN, RNN, LSTM, and Transformer). The proposed model is a combination of neutrosophic theory and several popular neural networks aimed at improving the predicted academic performance of students. Within the scope of this study, which evaluates the effectiveness of integrating neutrosophic theory and neural networks to address the problem of predicting students' scores, the dissertation employs five commonly used neural networks: DNN, CNN, RNN, LSTM, and Transformer. For ease of comparison, the dissertation utilizes the Adam optimization algorithm in all models. The process of applying modern deep learning techniques to predict students' academic performance is carried out as follows:

Step 1: Model Construction

The structure of the model includes the following main layers:

Input Layer: The data processing layer is responsible for preparing raw

data for use in the neural network. This includes tasks such as data cleaning, converting data into a format that the neural network can understand, and organizing the data in a continuous timeline format. The output of this layer is then passed on to the next encoding layer.

Encoding Layer: This layer transforms the data using neutrosophic theory. From the output data of the input layer, the data is neutrosophicized using corresponding neutrosophic membership functions to represent uncertainty, indeterminacy, and inconsistency in the datasets. Trapezoidal neutrosophic membership functions used in encoding neutrosophic data (details on some neutrosophic functions can be found in Subsection 2.4.1).

Hidden Layer: The objective of this layer is to evaluate and consider the uncertainty, indeterminacy, and inconsistency factors in deep learning models. In this dissertation, the hidden layer is examined for traditional neural network architectures such as DNN, CNN, RNN, LSTM, and Transformer.

With the structure of the available methods, determining appropriate input parameters and preprocessing data (data cleaning, sequence design, continuous timeline sorting, and neutrosophicization of input data) while highlighting the comparability of the predictive methods are crucial factors to consider when constructing the model in this dissertation.

The dissertation begins with data analysis using the DNN model, which serves as a baseline for comparing other deep learning approaches. It then introduces the RNN model, incorporating the temporal nature of academic data by analyzing the sequential performance of students across semesters. Data preprocessing involves cleaning, organizing, and structuring the raw data into time-series format based on actual semester records.

For neutrosophic-based models, neutrosophic functions are applied to convert the input into neutrosophic sets. The key challenge lies in adapting the input data format for each model and selecting appropriate hyperparameters to ensure effective training and accurate predictions. The LSTM and Transformer models are also employed to explore prediction capabilities using the same dataset. Sigmoid function are applied, and training is conducted using parameters such as batch size and epochs.

Decoder and Output Layer: The output of the hidden layers consists of neutrosophic values, and the objective of the decoder layer is to apply

neutrosophic defuzzification to generate the corresponding real values of neutrosophic membership. The final layer is the prediction layer, which provides the final prediction based on the input features received from the previous hidden layers.

Table 2. 1. Layer structure of DNN model

Model: "sequential"

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 18)	0
dropout (Dropout)	(None, 18)	0
dense (Dense)	(None, 128)	2432
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Table 2. 2. Layer structure of CNN model

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 1, 64)	1216
max_pooling1d (MaxPooling1D)	(None, 1, 64)	0
conv1d_1 (Conv1D)	(None, 1, 64)	4160
max_pooling1d_1 (MaxPooling1D)	(None, 1, 64)	0
flatten (Flatten)	(None, 64)	0
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 128)	8320
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Table 2. 3. Layer structure of CNN model

Model: "sequential"

Layer (type)	Output Shape	Param #
simple_rnn (SimpleRNN)	(None, 1, 312)	103272
simple_rnn_1 (SimpleRNN)	(None, 1, 56)	20664
simple_rnn_2 (SimpleRNN)	(None, 1, 88)	12760

simple_rnn_3 (SimpleRNN)	(None, 1, 504)	298872
simple_rnn_4 (SimpleRNN)	(None, 264)	203016
dropout (Dropout)	(None, 264)	0
dense (Dense)	(None, 1)	265

Table 2. 4. Layer structure of LSTM model

Model: "sequential"

Layer (type)	**Output Shape**	**Param #**
lstm (LSTM)	(None, 1, 128)	75264
lstm_1 (LSTM)	(None, 64)	49408
flatten (Flatten)	(None, 64)	0
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 128)	8320
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Table 2. 5. Layer structure of Transformer model

Model: "sequential"

Layer (type)	**Output Shape**	**Param #**
input_1 (InputLayer)	[(None, 1, 18)]	0
layer_normalization (LayerNormalization)	(LayerNorma (None, 1, 18)	36
multi_head_attention (MultiHeadAttention)	(None, 1, 18)	0
dropout (Dropout)	(None, 1, 504)	298872
layer_normalization_1 (LayerNormalization)	(None, 1, 18)	36
conv1d (Conv1D)	(None, 1, 4)	76
dropout_1 (Dropout)	(None, 1, 4)	0
conv1d_1 (Conv1D)	(None, 1, 18)	90
layer_normalization_2 (LayerNormalization)	(None, 1, 18)	36
multi_head_attention_1 (MultiHeadAttention)	(None, 1, 18)	76818
dropout_2(Dropout)	(None, 1, 18)	0
layer_normalization_3 (LayerNormalization)	(None, 1, 18)	36
conv1d_2 (Conv1D)	(None, 1, 4)	76
dropout_3 (Dropout)	(None, 1, 4)	0

conv1d_3 (Conv1D)	(None, 1, 18)	90
layer_normalization_4 (LayerNormalization)	(LayerNor (None, 1, 18)	36
multi_head_attention_2 (MultiHeadAttention)	(None, 1, 18)	76818
dropout_4 (Dropout)	(None, 1, 18)	0
layer_normalization_5 (LayerNormalization)	(None, 1, 18)	36
conv1d_4 (Conv1D)	(None, 1, 4)	76
global_average_pooling1d (GlobalAveragePooling1D)	(None, 1)	0
dense (Dense)	(None, 128)	256
dropout_8 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Step 2: Model Training

The principle of this step is to compute the weights and address the optimization problem. In this process, parameters (weights $\boldsymbol{w}$ and biases - the deviations of each node) are learned by the machine to suggest the optimal results. The backpropagation problem uses the Adam optimization algorithm and employs MAE (Mean Absolute Error) during the model training process. The model parameters are detailed in Table 2.6**.**

Table 2. 6. **Model parameters**

Hyper-parameters	Selection
Learning rate α	0.0003
Drop-out rate	0.3
Number of epochs	1000
Loss function	Mean Absolution Error
Optimizer	Adam

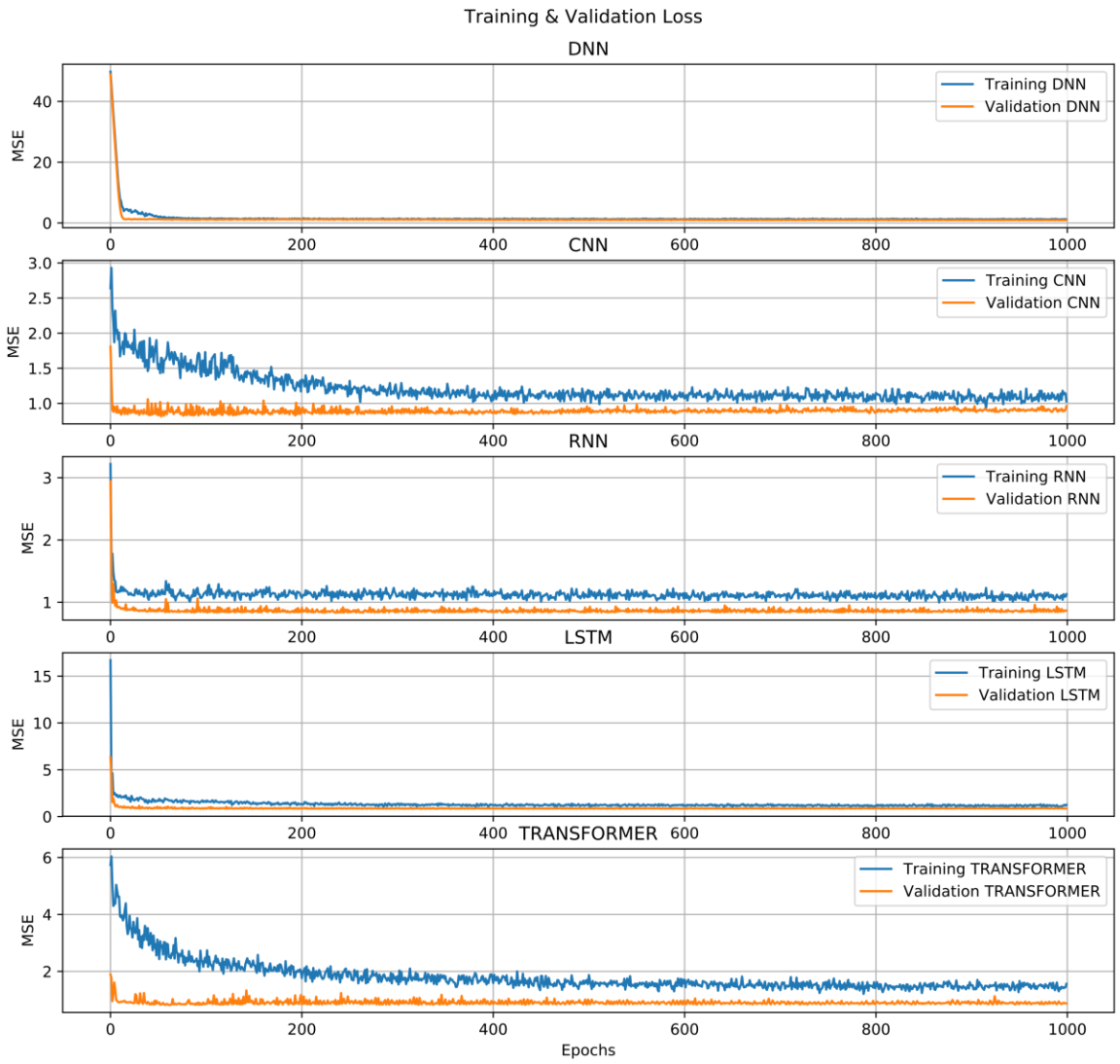


Figure 2. 5. Loss function value chart for training and validation of models

Figure 2.5 shows the loss function value chart for the training and validation phases of the DNN, CNN, RNN, LSTM, and Transformer models with neutrosophic values. From Figure 2.5, it can be observed that the loss function values for both training and validation are converging towards zero. In the proposed framework, the dissertation uses a layer to convert real data into neutrosophic numbers by applying neutrosophic conversion functions, and the model training process uses these neutrosophic numbers. Before making the final predictions, a defuzzification layer is used (to convert neutrosophic numbers back to real values) to provide the corresponding output.

Step 3. Prediction, Testing

After obtaining the model with the computed parameters from Step 2, the dissertation inputs the test data and measures the error. The errors considered here include RMSE, MAE, and R^2.

In summary, the process of applying modern deep learning techniques to predict student's academic performance can be outlined as follows: First, the dissertation utilizes existing libraries of DNN, CNN, RNN, LSTM, and Transformer, incorporates reasonable parameters in preprocessing, and specifically applies neutrosophic sets in handling input data. It then connects (reads) the input data from an Excel file that has been appropriately processed for the application model. Next, the model setup step is performed, followed by training the model to extract parameters (weights, biases). Once the model parameters have been learned, the test data is inputted, predictions are made, and errors are measured.

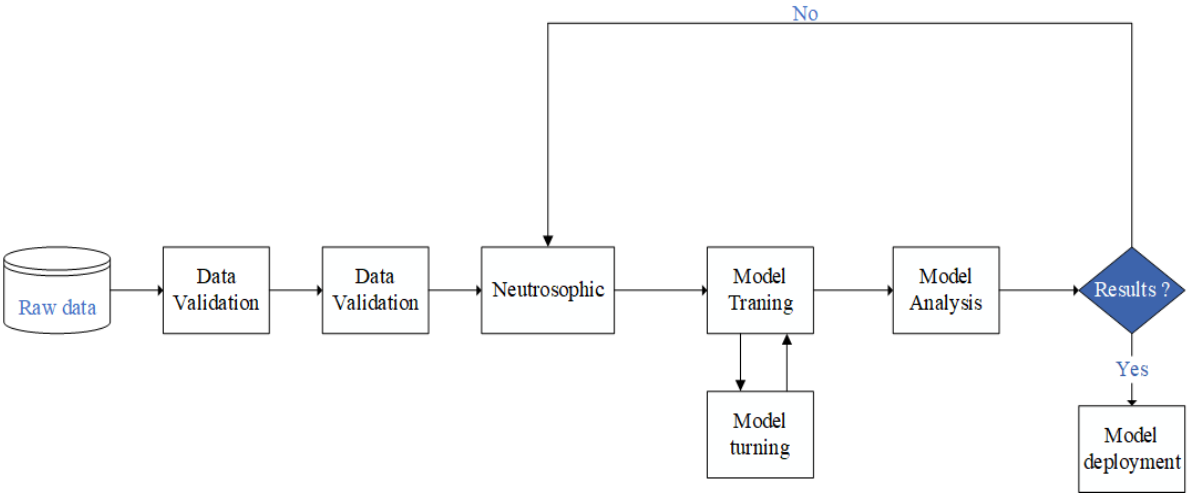


Figure 2. 6. The pineline of NeutroDL model

The integration of the neutrosophic function to neutrosophicize the input and generate the real output creates a novel approach for the models. Results from previous terms (or previous semesters) are inputted to predict the SGPA for the upcoming term. The comparison of prediction ability and accuracy is specifically demonstrated in Subsection 2.2.3, with experiments conducted on the test dataset from Hanoi Metropolitan University, Hanoi, Vietnam.

Details about the layers of each network are provided in the experimental section. Experimental results show that the proposed hybrid model yields better results than traditional models in predicting student's SGPA.

The computational complexity of Algorithm 1 is primarily influenced by three core components: the neutrosophic encoding stage, the model construction and training phase, and the evaluation step.

Algorithm 2.1. Neutrosophic Deep Learning for Student Performance Prediction

Input: X are historical student records; H is prediction horizon;
    F_n: Neutrosophic membership functions;
    Model ∈ {DNN, CNN, RNN, LSTM, Transformer};
    Hyperparameters: learning rate η, dropout rate d, epochs E
Output: $\hat{y}$ Predicted student performance score
Preprocess the raw student data: clean, normalize, order by time
For each input $x_i \in X$ **do**
    Encode x_i using neutrosophic trapezoidal function:
$$[\mathrm{T}(x_i), \mathrm{I}(x_i), \mathrm{F}(x_i)] \leftarrow F_n()$$
end for
Construct model with:
    Input layer (neutrosophic vector $[T, I, F]$)
    Encoder (neutrosophic transformation)
    Hidden layers based on selected model (model ∈ {DNN, CNN, RNN, LSTM, Transformer})
    Decoder (neutrosophic defuzzification)
    Output layer (regression head)
Train the model using Adam optimizer with MAE loss
Run training for E epochs on training data
Evaluate model on test data using RMSE, MAE, R^2
Return $\hat{y}$

The computational complexity of the proposed model depends on the complexity of the underlying architectures such as DNN, CNN, RNN, LSTM, or Transformer. For instance, the computational cost of LSTM and Transformer models is $O(n \cdot d^2)$, where n denotes the length of the input learning sequence and d represents the hidden vector dimension.

2.2.3. Experiment

2.2.3.1. Training dataset

The dataset used in this section is HNMU1 dataset. The structure of the training dataset and input of the models can ben seen in Table 2.7.

Table 2. 7. Description of the training dataset

Dataset	Number of samples	Cases	Real Input attributes	Neutrosophic Input attributes
HNMU1	932	Case 1	1	18
		Case 2	2	36
		Case 3	3	54

Six neutrosophic sets for neutrosophic inputs are Excellent, Very Good, Good, Medium, Poor, and Very Poor. We use the trapezoidal neutrosophic functions:

$$\text{Very Poor} = [0, 0.2, 3.7, 4.1; 1, 0, 0]; \ \text{Poor} = [3.8, 4.2, 4.7, 5.1; 1, 0, 0];$$

$$\text{Medium} = [4.8, 5.2, 5.7, 7.1; 1, 0, 0]; \text{Good} = [6.8, 7.2, 7.7, 8.1; 1, 0, 0];$$

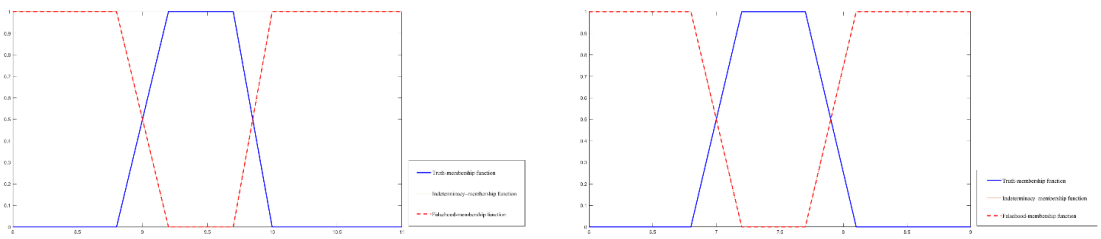
$$\text{Very Good} = [7.8, 8.2, 8.7, 9.1; 1, 0, 0]; \ \text{Excellent} = [8.8, 9.2, 9.7, 10.0; 1, 0, 0].$$

$$T_{Very\,Poor}(x) = \begin{cases} 0 & if \quad x \le 0, x > 4.1 \\ 5x & if \quad 0 < x \le 0.2 \\ 1 & if \quad 0.2 < x \le 3.7 \\ 10.25 - 2.5x & if \quad 3.7 < x \le 4.1 \end{cases}$$

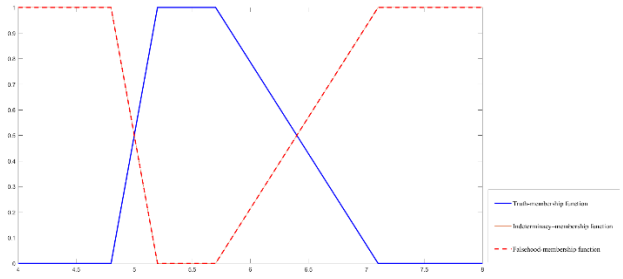
$$I_{Very\,Poor}(x) = \begin{cases} 1 & if \quad x \le 0, x > 4.1 \\ 1 - 5x & if \quad 0 < x \le 0.2 \\ 0 & if \quad 0.2 < x \le 3.7 \\ 2.5x - 9.25 & if \quad 3.7 < x \le 4.1 \end{cases}$$

and

$$F_{Very\,Poor}(x) = \begin{cases} 1 & if \quad x \le 0, x > 4.1 \\ 1 - 5x & if \quad 0 < x \le 0.2 \\ 0 & if \quad 0.2 < x \le 3.7 \\ 2.5x - 9.25 & if \quad 3.7 < x \le 4.1 \end{cases} \qquad (2.3).$$



a) b)



c)

Figure 2. 7. The neutrosophic functions for the concepts a) Good, b) Very Good and c) Excellent

2.2.3.2. Experimental implementation

The student academic performance dataset obtained from Hanoi Metropolitan University is not linearly separable, making certain machine learning methods such as linear regression and Perceptron Learning Algorithm inapplicable. Therefore, this dissertation proposes the use of several modern deep learning models, which are particularly well-suited for handling time series data, as processed in this study. The deep learning methods employed for data analysis include classical DNN, CNN, RNN, LSTM, and Transformer models.

In this section, the CNN model applied consists of one convolutional layer followed by sequential layers, three max-pooling layers, and fully connected layers. The model also utilizes the ReLU activation function and dimensionality reduction techniques to generate predictions of student scores.

The underlying idea of RNN is to process sequential data. RNN are so named because, for each element in a sequence, the output is calculated based on previous computations. In theory, RNN, LSTM and Transformer can process long sequences, but in practice, they are limited to looking back only a few steps. This allows the model to take into account the consistency of the scores achieved by students. In this case, the dissertation uses a sequence of student scores as the input to the neural network. Since the network performs the same task for each element in the sequence, it processes the entire student record to output the most accurate predicted score. Theoretically, the network retains a memory of the student's SGPA.

All experiments were implemented in Python 3.11 within a Conda environment, using common libraries such as NumPy, SciPy, Pandas, Scikit-learn, PyTorch, and TensorFlow/Keras. Classical ML models (e.g., LR, SVM, KNN, RF) were built with Scikit-learn, while deep learning architectures (LSTM, Transformer) were trained with PyTorch and TensorFlow. GPU acceleration was enabled via CUDA Toolkit 11.8 to optimize training efficiency. The experiments were conducted on a workstation equipped with an Intel Core i7-12700KF CPU, NVIDIA RTX 3060 GPU, and 32GB RAM.

Evaluative Metrics: The Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), and R^2-score to estimate the performance and the difference between the student's actual SGPA and predicted SGPA, allowing for appropriate model selection based on the situation.

2.2.4. Results and discussions

Table 2.8 shows results of neutrosophy deep learning models. The average errors are evaluated 10 times on the 10 folds corresponding to Cases 1, 2 and 3. When combining neutrosophic functions to fuzzify input data, calibrating parameters of applied deep learning algorithms, with the training data sample accounting for 80% and 20% of the data used for testing, we get the predicted results as shown in Figures 2.8, 2.9 and 2.10 (for 01 test). The test is devided in to different cases: (Case 1) Using the previous semester's average data to predict the next semester's average score; (Case 2) Using the average data of the 2 previous semesters to predict the average score of the third semester; (Case 3) Using the average data of the 3 previous semesters to predict the average score of the fourth semester.

Table 2. 8. Average error for cases 1, 2, 3 with Neutrosophic approach

Model /Metric	RMSE			MAE			R^2(%)		
	Case 1	*Case 2*	*Case 3*	*Case 1*	*Case 2*	*Case 3*	*Case 1*	*Case 2*	*Case 3*
DNN	0.89 ± 0.09	0.57 ± 0.05	0.87 ± 0.05	0.75± 0.06	0.47 ± 0.05	0.74 ± 0.04	12.76 ± 4.90	48.42 ± 5.74	59.45 ± 4.00
CNN	0.90± 0.07	0.58 ± 0.04	0.80 ± 0.08	0.74± 0.04	0.46 ± 0.02	0.62 ± 0.07	11.40 ± 5.06	47.03 ± 5.80	62.04 ± 5.36
RNN	0.92 ± 0.07	0.60 ± 0.05	0.80 ± 0.08	0.73 ± 0.04	0.45 ± 0.03	0.60± 0.06	12.39 ± 6.06	46.16 ± 6.82	62.14 ± 7.67
LSTM	0.91 ± 0.07	0.57 ± 0.04	0.76 ± 0.11	0.74 ± 0.04	0.45 ± 0.03	0.59 ± 0.07	12.73 ± 4.97	49.51 ± 5.40	65.28 ± 8.93
Transformer	**0.89 ± 0.08**	**0.59 ± 0.06**	**0.79 ± 0.06**	**0.74 ± 0.04**	**0.47 ± 0.06**	**0.59 ± 0.05**	**13.13 ± 7.65**	**45.54 ± 8.92**	**65.95 ± 4.33**

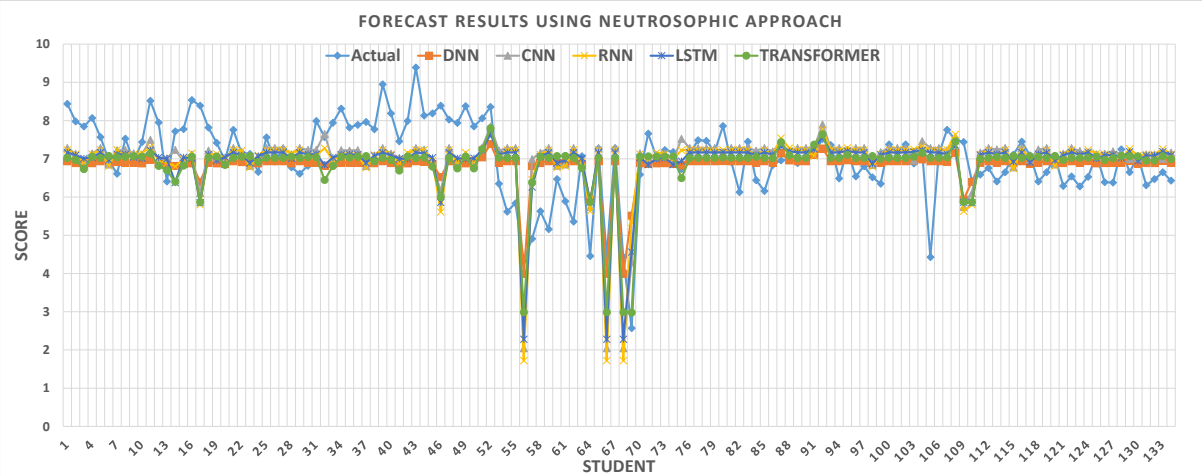


Figure 2. 8. Graph of prediction (neutrosophic data, Case 1)

For Case 1: Using the previous semester's average data to predict the next semester's average score, we estimate the errors (average after 10 tests) of the algorithms as shown in Table 2.9. In this case, the R^2 scores of all five methods are notably low (just slightly above 10%), indicating that this dataset is not suitable for real-world forecasting applications.

The poor performance can be attributed to the limited input data, as only one feature (a single semester score) was used. Although neutrosophic transformation expands this into 18 features - representing 18 values of neutrosophic membership functions - they still lack sufficient diversity to serve as effective inputs for deep learning models.

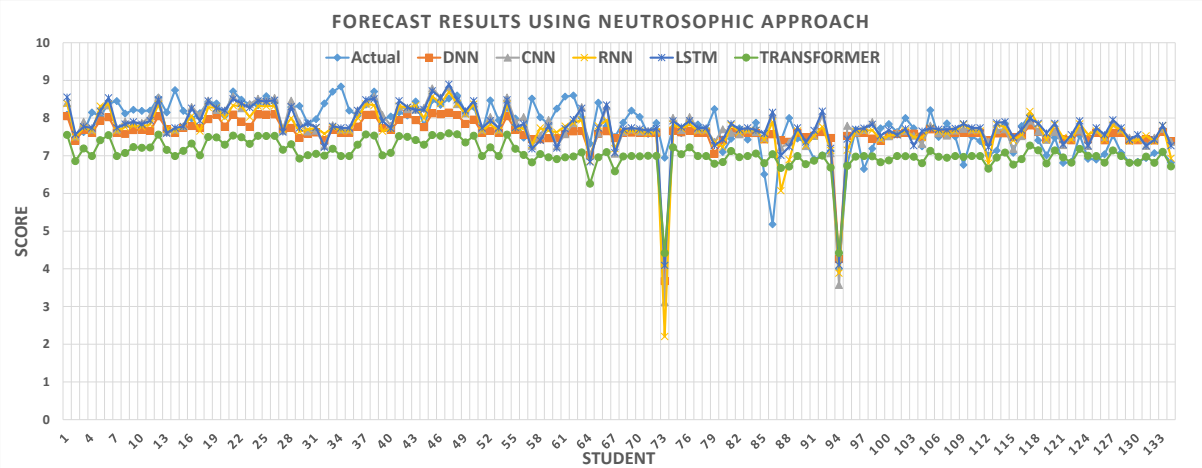


Figure 2. 9. Graph of prediction (neutrosophic data, Case 2)

For Case 2: Using the average data of the 2 previous semesters to predict the average score of the third semester, we get the predicted results as shown in Figure 2.10 (for 01 test) and the errors (average after 10 tests) of the algorithms in Table 2.9. In this Case, we can see that the quality of forecasting methods increases significantly compared to Case 1. In this case the R^2 metrics of all five method are approximately 50% for neutrosophic cases, so in cases where much information is not collected, we can also use this model for the problem of predicting student scores.

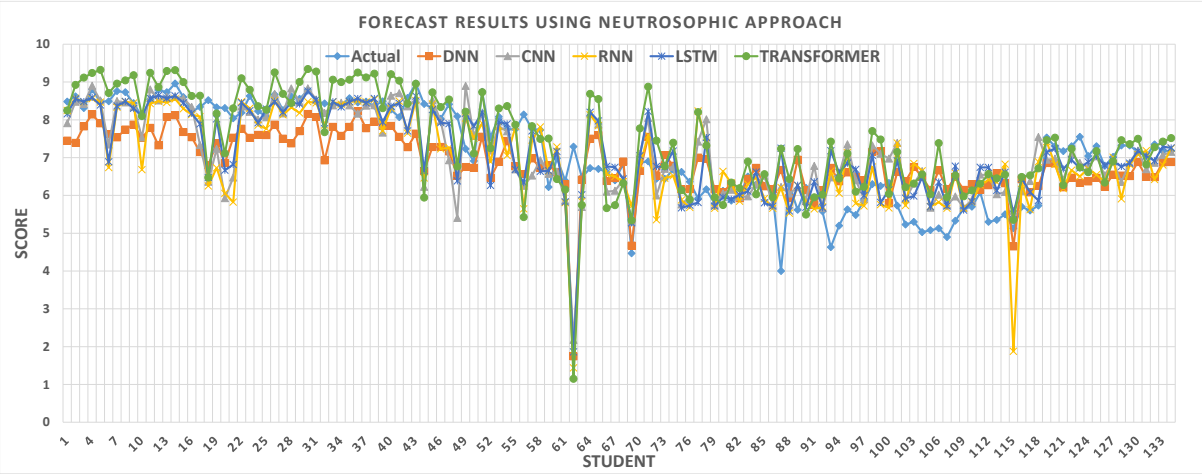


Figure 2. 10. Graph of prediction (neutrosophic data, Case 3)

For Case 3: We use the grade data of the previous three semesters to forecast the SGPA for the fourth semester's grades. From the comparison table 2.9, we see that in the case of using the Neutrosophic approach, all metrics are improved compared to the real case (the case without using the neutrosophic approach). Moreover, from Table 2.9 we can see that, in the case of using the

neutrosophic approach, the R^2-score parameter of all 5 methods is greater than 60%, which proves that all five methods are suitable for the data and the problem.

Comparison results estimate the errors (average after 10 tests) of the algorithms as shown in Table 2.9. From the results are presented in Table 2.9, we conclude that the numerical results that are highlighted in "bold" indicate that the corresponding forecasting method has better results than the other method. It seems that, RNN, LSTM, and Transformer methods significantly outperform the other methods, confirming the effectiveness of our approach as we achieved consistent results on the validation set.

Table 2. 9. Average error comparison for cases 1, 2, 3

Model/Metric		RMSE		MAE		R^2 (%)	
		Real input	**Neutro. Approach**	**Real input**	**Neutro. Approach**	**Real input**	**Neutro. Approach**
Case 1	**DNN**	1.06 ± 0.33	**0.89** ± 0.09	1.07 ± 0.11	**0.75**± 0.06	48.26 ± 32.00	12.76 ± 4.90
	CNN	0.92 ± 0.06	**0.90**± 0.07	**0.73** ± 0.04	0.74± 0.04	8.52 ± 4.65	**11.40** ± 5.06
	RNN	**0.89**± 0.05	0.92 ± 0.07	**0.72**± 0.04	0.73 ± 0.04	12.6 ± 8.49	12.39 ± 6.06
	LSTM	**0.90** ± 0.05	0.91 ± 0.07	0.74 ± 0.03	**0.74** ± 0.04	9.72 ± 5.39	**12.73** ± 4.97
	Transformer	0.90 ± 0.04	**0.89** ± 0.08	0.74 ± 0.03	**0.74** ± 0.04	26± 5.40	**13.13** ± 7.65
Case 2	**CNN**	**0.53** ± 0.05	0.58 ± 0.04	**0.41** ± 0.03	0.46 ± 0.02	46.39 ± 6.51	**47.03** ± 5.80
	RNN	**0.55** ± 0.07	0.60 ± 0.05	**0.42** ± 0.05	0.45 ± 0.03	37.12 ± 18.19	**46.16** ± 6.82
	LSTM	**0.52** ± 0.04	0.57 ± 0.04	**0.40** ± 0.03	0.45 ± 0.03	**52.42** ± 9.95	49.51 ± 5.40
	Transformer	0.63 ± 0.07	**0.59** ± 0.06	0.48 ± 0.04	**0.47** ± 0.06	26.83 ± 5.96	**45.54** ± 8.92
Case 3	**CNN**	0.86 ± 0.08	**0.80 ±** 0.08	0.67 ± 0.07	**0.62 ±** 0.07	59.01 ± 7.00	**62.04 ±** 5.36
	RNN	0.82 ± 0.13	**0.80 ±** 0.08	0.62 ± 0.12	**0.60**± 0.06	60.69 ± 9.20	**62.14** ± 7.67
	LSTM	0.88 ± 0.13	**0.76 ±** 0.11	0.71 ± 0.15	**0.59** ± 0.07	58.51 ± 1.54	**65.28** ± 8.93
	Transformer	0.93± 0.07	**0.79 ±** 0.06	0.77 ± 0.06	**0.59** ± 0.05	53.05 ± 7.60	**65.95** ± 4.33

Based on Table 2.9, the Neutrosophic approach demonstrates consistent improvements over the Real input baseline. In Case 2, the Transformer reduced MAE from 0.60 to 0.47 and increased R^2 from 26.83% to 45.54%; similarly, LSTM improved R^2 from 32.42% to 49.51%. In Case 3, both LSTM and Transformer achieved R^2 values above 65%, approximately 7-10% higher than the real-input approach. These results indicate that incorporating Neutrosophy helps reduce prediction errors and enhances model fit across different educational data scenarios.

Models performance are compared by aligning the actual values from a subset of the test data with the predicted SGPA, followed by an analysis of the resulting errors. To support this comparison, the RMSE metric to quantify the deviation between actual and predicted SGPA, enabling the selection of the most context-appropriate model. In general, the Neutrosophy-LSTM and Transformer methods produced comparable and superior results compared to other approaches.

Among the proposed methods, the Transformer model demonstrated outstanding performance, with a notable margin of 0.37 in the initial trial. These results were consistently confirmed on the validation dataset, with nearly identical scores, highlighting the stability and effectiveness of the proposed approach. Regarding the impact of classification and prediction layers, the addition of a dense layer slightly improved accuracy, as reflected in the performance gaps between models such as RNN versus LSTM and Transformer.

The dataset for SGPA prediction is heavily influenced by the specific academic major of the students. Hanoi Metropolitan University (HNMU) is a young, medium-sized institution located in Hanoi, Vietnam, with approximately 7,000 students currently enrolled. The HNMU1 dataset focuses on students in the Primary Education major, which has an annual intake of only around 200 - 300 students. As a result, the size of the HNMU1 dataset is relatively small, and the outcomes achieved in this context are deemed acceptable.

However, the overall accuracy remains below 66%, which is not sufficient for large-scale practical deployment. Therefore, future improvements should focus on expanding feature sets (including both academic/score-based and non-academic factors), integrating data generation mechanisms or hybrid architectures, to further improve predictive performance and provide stronger support for students, instructors, and administrators.

In response, we propose the development of hybrid deep learning models that leverage the strengths of neutrosophic-integrated architectures while enhancing predictive performance. This direction will be further elaborated in the next section - the NeutroGNT model.

2.3. NeutroGNT model

2.3.1. The theoretical basis for model selection

Based on the findings presented in Section 2.2, it is evident that the Transformer not only maintains performance stability but also effectively leverages neutrosophic representations to improve prediction accuracy. This

highlights the fact that the effectiveness of neutrosophic logic is highly dependent on the underlying model architecture, and that the Transformer emerges as the most appropriate model in this context.

However, overall prediction accuracy remained below 66% (in section 2.2), which -though acceptable for small educational datasets like HNMU1 - falls short of practical deep learning expectations. This limitation is mainly due to two challenges:

(i) the limited size and class imbalance of educational datasets, and

(ii) the inherent uncertainty and subjectivity in academic evaluation.

To address these issues, the dissertation proposes a hybrid deep learning framework that integrates:

- CGAN, to augment data and improve class balance by generating realistic synthetic samples;
- Neutrosophic representations, to model uncertainty and better reflect ambiguity in educational assessments;
- A noise-injection strategy, to enhance robustness and prevent overfitting under noisy, small-scale conditions.

Together Transformer, these components form the basis of the proposed NeutroGNT model, offering a comprehensive solution to the challenges of sparsity and uncertainty in real-world educational data.

2.3.2. Proposed model

In this section, the dissertation proposes a hybrid deep learning framework that integrates the Transformer architecture, CGAN, and neutrosophic input representation. Additionally, a noise-injection strategy is incorporated to enhance the generalization capability of the model. The overall model is presented in Figure 2.11.

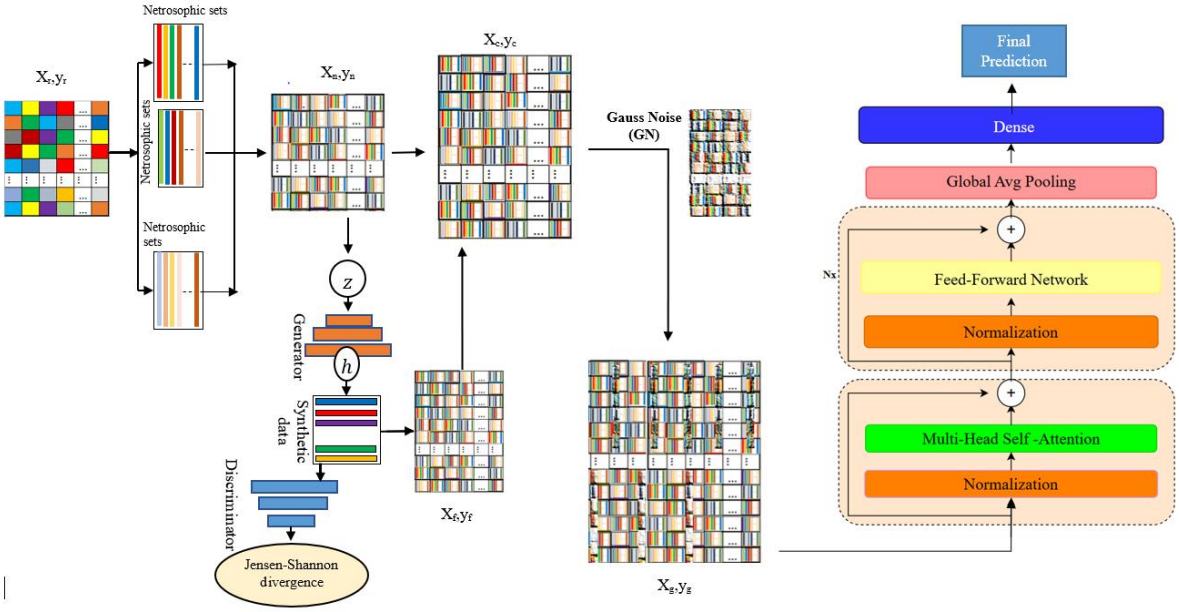


Figure 2. 11. NeutroGNT model [CT5]

Figure 2.11 illustrates the general architecture of the neutrosophic neural network, incorporating CGAN and Transformer models. A noise-injection strategy is incorporated to improve the robustness and generalization capabilities of the predictive model. The functioning of the model illustrated in Figure 2.11 is as follows: Given the real dataset (X_r, y_r), we apply trapezoidal neutrosophic functions to capture uncertainty, indeterminacy, and inconsistency in the data to construct a new dataset denoted as (X_n, y_n). Although neutrosophiclize process is utilized, deep learning models generally require large datasets. To fully leverage deep learning effectiveness, we further incorporate a CGAN to generate synthetic samples and augment the training dataset, forming (X_f,y_f). CGAN is used because it captures the underlying distribution of the original dataset, allowing for an expanded training set. The two datasets (X_n, y_n) and (X_f,y_f) are then concated to form (X_c, y_c). On this consolidated dataset (X_c, y_c), a noise-injection strategy is incorporated to improve the robustness and generalization capabilities of the predictive model, forming (X_g,y_g). This approach is beneficial as it increases diversity, reduces computational complexity, improves prediction performance, enhances robustness to noise, handles missing data more effectively, aids feature discovery, and is particularly effective for small training sets.

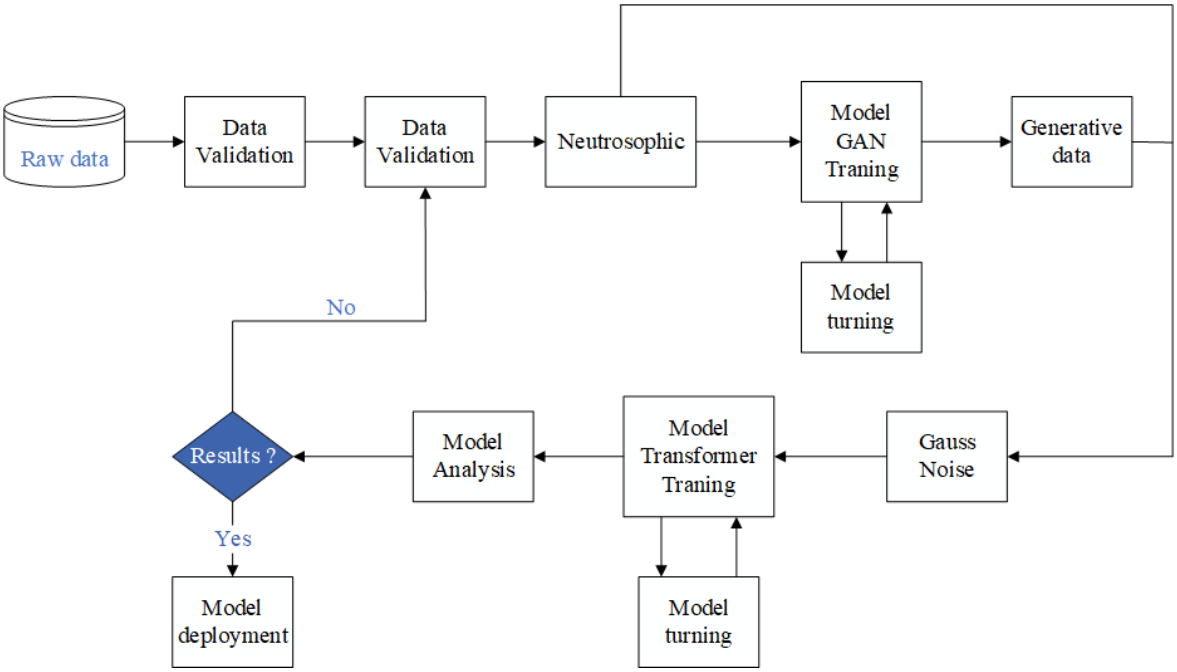


Figure 2. 12. The pineline of NeutroGNT model

Given that educational data is often incomplete or imbalanced across academic performance levels, CGAN is employed to generate conditionally sampled tabular data. This approach helps expand the learning space and improves the generalizability of deep learning models. CGAN also addresses data scarcity issues among minority student groups (e.g., those with very low or very high grades), which could otherwise bias the model's predictions if left unaddressed.

The use of neutrosophic logic allows the model to directly encode the three components of uncertainty: truth, indeterminacy, and falsity, an explicit modeling that traditional fuzzy systems have not adequately addressed. Furthermore, the introduction of controlled noise improves model stability, reduces overfitting, and enhances robustness against measurement errors or noisy data.

Synthesizing both theoretical and empirical analyses, the proposed Transformer-based model emerges as the optimal architecture for handling uncertain educational data. This is achieved through its effective integration with neutrosophic representations and the data generation mechanism provided by CGAN. The principal contribution of this study lies in the design of an integrated neutrosophic encoding/decoding mechanism within a deep learning architecture, significantly enhancing prediction accuracy for early academic performance forecasting and supporting more informed educational decision-making.

The principal contribution of this model lies in the integration of a neutrosophic encoder-decoder mechanism within a deep learning architecture, facilitating more accurate early prediction and identification of students at risk of academic failure. This, in turn, enables timely educational interventions and strategic support planning while also assisting institutions in identifying high-achieving students for advanced academic opportunities.

Algorithm 2.2. NeutroGNT - SGPA prediction with Neutrosophic logic, CGAN, and Transformer

1: Input: D_{real} : Real dataset of student academic records and SGPA
2: Z : Latent noise vector for CGAN
3: G : Number of synthetic neutrosophic samples to generate
4: T_{Neutro}: Transformer model with neutrosophic encoding and noise injection
5: Output: $\hat{Y}$: Predicted SGPA values for test set

6: $[X_r, y_r] \leftarrow$ Preprocess(D_{real}) ▷ Clean, scale, sort by semester
7: $X_{Neutro} \leftarrow$ NeutrosophicTransform(X_r) using trapezoidal membership functions
8: $[G_{CTGAN}, D_{CTGAN}] \leftarrow$ Train CGAN($[X_{Neutro}, y]$, Z)
9: for $i = 1$ **to** G **do**
10: $z_i \leftarrow$ Sample(Z)
11: $y_i \leftarrow$ SampleLabelDistribution(y_r)
12: $X_f[i] \leftarrow G_{CTGAN}(z_i, y_i)$ ▷ Generate synthetic neutrosophic input
13: $y_f[i] \leftarrow y_i$
14: end for
15: $D_{aug} \leftarrow$ Concatenate($[X_{Neutro}, y_r]$, $[(X_f, y_f])$
16: $D_{aug} \leftarrow$ InjectNoise(D_{aug}) ▷ Gaussian noise injection
17: $T_{Neutro} \leftarrow$ TrainTransformer(D_{aug})
18: $\hat{Y} \leftarrow$ Predict(T_{Neutro}, X_{test})
19: return $\hat{Y}$

The Transformer model operates combined to capture complex patterns and dependencies within the data ($\boldsymbol{X_g}$,$\boldsymbol{y_g}$). Finally, performs defuzzification to convert neutrosophic values back to real values and outputs the final prediction. For ease of comparison, all models in this study utilize the Adam optimization algorithm. The errors are evaluated using MSE, MAE, RMSE, R^2.

Table 2. 10. The parameters of the models

Hyper-parameters	Selection
Learning rate α	0.001
Drop-out rate	0.2
Number of epochs	200
Batch size	32
Loss function	Mean square error
Optimizer	Adam

The computational complexity of the proposed model depends on its core architectures, namely Transformer and CGAN. Specifically, the Transformer has a complexity of $O(n \cdot d^2)$, where n is the length of the input learning sequence and d is the hidden vector dimension. For CGAN, training involves two networks: Generator (G) and Discriminator (D), with a complexity of $O(E \cdot (|G| + |D|))$, where E denotes the number of training epochs.

Although the computational cost is higher than that of baseline machine learning models, it remains moderate compared to modern deep learning architectures. The total training time largely depends on the number of epochs and the amount of generated samples.

2.3.3. Experiments

The experiments were conducted in a Conda environment with Python 3.11, integrating libraries such as NumPy, SciPy, Pandas, Scikit-learn, PyTorch, and TensorFlow/Keras. Transformer were trained on GPU using CUDA 11.8. All computations were performed on a workstation with Intel Core i7-12700KF CPU, NVIDIA RTX 3060 GPU, and 32GB RAM.

In this section, we use 06 datasets. 02 datasets among them are collected from Hanoi Metropolitan University [CT3] and Vietnam National University, Hanoi [CT4]. The remaining datasets were obtained from Covenant University in Nigeria ([61]), the University of Malaya in Malaysia ([62]), and the well-known Portuguese school performance dataset ([64]).

Malaya-Stud dataset

This dataset ([62]), provided by Universiti Malaya and licensed under the Creative Commons Attribution 4.0 International License, includes data on 493 students across 33 features. These features encompass demographic details

(such as gender, financial status, and living conditions), study habits, and key academic indicators. The SSC Grade (Secondary School Certificate) represents academic grades at the lower secondary level, while the HSC Grade (Higher Secondary Certificate) reflects academic achievement at the upper secondary level, both serving as foundational indicators of pre-university readiness. The last semester grade captures the student's most recent SGPA at the university level, offering insights into current SGPA and adaptability to higher education. The overall grade, denoting the cumulative SGPA, is used as the primary target variable in predictive modeling to assess overall academic success.

Portugal dataset

The Portugal dataset ([64]) was collected from two Portuguese secondary encompassing academic records of 395 Mathematics students (Portugal-Math dataset) and 649 Portuguese Language students (Portugal-Lang dataset). Alongside demographic and behavioral data (e.g., study time, absenteeism, parental support), the standout feature of these datasets is the sequential recording of student performance at three critical stages of the academic year: G1 (first semester grade), G2 (mid-second semester grade), and G3 (final year grade). This time-series format clearly reflects the academic progression of each student and provides a strong foundation for developing machine learning models capable of forecasting future grades based on previous grades.

Covenant-Priv dataset

This dataset ([61]) comes from Covenant University, Nigeria. This large-scale educational dataset contains academic information from 1841 undergraduate students majoring in engineering from 2002 to 2014. It includes records of students from seven disciplines: Chemical Engineering, Civil Engineering, Computer Engineering, Electrical and Electronics Engineering, Information and Communication Engineering, Mechanical Engineering, and Petroleum Engineering. The data includes semester GPA from the first to the fifth year, along with the cumulative GPA (CGPA), with scores ranging from 0 to 5.

The details of the datasets are described in Table 2.11.

Table 2. 11. Training dataset description

	Name	M	S	K	Case	X	Input feature	Output
1	HNMU2	551	52	88	1	1	GPA Semester 1	GPA Semester 2
					2	2	GPA Semester 1, GPA Semester 2	GPA Semester 3
					3	3	GPA Semester 1, GPA Semester 2, GPA Semester 3	GPA Semester 4
2	VNU	271	43	91	2	2	GPA Semester 1, GPA Semester 2	GPA Semester 3
					3	3	GPA Semester 1, GPA Semester 2, GPA Semester 3	GPA Semester 4
3	Malaya- Stud	493	4	16	3	3	HSC, SSC, Last	Overall
4	Portugal-Math	395	3	33	2	2	G1, G2	G3
5	Portugal-Lang	649	3	33	2	2	G1, G2	G3
6	Covenant-Priv	1841	6	9	3	3	First Year GPA, Second Year GPA, Third Year GPA	Fourth Year GPA

Data's name, Sample size (M), Number of score-related features (S), the total of features (k), Input feature count (X), case using, name input feature, name output feature and web- link to data sources.

2.3.4. Results and discussions

In this subsection, we present three case studies of student data to illustrate the proposed method. Prior to experimentation, all records were preprocessed to remove missing values and eliminate scores outside the 0–10 range. The datasets were then split into 80% for training and 20% for testing.

Several experimental integration scenarios were designed for comparison, including:

- Integrating the neutrosophic framework with Transformer (Neutro_T),
- Integrating the neutrosophic framework and CGAN with Transformer (NeutroCT),
- The combination of all three components: the neutrosophic framework, CGAN and noise injection with Transformer (NeutroGNT).

Detailed experiments are presented for each case as follows.

Results for Case 1. The experimental results for Case 1 are presented in Table 2.12 for HNMU2 dataset.

Table 2. 12. Demonstrated errors (averaged over 10 runs - case 1)

Dataset		Real_T	Neutro_T	NeutroCT	NeutroGNT
HNMU2	MSE	0.519 ± 0.028	0.474 ± 0.040	0.469 ± 0.031	**0.458 ± 0.011**
	MAE	0.576 ± 0.014	0.560 ± 0.029	0.558 ± 0.022	**0.548 ± 0.010**
	R^2	-0.087±0.058	0.008 ± 0.085	0.017 ± 0.064	**0.041 ± 0.022**

Although NeutroGNT achieved the lowest MSE (0.458 ± 0.011) and showed a notable 12.8% improvement in R^2 compared to the Real_T model, the resulting R^2 value remains relatively low (0.041 ± 0.022), hovering near the threshold where the model fails to explain the variance in the data. This suggests that despite the reductions in absolute and squared errors, the model's generalization and explanatory capabilities are still limited, especially in real-world scenarios like the HNMU2 dataset, which involves high levels of noise and uncertainty.

Results for Case 2. The experimental results for Case 2 are presented in Table 2.13.

Table 2. 13. Demonstrated errors (averaged over 10 runs - case 2)

Dataset		Real_T	Neutro_T	NeutroCT	NeutroGNT
HNMU2	MSE	0.323 ± 0.101	0.183 ± 0.024	0.208 ± 0.052	**0.181 ± 0.030**
	MAE	0.459 ± 0.085	0.339 ± 0.025	0.363 ± 0.053	**0.338 ± 0.035**
	R^2	0.077 ± 0.288	0.478 ± 0.069	0.407 ± 0.147	**0.482 ± 0.084**
VNU	MSE	0.302 ± 0.031	0.320 ± 0.042	0.321 ± 0.044	**0.260 ± 0.04**6
	MAE	0.441 ± 0.032	0.453 ± 0.039	0.451 ± 0.042	**0.381 ± 0.054**
	R^2	0.201 ± 0.083	0.153 ± 0.112	0.150 ± 0.116	**0.202 ± 0.140**
Portugal-Math	MSE	2.536 ± 2.129	1.263 ± 0.080	1.409 ± 0.135	**1.197 ± 0.074**
	MAE	1.065 ± 0.567	0.770 ± 0.069	0.844 ± 0.077	**0.725 ± 0.043**
	R^2	0.505 ± 0.415	0.754 ± 0.016	0.725 ± 0.026	**0.767 ± 0.014**
Portugal-Lang	MSE	0.704 ± 0.550	**0.423 ± 0.014**	0.435 ± 0.032	0.440 ± 0.033
	MAE	0.528 ± 0.241	**0.403 ± 0.004**	0.413 ± 0.027	0.425 ± 0.027
	R^2	0.711 ± 0.225	**0.826 ± 0.006**	0.822 ± 0.013	0.820 ± 0.013

In Case 2, the proposed neutrophilization-based deep learning models consistently demonstrate superior and stable performance across all four benchmark datasets. Notably, the NeutroGNT model achieves remarkable results on the VNU dataset, with an MSE of 0.260 ± 0.046 and MAE of 0.381 ± 0.054. More importantly, on the HNMU2 dataset, the R^2 score of NeutroGNT increases by more than 40.5% compared to the baseline Real_T model,

highlighting a substantial enhancement in predictive capability. These findings reinforce the robustness and effectiveness of the NeutroGNT model, positioning it as a strong candidate for broader applications in future educational datasets.

On the Portugal-Math dataset, the NeutroGNT model shows a notable improvement in the R^2 score, reaching 0.767 ± 0.014, which represents a 26.2% increase over Real_T, while the MSE is also reduced by 1.339.

Although NeutroGNT achieved the best performance in terms of MSE and MAE across most datasets (HNMU2 and Portugal-Math), several limitations remain noteworthy. On the HNMU2 dataset, despite NeutroGNT having the highest R^2 (0.482 ± 0.084), the relatively large standard deviation reflects a lack of stability across multiple runs. This is particularly critical in real-world educational settings, where models must ensure consistency and high reliability.

Results for Case 3

The experimental results for Case 3 are presented in Table 2.14. For the dataset from Universiti Malaya, three input features - "HSC", "SSC", and "Last" - are used to predict the "Overall" score, and all values are normalized to a 10-point scale to ensure consistency. The Covenant-Priv dataset is retained in its original scale for comparison with previous studies, where the input features are "First Year GPA", "Second Year GPA", and "Third Year GPA", and the output is the "Final CGPA".

Table 2. 14. Demonstrated errors (averaged over 10 runs - case 3)

Dataset		Real_T	Neutro_T	NeutroCT	NeutroGNT
HNMU2	MSE	0.212 ± 0.088	0.208 ± 0.081	0.175 ± 0.082	0.152 ± 0.025
	MAE	0.374 ± 0.078	0.382 ± 0.083	0.347 ± 0.081	0.322 ± 0.029
	R^2	0.047 ± 0.393	0.068 ± 0.364	0.216 ± 0.367	0.319 ± 0.111
VNU	MSE	0.119 ± 0.037	0.109 ± 0.041	0.121 ± 0.061	**0.088 ± 0.017**
	MAE	0.281 ± 0.039	0.271 ± 0.051	0.282 ± 0.074	**0.242 ± 0.026**
	R^2	0.549 ± 0.140	0.588 ± 0.154	0.541 ± 0.230	**0.666 ± 0.064**
Malaya – Stud	MSE	0.495 ± 0.563	**0.342 ± 0.03**8	0.412 ± 0.063	0.400 ± 0.055
	MAE	0.505 ± 0.249	**0.434 ± 0.025**	0.485 ± 0.048	0.473 ± 0.036
	R^2	0.788 ± 0.241	**0.854 ± 0.016**	0.824 ± 0.027	0.829 ± 0.024
Covenant - Priv	MSE	0.023 ± 0.001	0.022 ± 0.001	0.023 ± 0.003	**0.019 ± 0.00**2
	MAE	0.116 ± 0.003	0.114 ± 0.002	0.118 ± 0.008	**0.107 ± 0.005**
	R^2	0.949 ± 0.002	0.952 ± 0.001	0.950 ± 0.007	**0.958 ± 0.003**
	RMSE	0.152 ± 0.003	0.147 ± 0.002	0.150 ± 0.009	**0.138 ± 0.005**

On the Malaya-Stud dataset, all models performed well with consistently high R^2 scores, with Neutro_T standing out by achieving the highest R^2 (0.854) and the lowest MSE (0.342), reflecting both stability and high prediction accuracy. Meanwhile, VNU proved more challenging, as all models yielded relatively low and fluctuating R^2 scores.

However, in more complex datasets such as HNMU2, the gains are limited (R^2 remains below 0.32), reflecting the strong presence of noise, contextual variability, and subjectivity in the data. This indicates that, despite incorporating GAN-based data generation and multi-attribute features, the model still faces challenges when dealing with inconsistent grading standards, institutional heterogeneity, and hidden latent factors not captured in the datasets.

Among the evaluated models, NeutroGNT stands out for achieving the best balance between accuracy and robustness on the Covenant-Priv dataset. It recorded the highest average R^2 score (0.958 ± 0.003), clearly outperforming other models. Notably, its average RMSE (0.138 ± 0.005) is 0.138 lower than that of the Real_T model.

Furthermore, it achieved a minimum RMSE of 0.1342, which is lower than the best result previously reported by Aderibigbe et al (2019). Additionally, its minimum MSE of 0.018 is the lowest across the entire study, and the maximum R^2 of 96.05% surpasses all prior benchmarks. These results confirm the superior predictive performance and effectiveness of the NeutroGNT model.

These results underscore the superior overall performance of the NeutroGNT model under average evaluation. Collectively, the Neutro approach emerges as a reliable and effective predictive framework for all cases 1, 2 and 3, particularly well-suited for international educational datasets and complex academic forecasting tasks.

2.4. Appendix to Chapter 2

2.4.1. Overview of Neutrosophy theory

Neutrosophy, first introduced by Florentin Smarandache ([74]; [75]), is a philosophical framework and mathematical foundation designed to handle uncertainty, imprecision, indeterminacy, and inconsistency in knowledge representation. Unlike classical logic, which operates under binary true/false

conditions, and fuzzy logic, which introduces a degree of truth, neutrosophy simultaneously considers three components for any proposition or statement: truth (T), indeterminacy (I), and falsity (F).

Definition 2.1. ([74]) A neutrosophic set (NS) A, defined on the universe of discourse X and denoted generally by x, can be represented in following form:

$$A = \{\big(x, T_A(x), I_A(x), F_A(x)\big): x \in X\} \tag{2.2}$$

where each element x in the set X is associated with three membership functions: $T_A(x)$ the truth membership function, $T_A: X \to [0,1]$, representing the degree of confidence or certainty that x belongs to the set, the indeterminacy membership function, $I_A: X \to [0,1]$, representing the degree of uncertainty or ambiguity about whether x belongs to the set, and $F_A: X \to [0,1]$: the falsity membership function, representing the degree of skepticism or disbelief that x belongs to the set. The sum of these membership values must satisfy the condition $0 \leq T_A(x) + I_A(x) + F_A(x) \leq 3$ for all $x \in X$, ensuring that the total membership remains within a valid range.

Example 2.1. The single valued trapezoidal neutrosophic number, $[a, b, c, d; T_N, I_N, F_N], a \leq b \leq c \leq d; 0 \leq T_N, I_N, F_N \leq 1,$ in the general formula, has following membership functions

$$T(x) = \begin{cases} 0 \quad if \quad x \leq a, x > d \\ \dfrac{(x-a)T_N}{b-a} \quad if \quad a < x \leq b \\ T_N \quad if \quad b < x \leq c \\ \dfrac{(d-x)T_N}{d-c} \quad if \quad c < x \leq d; \end{cases}$$

$$I(x) = \begin{cases} 1 \quad if \quad x \leq a, x > d \\ \dfrac{b-x+(x-a)I_N}{b-a} \quad if \quad a < x \leq b \\ I_N \quad if \quad b < x \leq c \\ \dfrac{x-c+(d-x)I_N}{d-c} \quad if \quad c < x \leq d; \end{cases} \tag{2.3}$$

$$F(x) = \begin{cases} 1 \quad if \quad x \leq a, x > d \\ \dfrac{b-x+(x-a)F_N}{b-a} \quad if \quad a < x \leq b \\ F_N \quad if \quad b < x \leq c \\ \dfrac{x-c+(d-x)F_N}{d-c} \quad if \quad c < x \leq d, \end{cases}$$

$x \in R,$ are used in this context, where T_N, I_N, F_N are the truth degree, the indeterminacy degree, and the falsity degree, respectively**.**

In this example, we give a simple example on the single-valued trapezoidal neutrosophic functions $N = [0,0.2,3.7,4.1; 1.0,0.1,0.5]$**.** According

to Definition 2.1, we receive the truth membership function, indeterminacy membership function and falsehood membership function as follows:

$$T_N(x) = \begin{cases} 0 & x \notin [0, 4.1] \\ \dfrac{x}{0.2} & 0 < x \le 0.2 \\ 1 & 0.2 < x \le 3.7 \\ \dfrac{4.1 - x}{0.4} & 3.7 < x \le 4.1. \end{cases};$$

$$I_N(x) = \begin{cases} 1 & x \notin [0, 4.1] \\ \dfrac{0.2 - 0.1x}{0.2} & 0 < x \le 0.2 \\ 0.1 & 0.2 < x \le 3.7 \\ \dfrac{0.9x - 3.29}{0.4} & 3.7 < x \le 4.1. \end{cases};$$

$$F_N(x) = \begin{cases} 1 & x \notin [0, 4.1] \\ \dfrac{0.2 - 0.5x}{0.2} & 0 < x \le 0.2 \\ 0.1 & 0.2 < x \le 3.7 \\ \dfrac{0.5x - 1.65}{0.4} & 3.7 < x \le 4.1. \end{cases}. \quad (2.4)$$

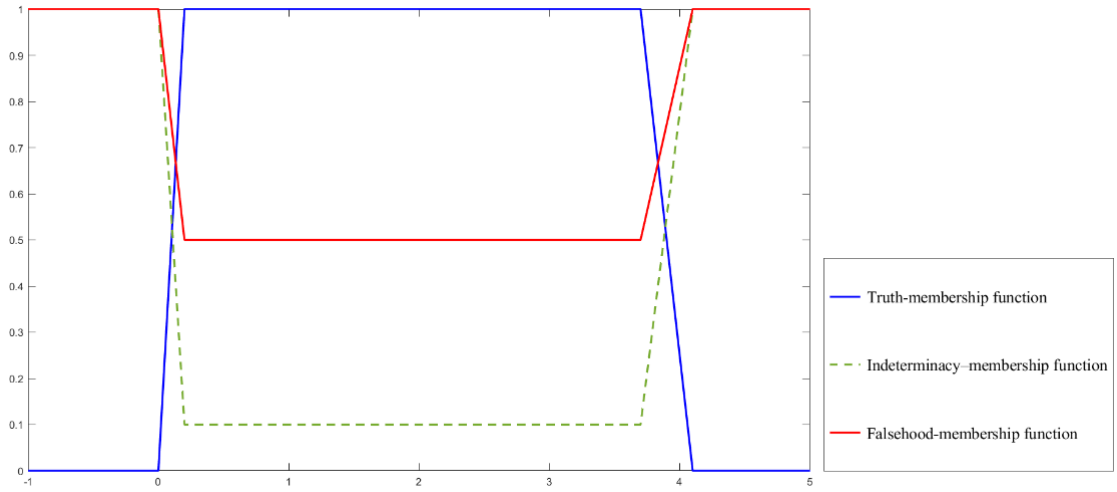


Figure 2. 13. The single-valued trapezoidal neutrosophic functions

The graphical representations of the neutrosophic number N is given in Figure 2.13. Leveraging the strength of fuzzy and neutrosophic sets in handling ambiguous data, recent studies have integrated neutrosophic sets

(NS) with machine learning and deep learning models. Ejegwa et al. ([73]) showed the value of fuzzy sets in pattern recognition using a soft computing approach. The feasibility of fuzzy sets in machine learning through soft computing methods and provided applications in pattern recognition problems of construction materials and mineral mines have been discussed.

In our framework, a neutrosophic encoder-decoder module applies neutrosophic logic to better manage uncertainty and indeterminacy in input and output data.

The idea of incorporating neutrosophic logic into deep learning models has also been proposed by some authors. For instance, in the work of Mayukh et al. ([76]), the authors utilized a neutrosophic approach in several extended LSTM and Transformer models for sentiment analysis tasks and demonstrated the potential and effectiveness of this combined approach.

2.4.2. Summary of GAN and CGAN

2.6.1.1. Generative Adversarial Networks

Generative Adversarial Networks (GAN) represent a powerful class of unsupervised deep learning models in which two neural networks, the Generator (G) and the Discriminator (D), engage in a dynamic adversarial process ([77]). As stated, the Generator seeks to map a latent noise vector $z \backsim p(z)$ to a data-like output $\tilde{x} = G(z)$, while the Discriminator learns to distinguish real data samples $x \backsim P_r$ from the synthetic ones $\tilde{x} \backsim \mathrm{P}_g$. The competition between Generator G and Discriminator D is formulated as a minimax objective, illustrated in equation (2.5):

$$\min_G \max_D E_{x \backsim \mathrm{P}_r}[logD(x)] + E_{\tilde{x} \backsim \mathrm{P}_g}[\log(1 - D(\tilde{x})], \tag{2.5}$$

where P_r is the data distribution and P_g is the distribution implicitly defined by the generator's output.

In case Discriminator D is trained to the optimal level before each parameter update of Generator G, minimizing the value function minimizes the Jensen-Shannon divergence between P_r và P_g. However, doing so often leads to the disappearance of the derivative when Discriminator D is saturated.

2.6.1.2. Conditional generative adversarial network (CGAN)

The Conditional Generative Adversarial Network (CGAN) is a variant of the original GAN ([19]). Since CGAN is a conditional generative model, both the Generator and Discriminator networks are trained simultaneously, with both receiving the label of the data as input, ensuring they generate and evaluate data that aligns with specific labels.

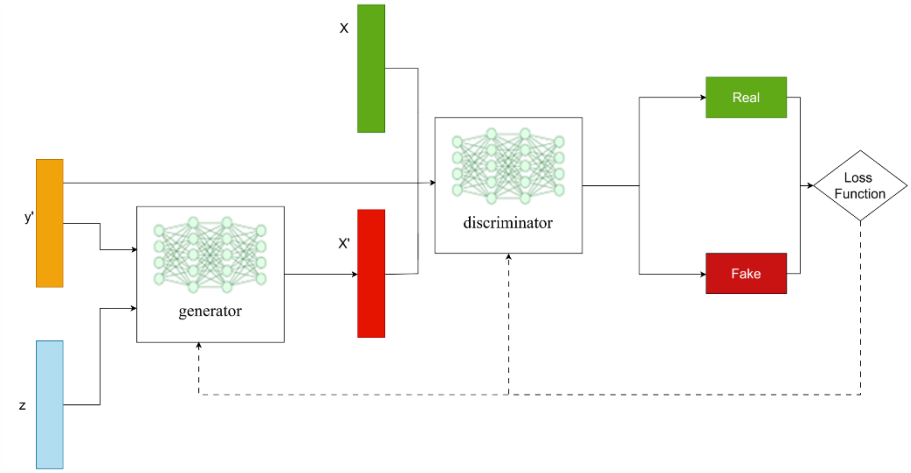


Figure 2. 14. CGAN model

The CGAN operates as follows: the Generator network takes as input a noise vector z and a condition label y', generating new data according to the condition provided by y. The real samples (x, y') and the newly generated samples (x', y') are then passed to the Discriminator network, which distinguishes between real and fake samples. This process is akin to a min-max game between two players, with the loss function calculated as

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x \mid y)] + E_{z\sim p_z(z)}[\log(1 - D(G(z \mid y) \mid y))]. \quad (2.6)$$

For real data input x and label y, $logD(x \mid y)$ is the probability that the Discriminator believes the data x (with label y) is real. The Discriminator's goal is to maximize this probability as much as possible. Meanwhile, $G(z \mid y)$ generates fake data using the Generator model from the noise matrix z, based on label y. $\log(1 - D(G(z \mid y) \mid y))$ is the probability that the Discriminator believes the newly generated data (with label y) is fake. The Discriminator aims to maximize this probability, while the Generator's goal is to make $D(G(z \mid y) \mid y)$ as close to 1 as possible, meaning it has successfully fooled the Discriminator.

The principle for model selection is that the CGAN model with the smallest FID value (FID is a method for assessing the difference between generated data and real data) will be chosen.

$$FID = ||\mu_r - \mu_g||^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}}, \quad (2.7)$$

where μ_r, μ_g are the average vector of features of real data and generated data. Σ_r, Σ_g are the variance matrix of real data and generated data. Tr is the trace of the matrix, i.e. the sum of the elements on the main diagonal of the matrix. ||.|| is the Euclidean distance between two vectors.

2.4.3. The Transformer model for the SGPA prediction task

The architecture of the Transformer model in this chapter is specifically designed to process tabular data composed entirely of continuous features. The model consists of three main components: a projection layer for the continuous input features, a stack of N Transformer blocks, and a multi-layer perceptron (MLP). Each Transformer block incorporates a multi-head self-attention mechanism and a position-wise feed-forward subnetwork. The overall architecture is illustrated in Figure 2.14.

Assume that a data sample is represented by the pair (x, y), where $x_{cont} \in R^P$ is a vector of P continuous features.

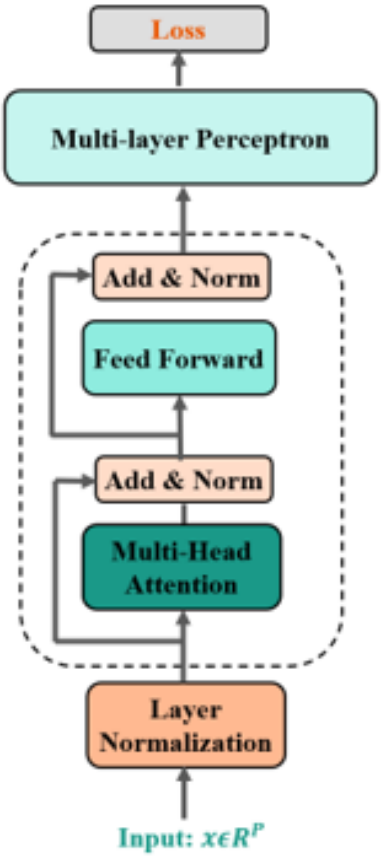


Figure 2. 15. The basic Transformer model for the SGPA prediction task.

First, the input vector x_{cont} is divided into a sequence of k values, and each value is projected into a d-dimensional space through a shared linear projection layer. This results in an input embedding matrix $X_{emb} \in R^{P\times d}$, where each row represents the score of a subject transformed into a feature vector. Each score is mapped to a feature vector via the linear projection, allowing the model to learn complex relationships between scores without the need for positional encoding or categorical embeddings.

The embedded matrix X_{emb} is then passed through a sequence of N Transformer layers. These layers iteratively update the representation of each element in the sequence by aggregating contextual information from the entire sequence. This process is denoted as the function f_θ

$$f_\theta(X_{emb}) = H = \{h_1, \ldots, h_P\}, h_i \in R^d, i = 1, .., P. \qquad (2.8)$$

After obtaining the contextualized embeddings H, an aggregation operation, such as averaging or flattening, is performed to produce a single vector $h_{final} \in R^{P\times d}$. This vector is then fed into a multi-layer perceptron (MLP) denoted as g_ψ to generate the final prediction output.

$$\hat{y} = g_\psi(h_{final}). \qquad (2.9)$$

In regression tasks utilizing the Transformer architecture, the MSE is widely regarded as the most appropriate and commonly used loss function. MSE effectively measures the average of the squared differences between predicted values and actual targets, making it well-suited for learning continuous-valued outputs. Its mathematical properties, being convex and differentiable, facilitate stable and efficient optimization through gradient-based algorithms such as Adam or SGD. Additionally, the output structure of Transformer models, particularly those adapted for tabular or time-series data, naturally aligns with the scalar or vector predictions required for MSE-based regression. As a result, MSE remains the default choice in most Transformer-based regression frameworks, ensuring both accuracy and optimization efficiency.

All parameters of the model, including the projection layer, Transformer layers θ, and MLP ψ, are trained end-to-end using first-order optimization algorithms such as Adam.

Transformer Layers

Each Transformer layer enables any given score embedding to attend to all other scores in the sequence, thereby allowing the model to learn inter-subject and inter-semester relationships as well as factors influencing students' academic progress. The structure of each layer includes a multi-head self-attention mechanism, followed by two linear feed-forward layers. Each step is equipped with residual connections and layer normalization to stabilize and enhance the learning process.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V . \quad (2.10)$$

Where $Q = X_{emb}W_Q$, $K = X_{emb}W_K$, $V = X_{emb}W_V$; W_Q, W_K, $W_V \in R^{d \times d_k}$ are learnable projection matrices. d_k is the dimensionality of the key/query vectors.

The conclusion of Chapter 2

The prediction of students' SGPA plays an essential role in monitoring academic progress, enabling early interventions, and supporting personalized educational planning. However, in real-world educational environments, SGPA is influenced by numerous uncertain, subjective, and evolving factors - ranging from multi-component assessment structures to shifts in teaching methods and the psychological states of learners. As such, SGPA should not be interpreted as a static or precise value but rather as a dynamic indicator affected by uncertainty and variability.

This chapter has highlighted the need for robust predictive models that can handle incomplete and uncertain educational data. Two key frameworks were introduced: **NeutroDLs** and **NeutroGNT.** NetroGNT achieved a minimum MSE of 0.018 and a maximum R^2 of 96.05%, significantly outperforming conventional methods.

The results of the SGPA prediction study (a short-term regression task aimed at monitoring academic progress and providing timely individual support) in this chapter lay the foundation and provide motivation for developing the graduation classification prediction (a long-term classification problem with strategic, system-wide implications that supports policy planning and quality enhancement in education under uncertain and data-scarce conditions), to be addressed in the next chapter.

CHAPTER 3: ENHANCING THE PERFORMANCE OF EARLY GRADUATION CLASSIFICATION MODELS

To further improve the performance of early graduation prediction models for university students, this chapter presents two advanced hybrid deep learning models: LATCGAd and AWG-GC. Both models are designed to address the challenges of limited and imbalanced educational data by automatically augmenting training data and leveraging state-of-the-art deep learning architectures to improve predictive capability. LATCGAd combines Transformer, CGAN, and Adaptive Layer Normalization (AdaLN) to improve data quality, stabilize training, and reduce overfitting, reaching 96.97% accuracy and 73.66% F1-score. AWG-GC integrates Autoencoder, Wasserstein GAN, and Graphormer for joint representation learning, data augmentation, and classification, achieving 98.54% accuracy and 99.25% F1-score, significantly surpassing baseline models. The contents of this chapter are based on the research presented in publications [CT2], [CT7] and [CT8].

3.1. Problem formulation

3.1.1. Early prediction of graduation classification problem

At a higher education institution, a student's graduation classification is determined based on their final GPA upon completion of all academic semesters.

The early prediction of graduation classification task refers to estimating a student's final graduation outcome (e.g., Excellent, Good, Medium…) based on academic data from their early semesters.

The conversion scale for graduation classification is shown in Table 3.1.

Table 3. 1. Graduation classification based on final GPA

Classification	10-Point Scale	4-Point Scale
Excellent	[9.0–10]	[3.6–4.0]
Very Good	[8.0–9.0)	[3.2–3.6)
Good	[7.0–8.0)	[2.5–3.2)
Medium	[5.0–7.0]	[2.0–2.5)
Poor	[4.0–5.0)	[1.0–2.0)
Very Poor	[0–4.0)	[0–1.0)

Practical Significance

- **For students**: Early awareness of their potential graduation classification enables them to adjust their study plans, select courses more effectively, improve academic performance, and make better-informed career decisions.
- **For institutions**: Early prediction helps identify students at risk of low graduation outcomes or delayed completion, allowing timely support and intervention. It also supports administrators in refining curriculum design, admission strategies, and academic policies.

Data Used: In addition to academic scores, predictive models can incorporate personal factors (e.g., gender, interests, soft skills), family background (e.g., parents' education, region), societal influences (e.g., learning habits, peer environment), and institutional characteristics (e.g., faculty quality, infrastructure, curriculum).

The problem is formulated as follows: Given input data encompassing personal information, study habits, environmental factors, and grades during the first and second years of undergraduate study, the goal is to accurately predict the student's final graduation classification.

Formally, let the dataset consist of M samples, each represented by P features. These features include students' academic scores from the first and second years of university, along with encoded survey data reflecting personal background, study habits, and environmental conditions.

$$X = \{x^j = (x_1^j, x_2^j, \dots, x_P^j) | x_i^j \in \mathbb{R}, i \in \{1, .., P\}, j \in \{1, \dots, M\}\}, \quad (3.1)$$

and a portion of the data is labeled

$$Y = \{y_1, y_2, \dots, y_L\}, y_i \in \{1, \dots, 6\}, i \in \{1, \dots, L\}, L \geq 1; L \ll M \quad (3.2)$$

associated with graduation classifications for each student, corresponding to the following categories: Excellent, Very Good, Good, Medium, Poor or Very Poor (already encoded).

Problem requirement

Model construction: Determine a mapping $f: \mathbb{R}^P \to \mathbb{N}$ such that it can accurately predict the graduation classification $y_j \in \{1, .., 6\} \subset \mathbb{N}$ for each student $x^j \in X \subset \mathbb{R}^P, j = 1, .., M$.

Building predictive models for early graduation classification is a valuable tool for personalized academic advising and strategic educational planning. By integrating both academic and non-academic data, such models offer a more holistic view of student potential, contributing to improved training quality and more effective educational management.

3.1.2. Learning Analytics with graph data

In addition to the related works on graduation classification discussed in Subsection 1.3.1, this subsection focuses on the LAGT model. The LAGT (Learning Analytics with Graph Convolutional Network and Transformer) framework, introduced in [CT2], constitutes a significant advancement in graduation classification prediction. Its architecture is organized into two main phases: a preprocessing phase using GCN to augment and normalize the training set by leveraging structural relationships among students, courses, and learning factors; and a prediction phase using Transformer to capture semantic representations and model complex spatio-temporal dependencies among input variables. This division allows each component to contribute its unique strengths - GCN for structural representation and feature enrichment, and Transformer for context-aware learning with high accuracy.

Experimental results on three datasets demonstrate that LAGT achieves accuracy of up to 92.73%, outperforming strong baselines such as DNN, GAT, and standalone Transformer. These findings validate the effectiveness of integrating GCN and Transformer for educational data and suggest that additional techniques, such as data augmentation (e.g., SMOTE), can further improve performance. Building on this foundation, the subsequent part of this dissertation extends the LAGT model into more advanced hybrid architectures (LATCGAd, AWG-GC) to address the persistent challenges of small-scale, imbalanced, and uncertain datasets, thereby enhancing the robustness and reliability of predictive outcomes.

3.2. The LATCGAd model

3.2.1. The theoretical basis for model selection

Early prediction of students' graduation classification in a fragmented, non-uniform, and small-scale educational dataset environment requires a deep

learning model capable of handling imbalanced data, extracting features effectively, and learning stably under low-data conditions.

In parallel, recent advancements in generative models, particularly GAN and their variant CGAN, have emerged as state-of-the-art solutions for generating high-quality synthetic data. This is especially relevant given that small and imbalanced datasets remain a major barrier to deploying effective large-scale LA systems.

Based on these requirements, the dissertation proposes the LATCGAd model, a hybrid architecture that integrates three main components: CGAN, Transformer Encoder, and Adaptive Layer Normalization (AdaLN). Each component supports specific learning functions and complements the others to enhance the overall performance of the model.

- Conditional data augmentation with CGAN: To address the issue of class imbalance commonly found in educational datasets, CGAN is employed as a data augmentation technique. In this approach, CGAN is employed to generate synthetic samples $x^* \sim p_g(x|y)$, where $y \in \{C_1, C_2, \dots, C_k\}$ represents the graduation classification labels. These labels correspond to distinct academic performance levels, which are often imbalanced in real-world datasets, such as the "medium" or "excellent" classes in the HNMU1 and HNMU2 datasets.

CGAN is trained to generate label-conditioned synthetic data, where the Generator learns to mimic real samples and the Discriminator distinguishes them based on the same label. This approach helps balance class distribution by augmenting minority classes, thereby improving model performance. (see Subsection 2.4.2, Chapter 2).

- Extracting complex relationships with Transformer Encoder: After the dataset has been augmented, the model employs a Transformer Encoder to learn nonlinear and long-range dependencies among input features. Unlike RNN or LSTM, Transformer do not rely on sequential processing; instead, they fully utilize the attention mechanism, enabling efficient learning even when the input data lacks sequential structure. At each encoder layer, the feature representation $X \in R^{n \times d}$ is updated through a multi-head attention mechanism. With multiple attention heads operating in parallel, the model can capture multi-dimensional

dependencies among input features such as GPA, course outcomes, or non-academic factors.

- Stabilizing training with Adaptive Layer Normalization (AdaLN). Although Transformers exhibit powerful learning capabilities, they tend to overfit and converge slowly on small datasets. To address this, the study incorporates AdaLN into each Transformer Encoder layer. AdaLN is an extension of Layer Normalization that adapts normalization parameters based on the characteristics of the input data. It is particularly useful in deep learning models like Transformers, where input distributions can vary throughout training. AdaLN enhances model performance by dynamically adjusting parameters to reduce bias and variance across layers. In standard Layer Normalization, normalization is applied by computing the mean μ and standard deviation σ of the inputs:

$$\widehat{x_i} = \frac{x_i - \mu}{\sigma}. \quad (3.3)$$

After normalization, learned parameters γ and β (scale and shift) are applied:

$$y_i = \gamma \widehat{x_i} + \beta. \quad (3.4)$$

AdaLN adapts these parameters for each input dynamically:

$$y_i = \gamma(x) \cdot \widehat{x_i} + \beta(x), \quad (3.5)$$

where $\gamma(x)$ and $\beta(x)$ are computed based on the input x for each layer. These parameters are learned through a sub-network, allowing for adjustment according to the data's characteristics. This is particularly useful in cases where the data has a non-uniform distribution, as it helps reduce internal covariate shift, improves convergence, and lowers the risk of overfitting. In Transformer models, AdaLN improves stability and efficiency, especially when working with small or heterogeneous datasets, ensuring effective learning across layers.

In general, the integration of CGAN, Transformer Encoder, and AdaLN into the hybrid LATCGAd model provides three main benefits: (a) Addressing label imbalance: CGAN conditionally generates samples for minority classes. (b) Learning powerful representations: The Transformer captures multi-dimensional nonlinear relationships among features. (c) Ensuring stable convergence and reducing overfitting: AdaLN adapts the normalization process contextually, supporting effective learning on small datasets.

3.2.2. Proposed model

Figure 3.1 illustrates the LATCGAd model, where the combination of the CGAN and Transformer Encoder provides an effective solution for accurately predicting graduation classification on small and imbalanced datasets. In this model, CGAN expands and balances the dataset by generating synthetic samples for specific labels, addressing the issue of data scarcity in underrepresented groups. Once the dataset is expanded and balanced, the Transformer Encoder learns from this diverse dataset, optimizing its ability to capture complex relationships among input features. Notably, to improve accuracy and ensure robust performance on small datasets, the model integrates AdaLN into each Transformer Encoder layer. AdaLN automatically adjusts normalization parameters based on the characteristics of the input data. It reduces bias and variance across network layers, enhancing convergence and overall model performance. As a result, the Transformer learns more stably and mitigates overfitting-a common challenge when working with small datasets. The tight integration of CGAN, AdaLN, and Transformer not only improves accuracy but also enhances precision, recall, and F1-score, enabling the model to make more reliable and comprehensive predictions.

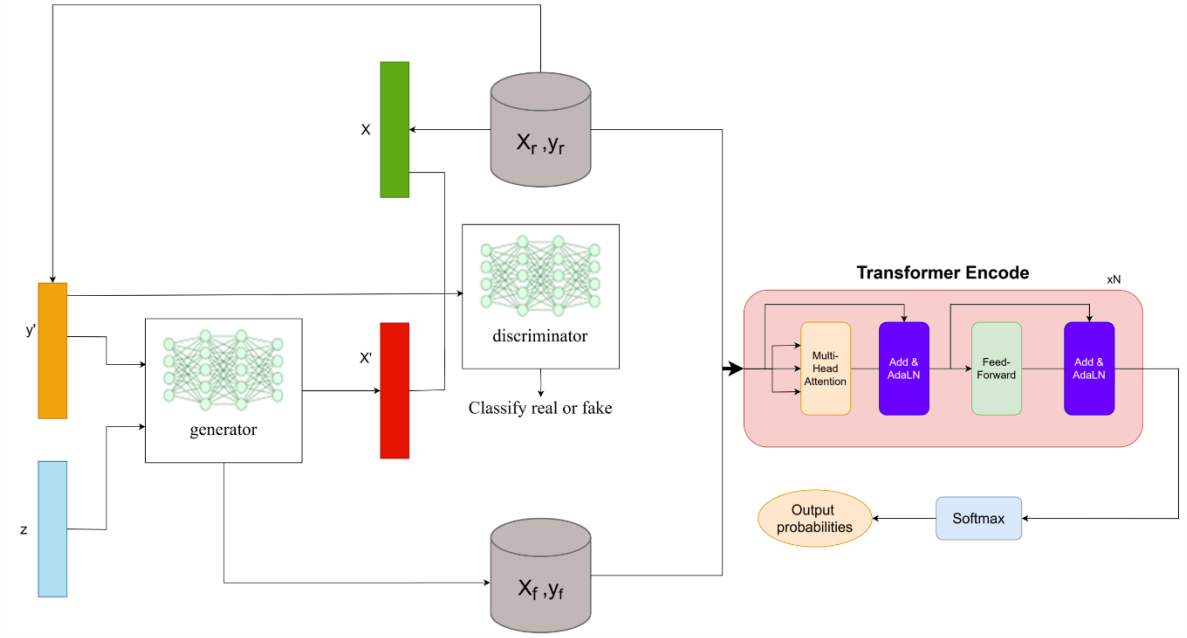


Figure 3. 1. The LATCGAd model ([CT7])

The operation of the proposed model is depicted in Figure 3.1**.** Real data samples X_r (which contain student-related information, such as survey data and academic scores from the first two years of study) are collected along with the labels y' (represents the actual academic ranking after students graduate, serving as the ground truth labels for training and evaluating the model). This label is used in both the CGAN and Transformer components. In CGAN, it acts as a condition during data generation to ensure that the synthetic data aligns with specific academic ranking categories. In the Transformer model, it serves as the target variable for the classification task.

To generate additional synthetic data and expand the training dataset, we propose integrating CGAN into the model. Using the original data, we train the CGAN model, where the generator takes in noise vectors z (z is a random input that allows the Generator to create diverse synthetic data instead of repeating a single sample for each label) and labels y' to create synthetic data. The discriminator then distinguishes between real and synthetic data using the labels y' (y' is the label that helps the Generator create synthetic data with the correct label for each class; meanwhile, the Discriminator uses y' to verify whether the generated data matches the assigned label), allowing the generator to improve and produce data that closely resembles the real data.

After training the CGAN, the generator will be used to generate additional new data X_f (synthetic student-related data generated by CGAN, including simulated survey responses and academic scores from the first two years, corresponding to each y_f) and y_f. CGAN allows the generation of data based on specific classification labels.

This augmented dataset is subsequently used to train the Transformer model enhanced with Adaptive Layer Normalization (AdaLN), which dynamically adjusts normalization parameters across layers to reduce bias and variance. This adaptive mechanism improves model performance in predicting student graduation classification.

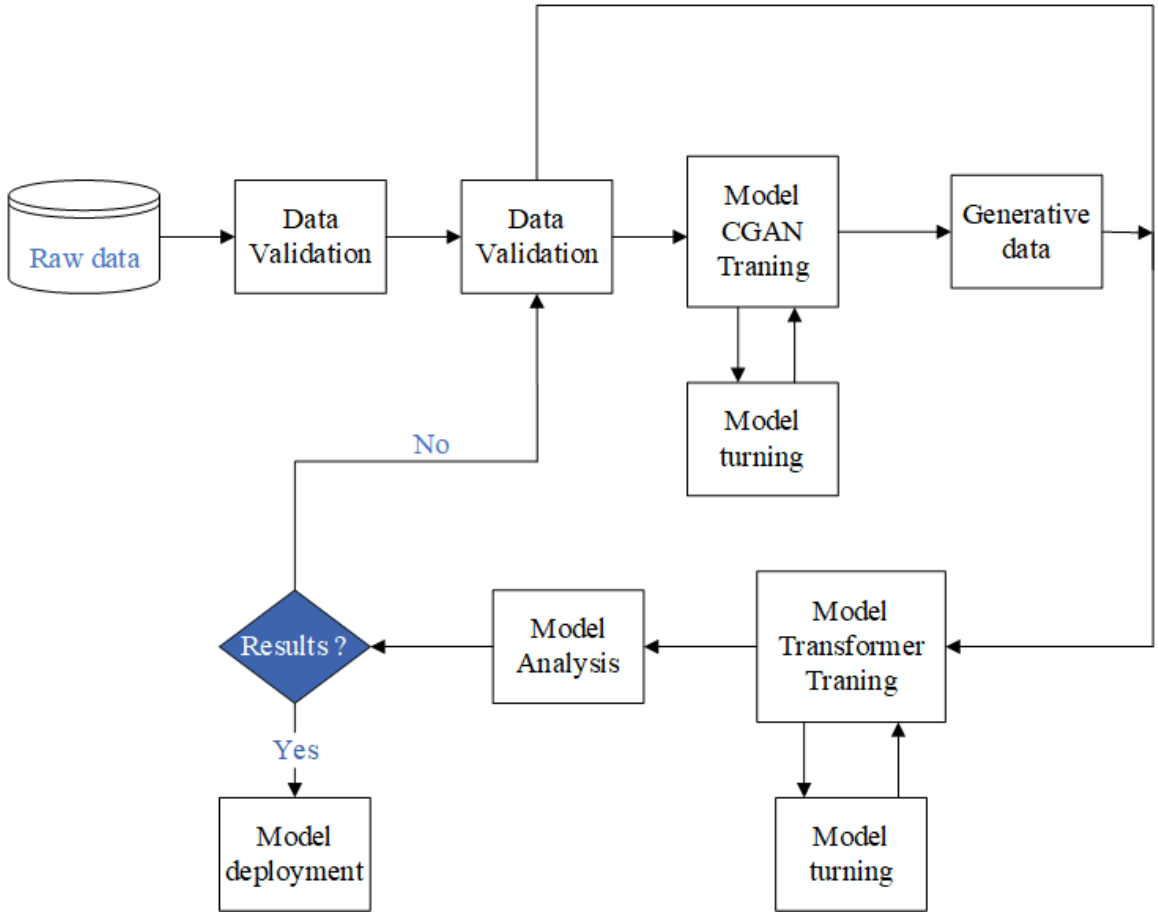


Figure 3. 2. The pineline of LATCGAd model

The proposed LATCGAd algorithm is given as follows.

Algorithm 3.1. LATCGAd - Learning Analysis with Transformer, CGAN, and Adaptive Layer Normalization

1: Input: D_{Real} : Real dataset of labeled student features and labels
2: Z : Latent noise vector for CGAN
3: G : Number of synthetic samples to generate
4: T_{AdaLN}: Transformer model with Adaptive Layer Normalization
5: Output: $\hat{Y}$: Predicted graduation classification labels for test set

6: $[X_r, y_r] \leftarrow$ Preprocess(D_{real})
7: $[G_{\text{CGAN}}, D_{\text{CGAN}}] \leftarrow Train_{CGAN}([X_r, y_r], Z)$
8: for $i = 1$ **to** G **do**
9: $z_i \leftarrow$ Sample(Z)
10: $y_i \leftarrow$ Sample_Label_Distribution(y_r)
11: $X_f[i] \leftarrow G_{CGAN}(z_i, y_i)$ ▷ Generate synthetic sample
12: $y_f[i] \leftarrow y_i$

13: end for

14: D_{aug}← Concatenate(**[**X_r, y_r], [X_f, y_f]) ▷ Augmented dataset

15: T_{AdaLN}← Train_Transformer(D_{aug})

16: $\hat{Y}$ ← Predict(T_{AdaLN}, X_{test})

17: return $\hat{Y}$

In this model, we utilize a generator with three hidden layers and a discriminator with four hidden layers. We apply the Adam optimizer, learning rate, and Beta_1.

For the Transformer model, we only use the Transformer Encoder. The final output will be a latent feature vector, which will then be passed through a fully connected layer for classification prediction. We use parameters such as multi-head attention, feed-forward layers, the number of Transformer encoder layers, the Adam optimizer, learning rate, and weight decay (see Subsection 3.2.3.2).

LATCGAd combines three key components: Transformer, CGAN, and AdaLN. The overall computational complexity is approximately $O(n \cdot d^2 + E \cdot (|G| + |D|))$, where n is the input sequence length, d the hidden dimension, and E the number of epochs.

The main bottlenecks lie in training the CGAN and Transformer, while AdaLN adds only minimal computational overhead.

Overall, the model achieves high predictive performance with a moderate computational cost compared to modern deep learning architectures.

3.2.3. Experiments

3.2.3.1. Datasets

To demonstrate the effectiveness of our proposed model, we will test it on these three datasets: HNMU1, HNMU2, and VNU. We will use all survey data and student scores from their first two years to predict student classification (see Section 1.4 Chapter 1).

Table 3. 2. Description of the training dataset

Dataset	**Number of samples**	**Survey-based attributes**	**Academic attributes** (first two academic years)	**Number of classes**
HNMU1	932	4	18	4
HNMU2	551	36	28	4
VNU	271	48	24	3

By using real data, we aim to show that our proposed model is effective in practical applications. The dataset is divided into train, validation, and test sets, with 60% of the data used for training, 15% for validation, and 25% for testing. All experiments were implemented on a workstation equipped with an Intel Core i7-12700KF CPU, NVIDIA RTX 3060 GPU, and 32GB RAM, ensuring reliable computational performance for deep learning training and evaluation.

3.2.3.2. ***Experimental setup***

CGAN Model: In this model, the Generator in the CGAN network consists of 3 layers to generate new data from latent space. Specifically, the first layer of the Generator has 256 neurons, the second layer has 512 neurons, and the third layer has 1024 neurons. The output of the Generator is 21 for HNMU1, 62 for HNMU2, and 72 for VNU. The activation function for HNMU1 and VNU is LeakyReLU with a coefficient of 0.2, and for HNMU2, it is ReLU. The Adam optimizer is used with a learning rate (lr) of 0.0002 and beta_1 of 0.5. The loss function is Binary Cross Entropy Loss, which helps the network learn nonlinear features effectively and avoid neuron death. The Discriminator also consists of 4 layers to evaluate the authenticity of the data generated by the Generator. Specifically, the first layer of the Discriminator has 1024 neurons, the second layer has 512 neurons, the third layer has 256 neurons, and the fourth layer has 64 neurons. The activation function for HNMU1 and VNU is LeakyReLU with a coefficient of 0.2, and for HNMU2, it is ReLU. The Adam optimizer is used with lr = 0.0002 and beta_1 = 0.5. The loss function is Binary Cross Entropy Loss. The output of the Critic is a single value representing the Discriminator's score for the input sample.

- For HNMU1, CGAN generates an additional 32 samples per class.
- For HNMU2, CGAN generates an additional 25 samples per class.
- For VNU, CGAN generates an additional 12 samples per class.

Table 3. 3. Number of samples before and after creation with CGAN

Datasets	**Labels**	**Medium**	**Good**	**Very Good**	**Excellent**	**Total**
HNMU1	Before generating	11	430	468	23	932
	After generating	43	462	500	55	1060

HNMU2	Before generating	19	337	191	4	551
	After generating	51	369	223	36	679
VNU	Before generating	0	46	187	38	271
	After generating	0	58	199	50	307

As shown in Table 3.4, the initial datasets suffer from significant class imbalance, with the *Medium* and *Excellent* categories represented by very few samples (for instance, only 4 *Excellent* cases in HNMU2 and none in the *Medium* category for VNU), while the *Good* and *Very Good* classes are predominant. This uneven distribution can lead the model to concentrate on the majority classes and neglect patterns associated with minority ones. After applying CGAN, the number of samples in the smaller classes increased significantly, such as *Excellent* in HNMU2 rising from 4 to 36, and *Medium* in HNMU1 from 11 to 43, leading to a more balanced distribution and improved model learning. However, some limitations remain, for instance, VNU still has no samples in the *Medium* class, and the generated data may not fully represent real data.

The parameters of the CGAN model (such as the number of layers, learning rate, and activation functions) are tailored for each dataset to ensure that the synthetic data closely resembles the real data. For HNMU1 and VNU, the LeakyReLU activation function is used in both the generator and discriminator, whereas ReLU is more effective for HNMU2. The number of synthetic data samples for each class is also adjusted differently for each dataset to ensure class balance without introducing excessive noise.

Table 3. 4. Generator model parameters on the HNMU1, HNMU2, and VNU datasets

Datasets	**First layer**	**Second layer**	**Third layer**	**Activation function**	**Output layer**	**Output activation function**
HNMU1	256	512	1024	LeakyReLU(0.2)	21	Tanh
HNMU2	256	512	1024	ReLU	62	Tanh
VNU	256	512	1024	LeakyReLU(0.2)	72	Tanh

Table 3. 5. Discriminator model parameters

Datasets	First layer	Second layer	Third layer	Fourth Layer	Activation function	Output activation function
HNMU1	1024	512	256	64	LeakyReLU(0.2)	Sigmoid
HNMU2	1024	512	256	64	ReLU	Sigmoid
VNU	256	512	256	64	LeakyReLU(0.2)	Sigmoid

Training parameters for the Generator and Discriminator models on the datasets as follows: Optimizer is Adam, Learning Rate=0.0002, Beta_1=0.5 and Loss function: Binary Cross Entropy Loss.

Transformer model

- For HNMU1, the Transformer uses 2 multi-heads. The feed-forward layer in each encoder layer has 64 units. The number of Transformer encoder layers is 1, with dropout = 0.6. This is followed by a fully connected network with an output of 4 (corresponding to the number of classes in the HNMU1 dataset).
- For HNMU2, the Transformer uses 7 multi-heads. The feed-forward layer in each encoder layer has 64 units. The number of Transformer encoder layers is 2, with dropout = 0.5. This is followed by a fully connected network with an output of 4 (corresponding to the number of classes in the HNMU2 dataset).
- For VNU, the Transformer uses 2 multi-heads. The feed-forward layer in each encoder layer has 128 units. The number of Transformer encoder layers is 1, with dropout = 0.6. This is followed by a fully connected network with an output of 3 (corresponding to the number of classes in the VNU dataset).

All models use the Adam optimizer with lr = 0.005 and weight decay = 0.0005.

The number of attention heads is selected based on the complexity of the feature space. For HNMU1 and VNU, two multi-heads are appropriate. The HNMU2 requires seven multi-heads to learn complex patterns from a larger feature space.

Feed-Forward Layer: For HNMU1 and HNMU2, the feed-forward layers have 64 units, whereas VNU requires 128 units due to the higher number of variables in the dataset. Learning Rate and Optimizer: The learning rate is set to 0.005 with the Adam optimizer after experimenting with different values and observing the convergence speed and accuracy.

During the model training process, several important parameters were utilized to optimize and adjust the model's learning capability, including Beta_1, learning rate (lr), and Dropout. A Beta_1 value of 0.5 was chosen to balance convergence speed and learning stability. The learning rate (lr) controls the adjustment speed of the model’s weights after each gradient update. This value was set differently for each model: 0.0002 for CGAN and 0.005 for

Transformer, ensuring optimal convergence speed while preventing oscillations or slow convergence. Additionally, Dropout was used as a regularization technique to mitigate overfitting by randomly dropping some neurons during training. The Dropout value was set at either 0.5 or 0.6, depending on the model and dataset, to enhance generalization and model robustness when applied to real data.

For LATCGAd model, the number of epochs for training the CGAN model for the three datasets HNMU1, HNMU2, and VNU is 5000 epochs. The training graphs of the models are shown respectively in Figure 3.3a), 3.3b), and 3.3c). The principle for model selection is that the CGAN model with the smallest FID value.

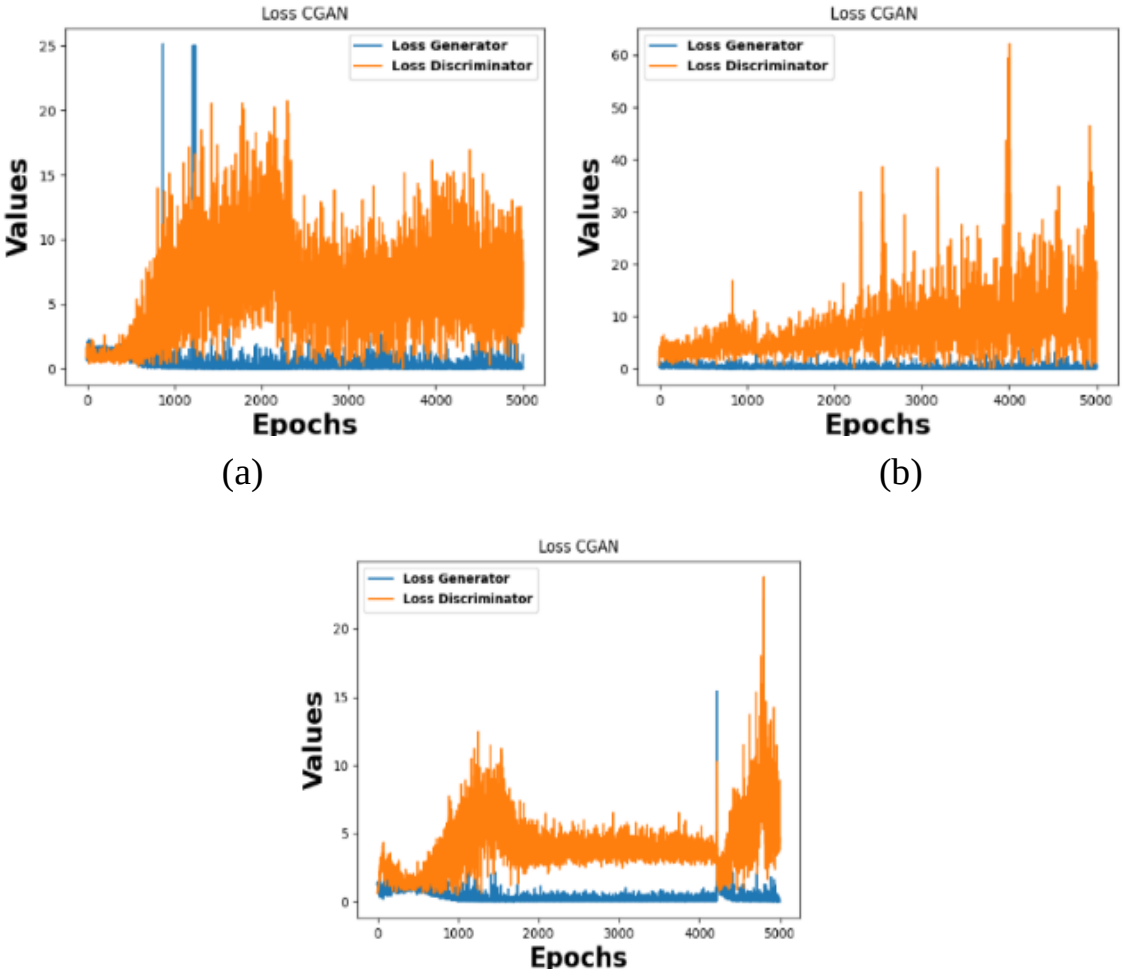


Figure 3. 3. Training the CGAN model (in the LATCGAd model)

a) On the HNMU1 dataset. b) On the HNMU2 dataset. c) On the VNU dataset.

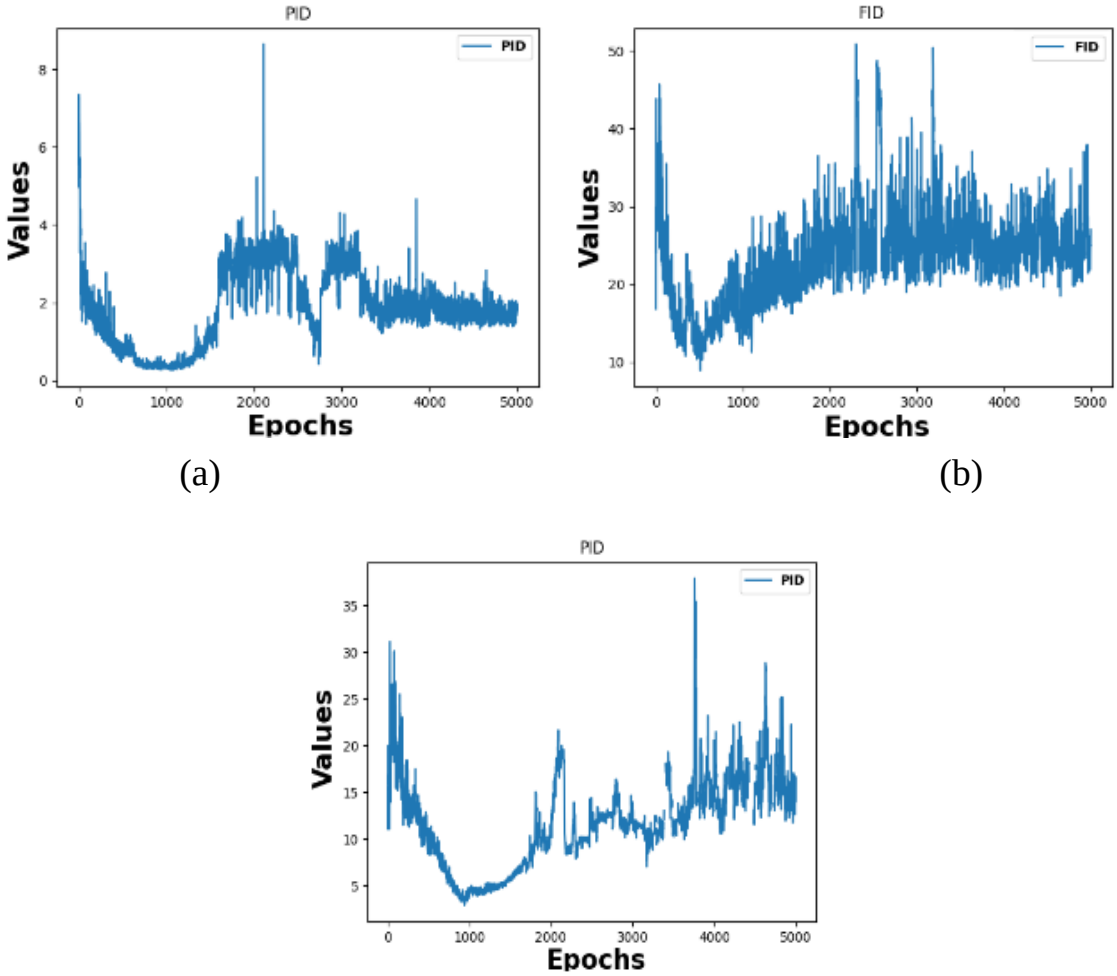


Figure 3. 4. FID values

a) On the HNMU1 dataset. b) On the HNMU2 dataset. c) On the VNU dataset.

We will then train the Transformer model for 1000 epochs for the three datasets HNMU1, HNMU2, and VNU. The training graphs for the models for the three datasets HNMU1, HNMU2, and VNU are shown respectively in Figure 3.5a), 3.5b), and 3.5c). The principle for selecting the best model is to take the average of the training loss and validation loss, and the epoch with the smallest value will be chosen. Based on this principle, for the model in Figure 3.5a), the model selected at epoch 71 has a train loss of 0.2677 and a validation loss of 0.1237. For the model in Figure 3.5b), the model selected at epoch 962 has a train loss of 0.0361 and a validation loss of 0.0018. For the model in Figure 3.5c), the model selected at epoch 61 has a train loss of 0.3878 and a validation loss of 0.2793.

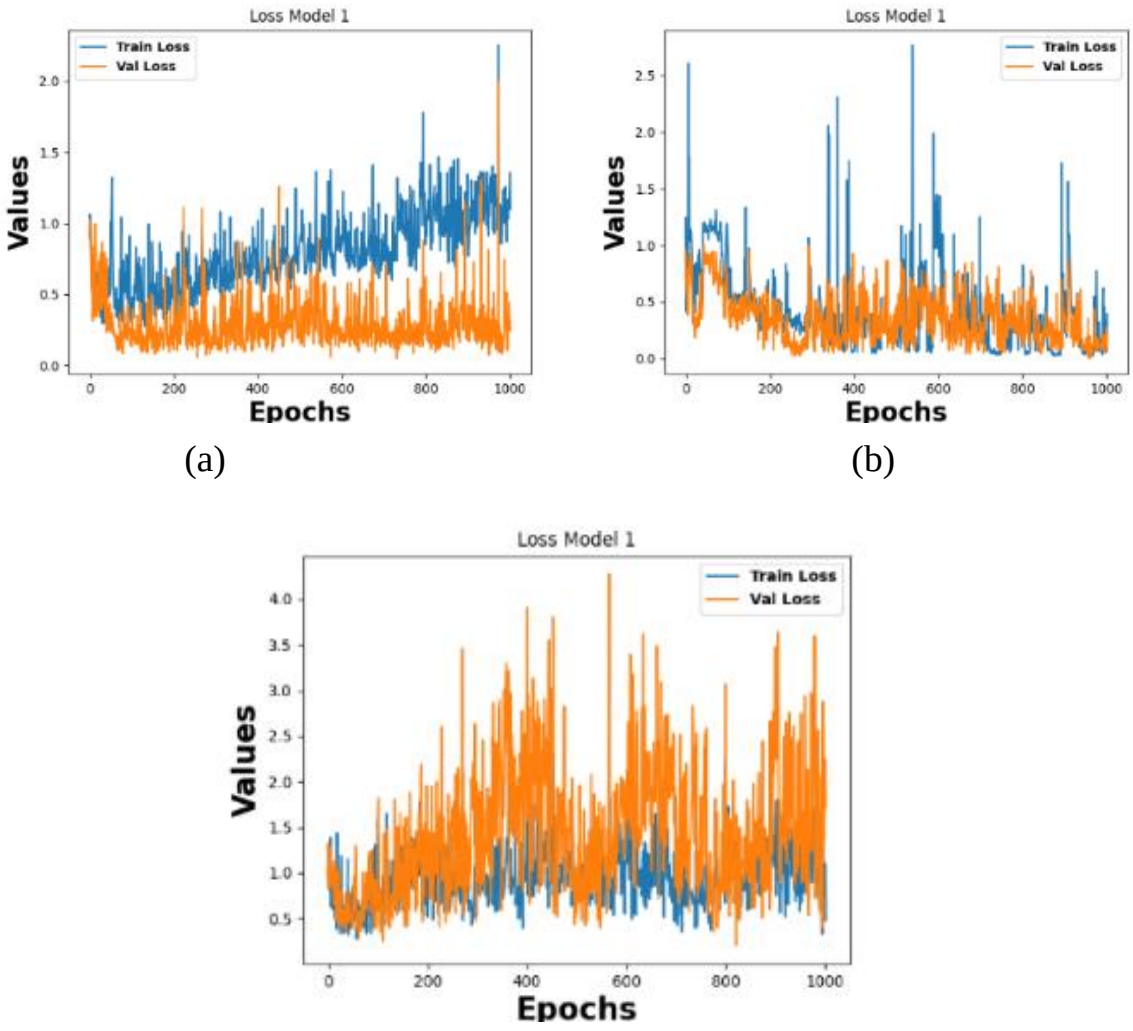


Figure 3. 5. Training the Transformer model (in the LATCGAd model)
a) On the HNMU1 dataset. b) On the HNMU2 dataset. c) On the VNU dataset.

We will compare the proposed model with three different deep learning algorithms (DNN, GAT and Transformer), and traditional machine learning methods (which are known to perform well with small datasets): DT, SVM and LR.

We trained the DNN, GAT and Transformer models with the three datasets HNMU1, HNMU2, and VNU, training the model for 1000 epochs. The principle for selecting the best model is that model with the lowest average of train loss and validation loss will be chosen.

3.2.4. Results and discussion

Experimental results on the three datasets (HNMU1, HNMU2, and VNU) show that the LATCGAd model outperforms traditional models (DT, SVM, LR) and deep learning models (DNN, GAT, and standard Transformer).

Table 3. 6. Prediction results on the HNMU1 dataset

Method	Accuracy	Precision	Recall	F1-Score
DT	89.64	39.77	48.09	42.90
SVM	84.64	35.95	45.31	38.77
LR	92.86	43.13	49.05	45.71
DNN	93.57	69.07	74.15	71.35
GAT	82.14	34.81	44.57	37.46
Transformer	93.57	44.71	47.96	46.26
LATCGAd	**95.56**	**72.50**	**74.78**	**73.61**

On the HNMU1 dataset, LATCGAd achieves an accuracy of 95.56%, significantly higher than DT (89.64%), SVM (84.64%), LR (92.86%), DNN (93.57%), and GAT (82.14%). In addition to accuracy, the model also improves Precision (72.50%), Recall (74.78%), and F1-score (73.61%), demon strating its ability to reduce errors and correctly classify almost all true positive samples, outperforming all compared models.

On the HNMU2 dataset, LATCGAd achieves the highest accuracy (96.97%), outperforming the standard Transformer (95.62%), DT (89.70%), and GAT (89.05%). However, while it maintains relatively balanced Precision (73.26%) and Recall (74.09%), these values are still noticeably lower than the high Precision of DT (94.65%) and its Recall (79.26%). This indicates that despite its stability and generalization ability, thanks to synthetic data generated by CGAN, LATCGAd lacks sharpness in accurately and comprehensively identifying the target class. This limitation may impact its effectiveness in real-world scenarios that require high discriminative performance (see Table 3.8). To address this, future work should focus on optimizing the CGAN-based data generation process to produce samples closer to real distributions and leveraging multi-level attention or graph-based features to better capture complex relationships within student data.

Table 3. 7. Prediction results on the HNMU2 dataset

Method	Accuracy	Precision	Recall	F1-Score
DT	89.70	94.65	79.26	82.48
SVM	80.29	41.38	41.81	40.55
LR	71.74	64.57	62.25	60.46
DNN	87.05	69.32	60.92	63.75

GAT	89.05	53.52	57.95	55.16
Transformer	95.62	72.77	60.99	64.79
LATCGAd	**96.97**	**73.26**	**74.09**	**73.66**

On the VNU dataset, LATCGAd achieves an accuracy of 87.65%, lightly outperforming DT (83.95%), and standard Transformer (86.76%). A key advantage is that Precision increases to 95.56%, significantly surpassing the other models, indicating its high reliability in predicting positive cases and minimizing false positives. However, Recall is 58.73%, slightly lower than Transformer (71.73%). This trade-off is justified by its optimized Precision, making it suitable for applications requiring high confidence in identifying critical cases. The F1-score of LATCGAd on the VNU dataset reaches 67.62%, surpassing most machine learning models, though slightly lower than standard Transformer (70.72%).

Table 3. 8. Prediction results on the VNU dataset

Method	**Accuracy**	**Precision**	**Recall**	**F1-Score**
DT	83.95	67.59	55.08	59.06
SVM	83.82	42.43	53.83	46.97
LR	76.47	70.88	74.06	59.19
DNN	75.36	67.26	53.39	55.44
GAT	80.88	51.60	50.52	51.00
Transformer	86.76	69.72	71.73	70.72
LATCGAd	**87.65**	**95.56**	**58.73**	**67.62**

A key factor influencing the experimental results is the differences in characteristics among the three datasets: HNMU1, HNMU2, and VNU. Each dataset varies in size, number of features, and class imbalance levels, which directly impact the models' performance.

Specifically, the HNMU1 dataset has the largest sample size (932 samples) but a limited number of features (21 features, with only 3 survey-based features). As a result, the model primarily relies on academic scores from the first two years. While this enables the model to quickly identify learning trends, it also increases the risk of missing additional insights from non-academic factors. By balancing the data with CGAN, LATCGAd significantly improves accuracy and F1-score compared to other models.

The HNMU2 dataset is smaller (551 samples) but contains 62 features (including 34 survey-based features), providing a more comprehensive view of students. This allows LATCGAd to capture multidimensional relationships between academic and non-academic data. As a result, HNMU2 achieves the highest accuracy (96.97%), while maintaining a balance between Precision and Recall, demonstrating that rich and diverse data plays a crucial role in enhancing model performance. The integration of AdaLN further improves stability and mitigates overfitting, enabling the model to perform consistently across diverse educational datasets.

In contrast, the VNU dataset has the smallest sample size (271 samples) but includes 72 features. Despite the dataset's high feature richness, its small size makes the model more susceptible to overfitting. The high Precision (95.56%) indicates that LATCGAd is highly effective in reducing false positives, but the low Recall (58.73%) suggests that some true positive samples were missed, likely due to the limited training data.

These confusion matrices clearly present the experimental results.

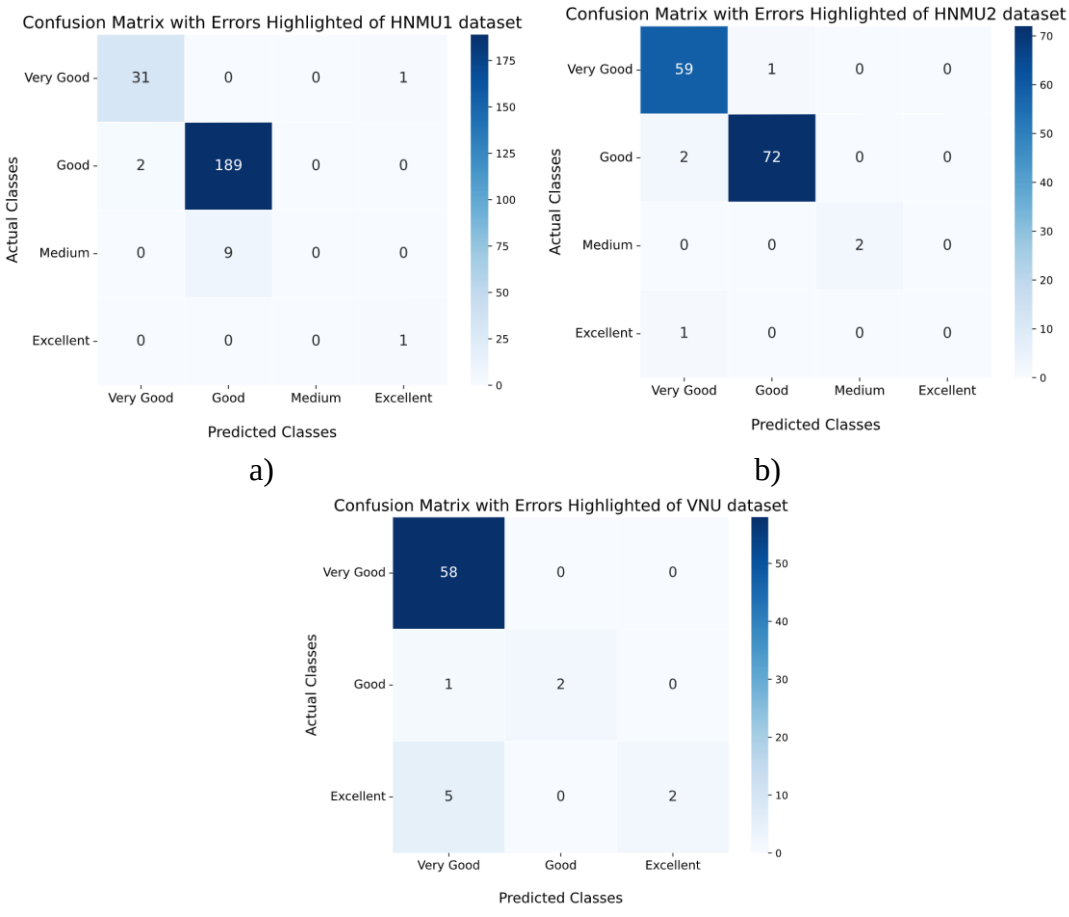


Figure 3. 6. Confusion Matrices (in the LATCGAd model) a) on the HNMU1 dataset. b) on the HNMU2 dataset. c) on the VNU dataset.

As illustrated in Figure 3.6, the confusion matrices reveal that most misclassifications occur between the *Good* and *Very Good* categories. This problem is largely due to class imbalance. In the HNMU2 dataset, which contains four performance categories (*Excellent*, *Very Good*, *Good*, and *Medium*), the distribution is dominated by the *Very Good* (337 samples, 61%) and *Good* (191 samples, 35%) classes. The overwhelming proportion of these two categories makes it challenging for the model to establish a clear separation between them. In the HNMU1 dataset, the imbalance issue also persists, where *Medium* students are sometimes misclassified as *Good.*

The differences in size and composition across these datasets highlight the necessity of data balancing and augmentation using CGAN, particularly when working with small or highly imbalanced datasets. Additionally, this underscores the importance of feature selection and analysis in optimizing deep learning model performance.

In summary, the LATCGAd model demonstrates superior accuracy and overall performance across all three datasets, particularly under conditions of small and imbalanced data. This improvement is attributable to the synergistic integration of data augmentation via CGAN and model optimization through AdaLN. Compared to the LAGT model presented in [CT2], LATCGAd effectively addresses the severe class imbalance observed in the HNMU2 dataset by employing CGAN to generate additional samples for each class, especially for underrepresented classes. This targeted data augmentation leads to a substantial enhancement in predictive performance.

However, on the VNU dataset, where class distribution is relatively balanced but the ratio of samples (271) to features (72) is disproportionate, the LATCGAd model performs less effectively. This indicates that the model is sensitive to situations where the number of samples is too small compared to the feature dimensionality, even when class proportions are not an issue.

Consequently, there arises a clear need to improve predictive performance in contexts that require additional feature selection and extraction. The subsequent model introduced in this work, AWG-GC, is designed to address this challenge effectively.

3.3. The AWG-GC model

3.3.1. The theoretical basis for model selection

Despite the strong performance of LATCGAd in predicting academic performance from small and imbalanced educational datasets, several technical

challenges remain unresolved. These limitations reveal the need for a more comprehensive approach to educational data modeling.

One key limitation lies in feature representation. Educational data is often noisy, sparse, and inconsistently structured, which reduces the ability of models like Transformers to extract meaningful patterns. Moreover, the reliance on labeled data presents difficulties when annotations are limited. This motivates the integration of an Autoencoder module, which can learn compressed, denoised representations in an unsupervised manner.

A second challenge is the quality and diversity of synthetic data used to balance training sets. LATCGAd uses CGAN to generate additional samples, but this approach suffers from unstable training and limited control over sample quality. From a theoretical perspective, WGAN improves the stability of GAN-based training, including CGAN, and helps avoid the mode collapse problem. Furthermore, the synthetic data generated by WGAN is generally of higher quality and closer to the real data distribution compared to that generated by CGAN. This makes WGAN a more suitable choice for generating reliable and diverse educational data.

Finally, educational data often contains rich relational structures,such as dependencies among courses, learning sequences, and behavioral interactions, that are naturally represented as graphs. Transformer-based models, including those used in LATCGAd, do not fully exploit these relationships. Prior work has shown that incorporating graph-based architectures can significantly enhance learning from such data.

To address these challenges, we propose AWG-GC, a hybrid deep learning model that combines Autoencoder, WGAN, and Graphormer. This architecture is designed to enhance feature learning, improve data generation, and capture complex relationships within educational data, thereby offering a more robust and effective solution for academic performance prediction.

Graphformer is a particularly promising candidate, as it fully integrates both Transformer mechanisms and graph-specific features, making it well-suited for educational datasets with rich relational structures. With multi-head self-attention and positional encoding mechanisms, Graphormer can effectively model long-range and hierarchical relationships, overcoming limitations such as over-smoothing found in traditional GNNs like GCN and GAT.

In summary, AWG-GC is a systematic extension of LATCGAd. It not only inherits the strengths of data generation and training optimization but also introduces critical components to fully address real challenges: complex feature

handling, limited labeled data, and multidimensional relationship modeling in educational data. As a result, the model provides an efficient hybrid deep learning framework with strong potential to support intelligent decision-making and enhance the quality of learning analytics.

3.3.2. Proposed model

This section presents the mapping $f: X \rightarrow Y$ in the form of the proposed AWG-GC model, which accurately predict the graduation classification $y \in Y$ for each student based on features $x \in X$. Figure 3.7 illustrates the AWG-GC model, which integrates an Autoencoder, WGAN, and Graphormer for graduation classification.

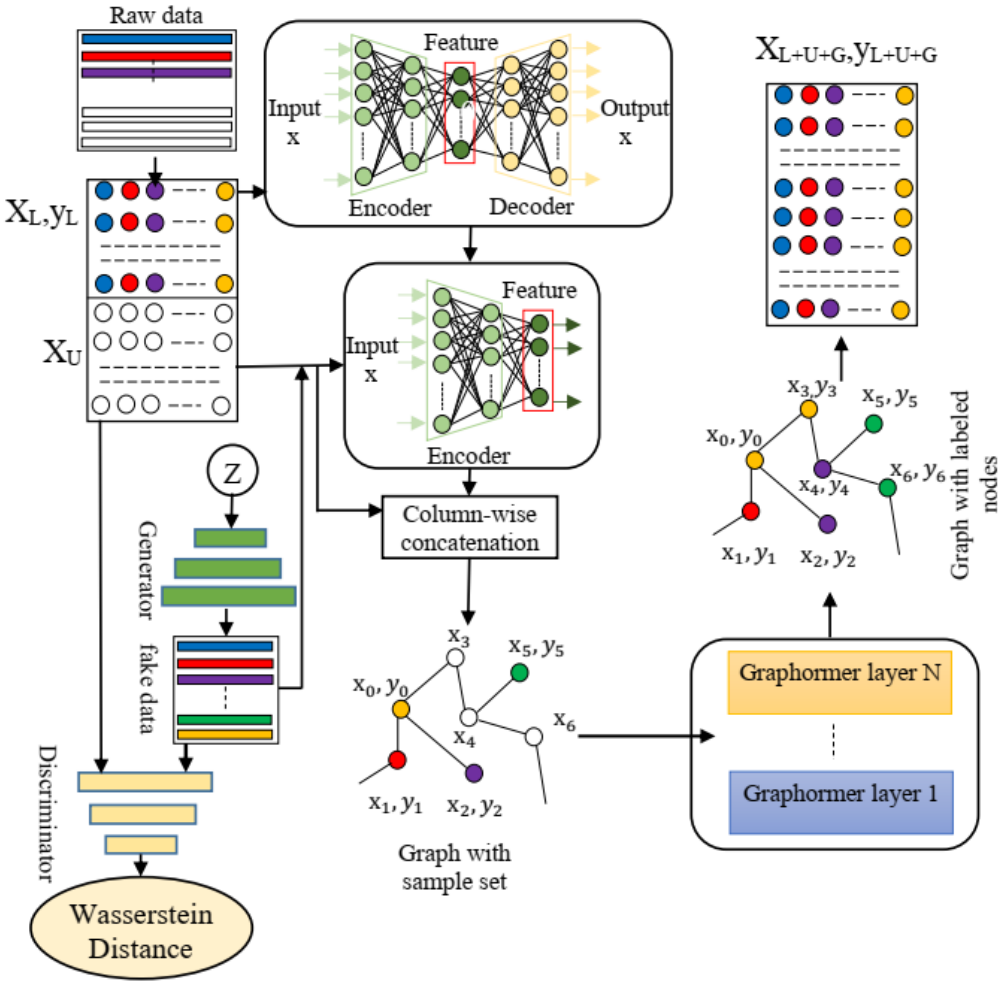


Figure 3. 7. The AWG-GC model ([CT8])

The implementation of the model, as illustrated in Figure 3.7, is carried out as follows:

After preprocessing, the raw data forms an initial sample set consisting of $(L + U)$ samples, where (X_L, y_L) are labeled samples and X_U are unlabeled

samples. Note that each sample in this set has a dimensionality of n. This dataset is used to train a deep Autoencoder neural network to learn the latent space representation. At the same time, the $(L+U)$ sample set is also used to train a Wasserstein Generative Adversarial Network (WGAN) to generate an additional synthetic sample set, X_G, consisting of G new samples.

After the sample generation process is completed, the dataset is expanded to $L+U+G$ samples, maintaining the same dimensionality of n. This expanded sample set is then fed into the encoder part of the Autoencoder to extract features and reduce the data dimension from n to m. Thus, each sample in the $L+U+G$ set has two representations: one in the original n-dimensional space and one in the m-dimensional feature space.

To construct the input graph for the Graphormer model, these two representations are combined by column concatenation, creating a dataset with a dimensionality of $n+m$. The neighborhood graph of the samples in the $(L+U+G)$ set is built using the KNN algorithm based on this combined feature space. The resulting graph is then fed into the Graphormer model, a Transformer-based variant designed to handle graph-structured data. Thanks to the global attention mechanism weighted by graph distance, Graphormer can efficiently learn the relationships between nodes, thus improving the accuracy in predicting students' graduation classifications.

In our model, the KNN algorithm is utilized in two different contexts. First, KNN is employed to construct the input graph structure for the Graphormer model. Specifically, after the raw data is passed through an Autoencoder to obtain compressed representations, we perform a column-wise concatenation of the original and compressed features to form a combined feature space. It is within this space that KNN is applied to identify the nearest neighboring nodes for building the input graph. Therefore, dimensionality reduction via the Autoencoder is carried out prior to graph construction using KNN. Second, KNN is also used as a baseline classification method in the experimental comparison. In this case, KNN is applied directly to the original feature space, without any dimensionality reduction or data augmentation. This allows us to assess the extent of improvement achieved by the proposed AWG-GC deep learning framework.

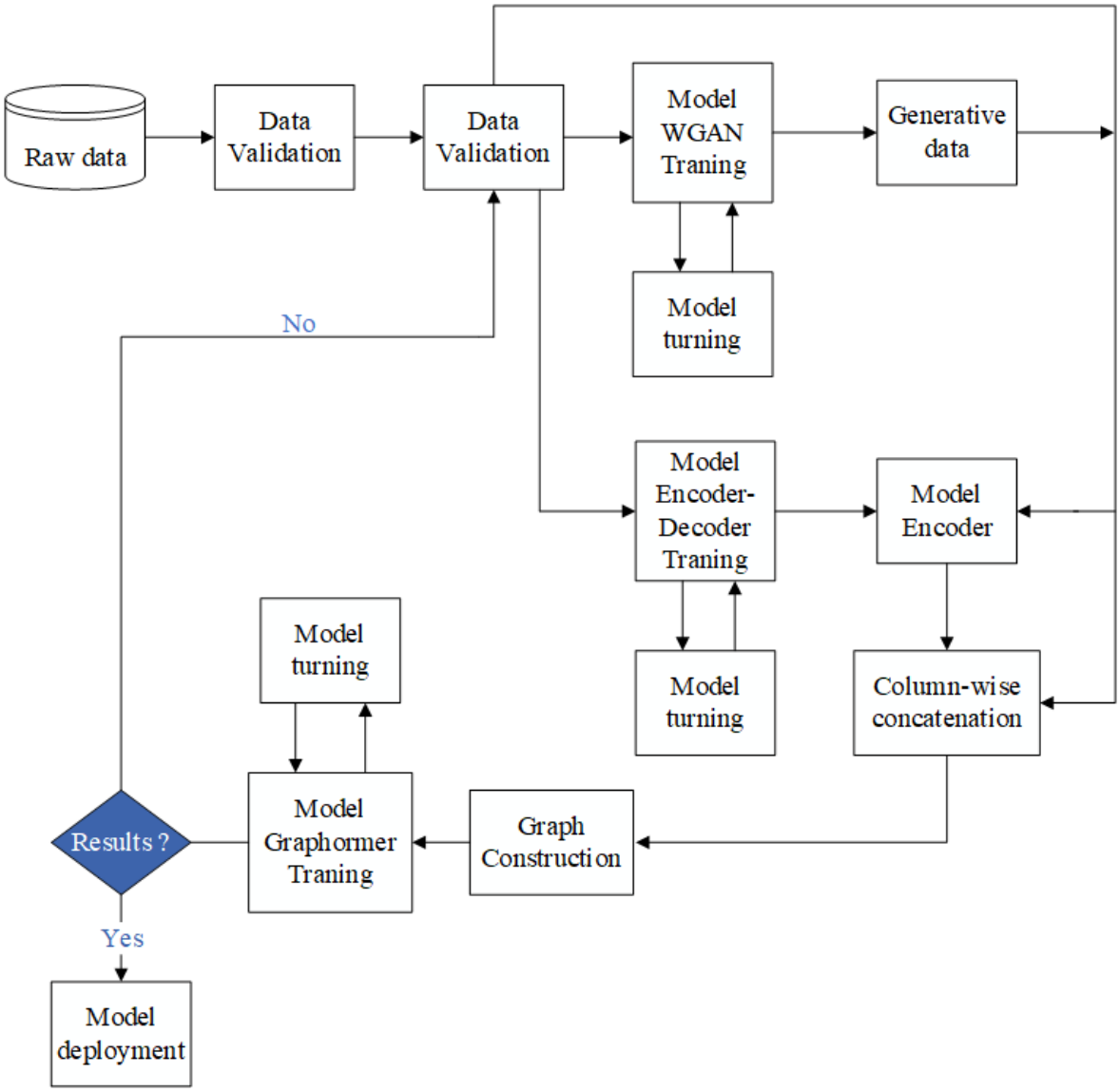


Figure 3. 8. The pineline of AWG-GC model

The AWG-GC model offers the following key benefits: (1) Leverages the Autoencoder network to extract features and reduce data dimensionality, improving training efficiency; (2) Uses the WGAN network to augment the dataset, enhancing the model's generalization ability; (3) Combines both original and extracted representations to construct the graph, thereby enhancing the performance of Graphormer in classifying students.

It is important to note that in the AWG-GC model, Autoencoder, WGAN, and Graphormer do not operate independently but work together in a unified process. The Autoencoder and WGAN help create a richer dataset, while Graphormer utilizes this dataset to improve the model's prediction accuracy.

The proposed AWG-GC algorithm will be given as follows.

Algorithm 3.2: AWG-GC – Integrating an Autoencoder, Wasserstein GAN, and Graphormer for Graduation Classification

1: Input: $\mathcal{D}_{\mathrm{L}}$ **:** Labeled dataset of student features and labels
2: $\mathcal{D}_{\mathrm{U}}$: Unlabeled dataset of student features
3: m : Number of samples in $\mathcal{D}_{\mathrm{U}}$
4: n : Number of samples in $\mathcal{D}_{\mathrm{L}}$
5: z : Latent feature dimension from Autoencoder
6: s : Number of synthetic samples generated by WGAN
7: Output: $\hat{\mathrm{Y}}$ **:** Predicted graduation classification labels

8: $[X_L, y_L] \leftarrow$ Preprocess($\mathcal{D}_L$)
9: $X_{\mathrm{U}} \leftarrow$ Preprocess($\mathcal{D}_{\mathrm{U}}$)
10: Train Autoencoder on $X_L \cup X_U$
11: $Z_{\mathrm{L}} \leftarrow$ Encode(X_L), $Z_{\mathrm{U}} \leftarrow$ Encode(X_{U})
12: $[\mathcal{G}, C] \leftarrow$ TrainWGAN(X_L, y_L)
13: **for** $i = 1$ **to** s **do**
14: $z_i \leftarrow$ SampleNoise()
15: $y_i \leftarrow$ SampleLabel(y_L)
16: $x_i^{gen} \leftarrow \mathcal{G}(z_i, y_i)$
17: $\mathcal{D}_S \leftarrow \mathcal{D}_S \cup (x_i^{gen}, y_i)$
18: **end for**
19: $\mathcal{D}_{all} \leftarrow \mathcal{D}_L \cup \mathcal{D}_U \cup \mathcal{D}_S$
20: $Z_{all} \leftarrow$ Encode($\mathcal{D}_{all}$)
21: $F_{combined} \leftarrow$ Concatenate(X_{all}, Z_{all})
22: $G_{knn} \leftarrow$ ConstructGraph($F_{combined}$)
23: Train Graphormer on (G_{knn}, y_L)
24: $\hat{\mathrm{Y}} \leftarrow$ Predict(Graphormer, X_{test})
25: **return** $\hat{\mathrm{Y}}$

AWG-GC combines three main components: an Autoencoder for dimensionality reduction, a WGAN for data augmentation, and a Graphormer for classification. Its computational complexity is approximately $O(E \cdot |AE| + n \cdot N^2 \cdot d^2 + |E_g|)$, where $|AE|$ denotes the computational cost per training epoch of the Autoencoder, E is the number of epochs, N the number of student nodes, and $|E_g|$ the number of graph edges.

The model is more computationally demanding than LATCGAd due to WGAN's gradient penalty and Graphormer's graph-aware attention. However, this higher complexity is justified by its improved accuracy and robustness in modeling inter-student relationships within educational graphs.

3.3.3 Experiments

3.3.3.1. Training dataset

In this section, we use three real datasets, HNMU2, VNU, and SATDAP, to evaluate the performance of the proposed model. The HNMU2 and VNU datasets contain information on students' grades and survey feedback from two major universities in Vietnam. Each dataset differs in scale, feature distribution, and classification labels, creating a diverse testing environment for our approach. The inclusion of the SATDAP dataset, collected in Portugal, serves to enhance the generalizability of our proposed method by incorporating data from a distinct international educational context.

The SATDAP dataset originates from the SATDAP program, Capacitação da Administração Pública, under the authorization of POCI-05-5762-FSE-000191, Portugal ([61]). This dataset consists of 4424 records and 36 features. It includes information on students' academic trajectories, demographic data, socio-economic factors, and SGPA over a five-year period.

The dataset contains variables related to demographic factors (such as age at enrollment, gender, marital status, nationality, postal code, special needs), socio-economic factors (such as whether the student works, parents' educational background, parents' occupations, parents' employment status, student scholarships, and tuition debt), and educational pathways (such as entrance exam score, number of years repeated in secondary school, program preference order, and type of secondary school course).

The academic information in this dataset is limited to observable factors prior to university enrollment, excluding any internal assessments after enrollment. Each student record is categorized into one of three groups: Success, Relative Success, or Failure, based on the time taken to complete the degree program. "Success" or "Graduate" refers to students who complete the program within the standard timeframe. "Relative Success" or "Enrolled" refers to those who graduate after up to three additional years. "Failure" or "Dropout" applies to students who take more than three additional years to graduate or do not graduate at all.

This classification reflects three levels of risk: Low-risk students are highly likely to succeed; Medium-risk students may benefit from institutional interventions to support their success, and High-risk students are those with a high likelihood of failure.

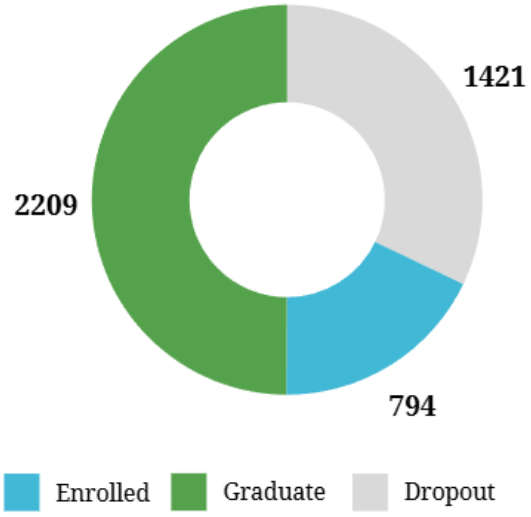


Figure 3. 9. Number of samples per class in the SATDAP dataset

The data underwent feature normalization using the StandardScaler() method, which adjusts features to have a mean of 0 and a standard deviation of 1. This ensures that all features share the same scale and units, thereby improving the performance of machine learning models. The formula for StandardScaler is:

$$x_{new} = \frac{x - x_{mean}}{x_{std}}. \quad (3.6)$$

For the labels, they were converted from text to numerical format using the LabelEncoder() function. This function encodes each unique label value as an integer, enabling machine learning models to process the labels numerically instead of as text strings.

To evaluate the efficacy of the AWG-GC model, this section focuses on conducting experiments on three distinct datasets: the HNMU2 dataset, the VNU dataset, and the SATDAP dataset. By evaluating the proposed model on these real datasets, we provide clear evidence of its effectiveness. We will sequentially extract train, validate, and test datasets from the data with the purpose of testing on the most recent graduate student data, enabling the best evaluation of the model's practical applicability when using past student data.

The datasets are divided into train, validation, and test sets, with 60% of the data used for training, 15% for validation, and 25% for testing.

To demonstrate the effectiveness of the proposed method, we divided it into scenarios as follows:

- SVM, KNN, RF, GAT, Transformer, Graphormer: using the original dataset allows us to evaluate the predictive capability of each model with the initial data.
- AutoGAT: applying dimensionality reduction using an Autoencoder before training the GAT model helps us understand the impact of the Autoencoder on predictions, particularly with datasets containing numerous fields and complex structures,
- AWG-GAT: combining dimensionality reduction via an Autoencoder with data augmentation using WGAN before training the GAT model demonstrates how WGAN-generated data can address the issue of small dataset sizes,
- AWG-GC: combining dimensionality reduction via an Autoencoder with data augmentation using WGAN before training the Graphormer model.

3.3.3.2. Set up of model parameters

Experiments were run on a workstation with Intel Core i7-12700KF, NVIDIA RTX 3060, and 32GB RAM, offering adequate resources for deep learning training and evaluation.

To ensure that the selected hyperparameters were both optimal and stable, we conducted a sensitivity analysis by varying key hyperparameters such as the number of neurons per layer, learning rate, dropout rate, and the number of attention heads within reasonable ranges based on previous studies. The learning rate was tested with values {0.0001, 0.001, 0.005, 0.01}, while the dropout rate was evaluated within the range {0.3, 0.5, 0.6, 0.7}. Experimental results showed that the model's performance remained stable across these configurations. The final hyperparameters were selected based on minimizing validation error and maximizing the F1-score, ensuring a balance between training effectiveness and generalization capability across datasets of varying sizes and structural characteristics.

Autoencoder

The Autoencoder network is structured with two main parts: an encoder and a decoder. The encoder comprises two stages. The first layer of the encoder has 128 neurons and is responsible for reducing the input data to a smaller space. Next, the second layer has 64 neurons and compresses the data into a hidden space. The hidden space of this Autoencoder network is flexibly structured with different sizes depending on each dataset: 10 neurons in the HNMU2 set, 10 neurons in the VNU set and 10 neurons in the SATDAP.

The Decoder part of the network also includes two layers, but it operates oppositely to the encoder part. The first layer of the decoder contains 64 neurons to expand data from the hidden space. Finally, the second layer has 128 neurons, which completes the process of reconstructing the data to their original form or close to the input data. This structure helps the Autoencoder network learn and compress information effectively and is capable of reconstructing data from hidden spaces with high accuracy.

All layers in the Autoencoder and the decoder use the ReLU activation function and the Adam optimizer with a learning rate equal to 0.005 and weight decay equal to 0.0005.

WGAN model

This WGAN is structured to include two main components: Generators and Critics. The Generator set of the WGAN network includes 3 layers for the HNMU2, VNU and SATDAP datasets, each with increasing sizes to generate new data from the hidden space. With the HNMU2, VNU and SATDAP datasets, the first layer of the Generator has 256 neurons, the second layer has 512 neurons, and the third layer has 1024 neurons. The generator output contains 62 features for the HNMU2 set, 72 features for the VNU set, and 36 features for the SATDAP set. All layers in the Generator use the LeakyReLU activation function with a coefficient of 0.2, which helps the network learn nonlinear features effectively and avoids neuronal death. We use the Adam optimizer with a learning rate equal to 0.0002 and betas equals (0.5, 0.9).

The Critic part of the WGAN network also includes 3 layers for the HNMU2, VNU and SATDAP datasets, but with decreasing size, to help evaluate the authenticity of the data generated by the Generator. Specifically, with Critic’s the HNMU2, VNU and SATDAP datasets, Critic's first layer had

512 neurons, the second layer had 256 neurons, and the third layer had 64 neurons. All layers in the Critic use the LeakyReLU activation function with a coefficient of 0.2. We use the Adam optimizer with a learning rate equal to 0.0002 and betas equals (0.5, 0.9). The Critic output is a single value representing the Wasserstein score of the input data sample. This structure helps the WGAN network achieve a balance between generating new data samples and evaluating the authenticity of the samples, which improves the quality of the generated data.

After training the model, we used the Generator to generate additional training data for GAT. For the HNMU2 dataset, we generated 100 samples per class, totaling 400 samples, to add to the training set. For the VNU dataset, we generated 64 samples per class, totaling 192 samples, to add to the training set. For the SATDAP dataset, we generated 1032 samples per class, totaling 3096 samples, to add to the training set The number of samples generated and added to each dataset is as follows: (1) HNMU2 dataset: 731 training samples and 137 testing samples; (2) VNU dataset: 355 training samples and 68 testing samples; and (3) SATDAP dataset: 6104 training samples and 855 testing samples.

Graphormer model

For HNMU2, the Graphomer model selects d_model = 64 to define the dimensionality of the hidden representations, ensuring a compact yet expressive embedding space for node features and facilitating efficient attention computations. We also configure max_distance = 12, which limits the hop distance used for spatial bias embeddings, effectively controlling the range of node interactions and reducing noise from overly distant connections. Attention dropout rate of 0.1, a multi-head attention value of 2, The number of Transformer encoder layers is 2, with a dropout rate of 0.4 at each layer. The network output is 4 (corresponding to the number of classes in the HNMU2 dataset). Models use the AdamW optimizer with lr = 0.005 and weight decay = 0.0005.

For VNU, the Graphomer model selects d_model = 64. We also configure max_distance = 12. Attention dropout rate of 0.1, a multi-head attention value of 4, The number of Transformer encoder layers is 1. The network output is 4 (corresponding to the number of classes in the HNMU2 dataset). Models use the AdamW optimizer with lr = 0.005 and weight decay = 0.0005.

For SATDAP, the Graphomer model selects d_model = 64. We also configure max_distance = 12. Attention dropout rate of 0.1, a multi-head attention value of 2, The number of Transformer encoder layers is 2. The network output is 4 (corresponding to the number of classes in the HNMU2 dataset). Models use the AdamW optimizer with lr = 0.005 and weight decay = 0.0005.

Graph construction

For graph construction, we used the method of selecting the 10 nodes with the lowest Euclidean distance to form neighboring nodes. If the Euclidean distance is above or below the farthest neighbor within the selected K neighbors, we do not include those nodes as part of the neighborhood. In this section, we do not use threshold-based neighbor selection.

Transformer model:

For HNMU2, the Transformer model selects a multi-head attention value of 4, with a dropout rate of 0.5 at each layer. The network output is 4 (corresponding to the number of classes in the HNMU2 dataset). Models use the Adam optimizer with lr = 0.005 and weight decay = 0.0005.

For VNU, the Transformer model selects a multi-head attention value of 2 and is combined with an ANN network structured in three layers: the first layer has 64 neurons, the second layer has 128 neurons, and the third layer has 64 neurons. The dropout rate is 0.4 at each layer. The network output is 3 (corresponding to the number of classes in the VNU dataset). Models use the Adam optimizer with lr = 0.005 and weight decay = 0.0005.

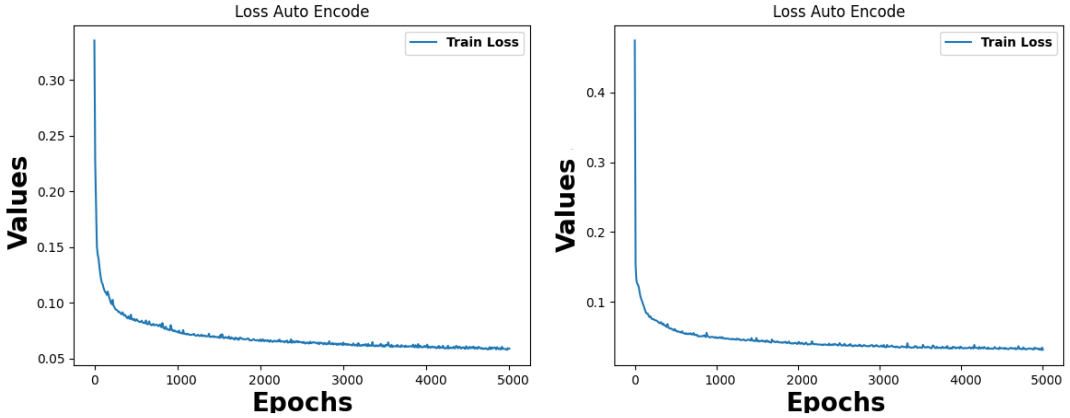
For the dataset from the SATDAP program, the Transformer model selects a multi-head attention value of 4, with a dropout rate of 0.4 at each layer. The network output is 3 (corresponding to the number of classes in the SATDAP dataset). Models use the Adam optimizer with lr = 0.005 and weight decay = 0.0005.

3.3.3.2. Model training

Train model for AutoGAT:

The number of epochs for training the Autoencode model on the HNMU2, and VNU datasets were all 5,000 epochs. The model training plots of these three datasets are shown in Figure 3.10a), Figure 3.10b) (on these figures, the values at the epochs that are divisible by 10 are shown). The principle of model selection is to select the Autoencode model with the smallest loss value.

On that principle, with the model in Figure 3.10a), the model is selected at the 4943rd epoch because it has a loss of 0.0557, and with the model in Figure 3.10b), the model is selected at epoch 4981 because it has a loss of 0.0310.

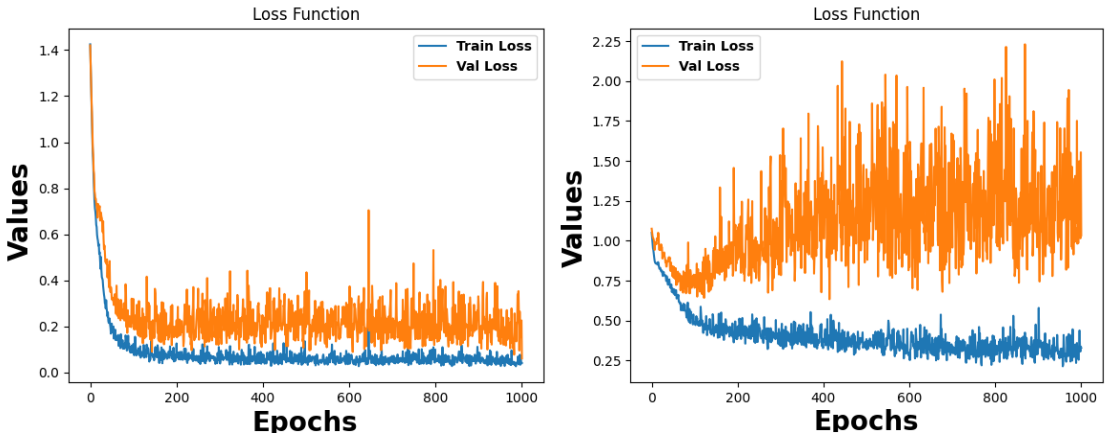


a) b)

Figure 3. 10. Autoencoder model training according to AutoGAT

a) on the HNMU2 dataset. b) on the VNU dataset.

In AutoGAT, we trained the GAT model on three datasets (1000 epochs). The training graphs of the models are shown in Figures 3.11a), 3.11b), respectively. The principle of selecting the best model is to take the average of the training and validation loss values. In which epoch gives the smallest value, the model is selected at that epoch.



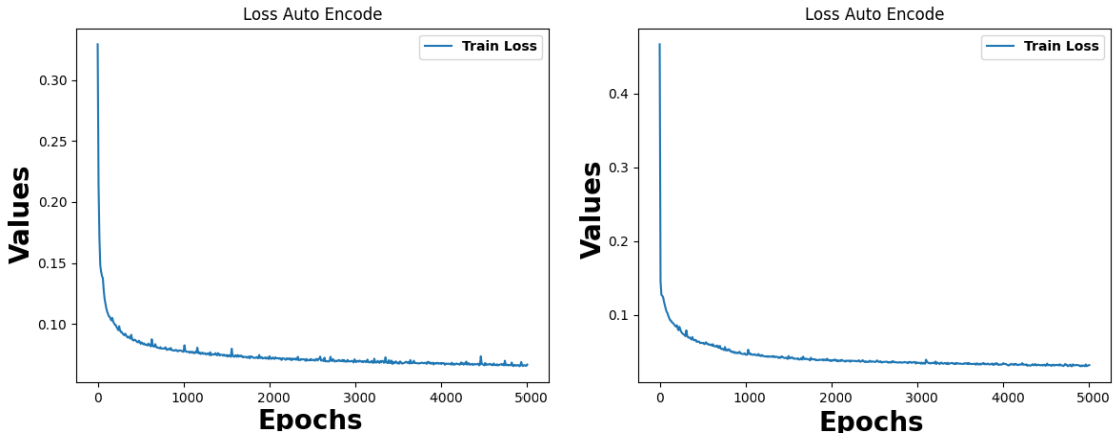
a) b)

Figure 3. 11. Training of the GAT model according to AutoGAT

a) on the HNMU2 dataset. b) on the VNU dataset.

Train model for AWG-GAT:

The number of epochs used to train the Autoencode model on the three datasets was 5,000 epochs. The training graphs of the models are shown in Figure 3.12 (on this figure, the values at the epochs that are divisible by 10 are shown). The principle of model selection is to select the Autoencode model with the smallest loss value.



a) b)

Figure 3. 12. Autoencoder model training according to AWG-GAT

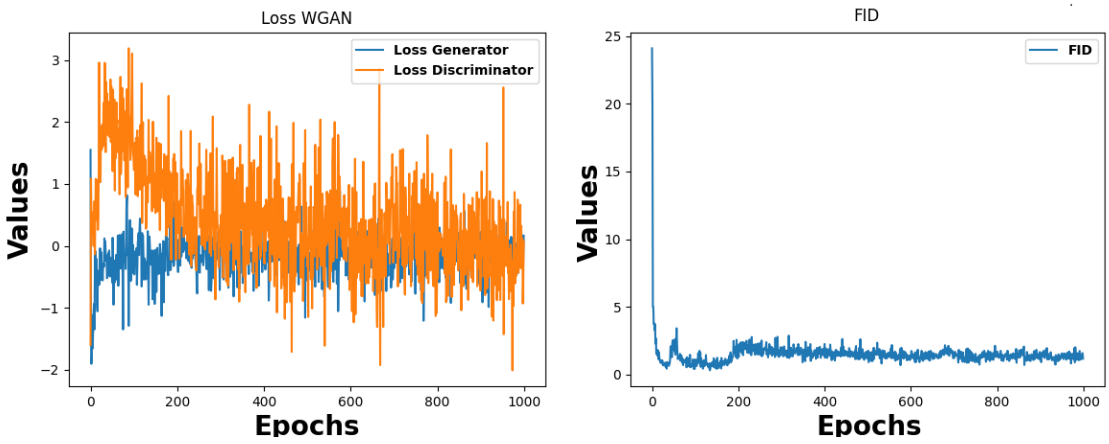
a) on the HNMU2 dataset. b) on the VNU dataset.

Figures 3.10, 3.11 and 3.12 illustrate the different training stages in the experimental scenarios and highlight the significant differences in the role of the Autoencoder and data processing. Specifically, Figure 3.10 demonstrates the training process of the Autoencoder in AutoGAT, where the model is trained on the original dataset, including both labeled and unlabeled samples, without any data augmentation techniques being applied. The goal at this stage is to extract latent features that help reduce dimensionality and enrich the input information for subsequent classification models. Figure 3.11 further illustrates the training process of the GAT model in the same scenario, using the output features from the Autoencoder in Figure 3.10.

Figure 3.12, belonging to AWG-GAT (the scenario involving synthetic data from WGAN), shows the Autoencoder training process, which is still conducted on the original dataset, without including any synthetic samples. Keeping the original dataset at the Autoencoder training stage ensures stability in learning feature representations and avoids potential bias from synthetic data that may not fully reflect the true distribution. Once the Autoencoder is trained, the WGAN model is then applied to generate additional data based on the same

original dataset. The entire expanded dataset (comprising both original and synthetic data) is subsequently passed through the Autoencoder's encoder to extract features for subsequent steps such as graph construction and classification. Therefore, in AWG-GAT, the role of the Autoencoder in the initial stage is similar to that in AutoGAT.

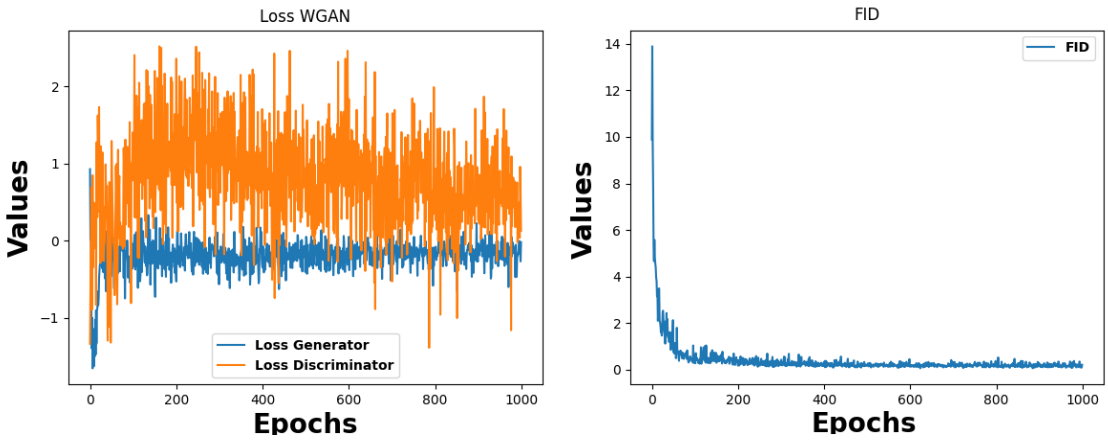
In AWG-GAT, the number of epochs selected to train the WGAN model on the three datasets was 1,000. The training graphs of the models are shown in Figures 3.13, 3.14, respectively. The principle of model selection is which WGAN model has the smallest FID value.



a) b)

Figure 3. 13. Training of the WGAN model on the HNMU2 dataset according to AWG-GAT.

a) Loss value. b) FID value.

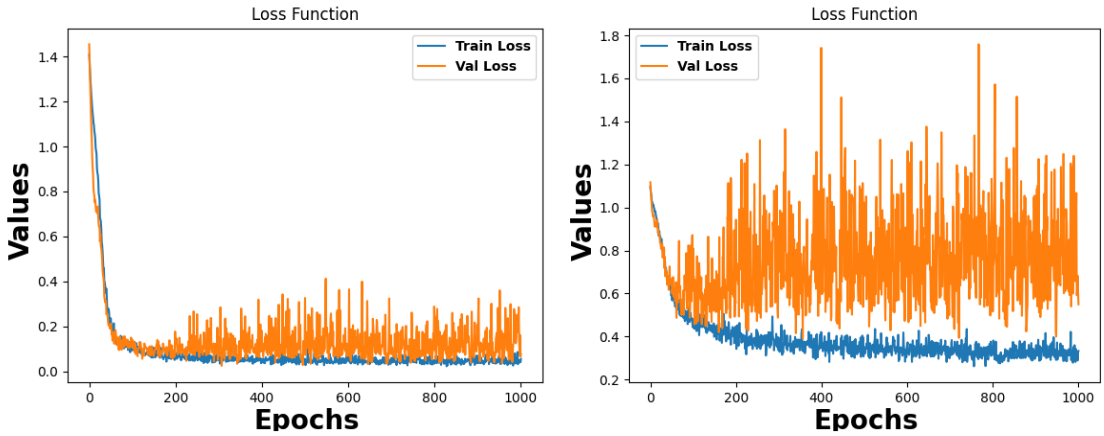


a) b)

Figure 3. 14. Training of the WGAN model on the VNU dataset according to AWG-GAT

(a) Loss value. (b) FID value.

In AWG-GAT, we trained the GAT model on the three datasets (1000 epochs). The training graphs of these models are shown in Figure 3.15a), Figure 3.15b). The principle of selecting the best model is to take the average of the training and validation loss values. In which epoch gives the smallest value, the model is selected at that epoch. On that principle, with the model shown in Figure 3.15a), the model selected at the 980th epoch has a training loss of 0.0338 and a validation loss of 0.0423. With the model shown in Figure 3.15b), the model selected at the 994th epoch had a training loss of 0.3369 and a validation loss of 0.5201.



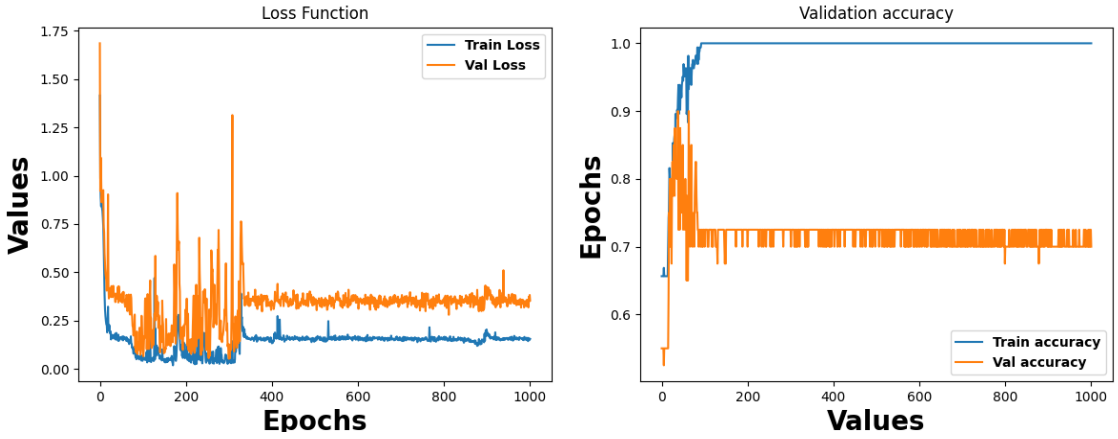
a) b)

Figure 3. 15. Training of the GAT model according to AWG-GAT:

a) on the HNMU2 dataset. b) on the VNU dataset.

Train model for Graphormer:

We trained the model with 1000 epochs on the HNMU2, VNU and SATDAP. The model training graphs for the three datasets HNMU2, VNU and SATDAP are shown in Figure 3.16. The principle of selecting the best model is to take the average of the training and validation loss values. In which epoch gives the smallest value, the model is selected at that epoch.

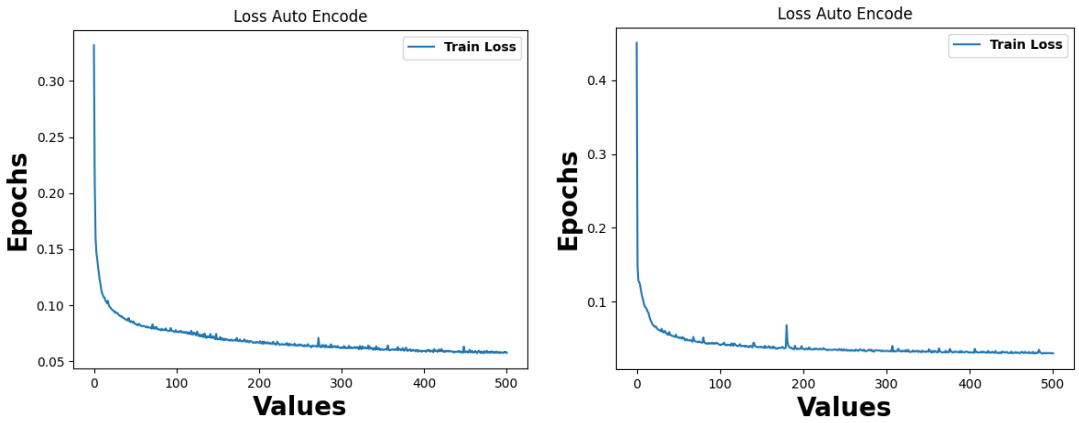


a) b)

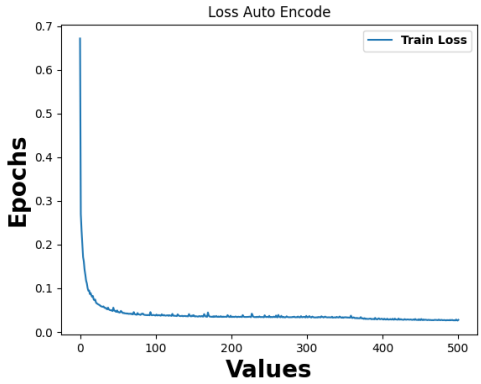
Figure 3. 16. Training of the Graphomer model
a) on the HNMU2 dataset. b) on the VNU dataset.

Train model for AWG-GC:

The number of epochs used to train the Autoencode model on the three datasets was 5,000 epochs. The training graphs of the models are shown in Figure 3.17a), Figure 3.17b), and Figure 3.17c) respectively (on these figures, the values at the epochs that are divisible by 10 are shown). The principle of model selection is to select the Autoencode model with the smallest loss value. On that principle, with the model in Figure 3.17a), the model is selected at the 4928nd epoch because it has a loss of 0.0567. With the model in Figure 3.17b), the model is selected at the 4822nd epoch because it has a loss of 0.0291. With the model in Figure 3.17c), the model is selected at epoch 4990th because it has a loss of 0.0261.



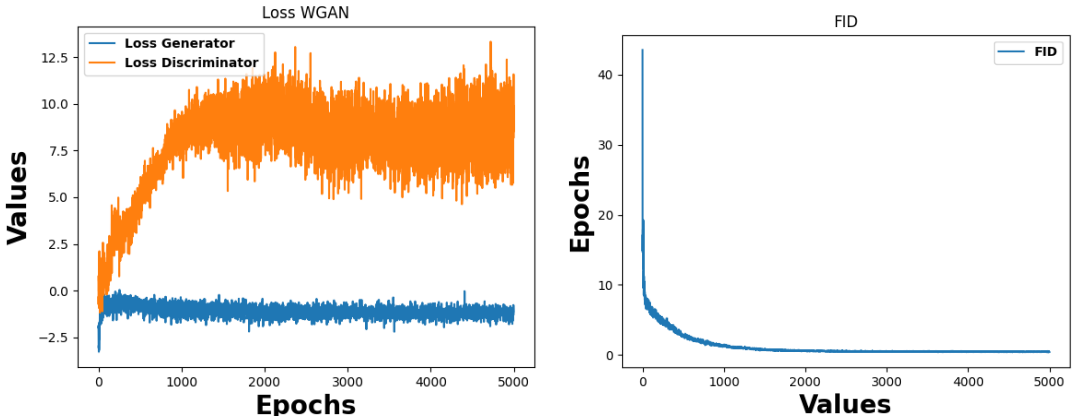
a) b)



c)

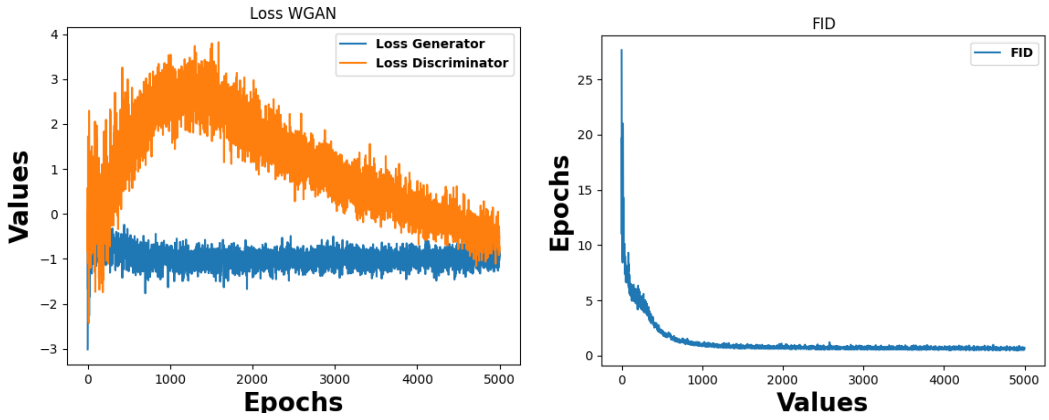
Figure 3. 17. Autoencoder model training according to AWG-GC a) on the HNMU2 dataset. b) on the VNU dataset. c) on the SATDAP dataset.

In AWG-GC with Graphomer, the number of epochs selected to train the WGAN model on the three datasets was 5,000. The training graphs of the models are shown in Figures 3.18, 3.19, and 3.20, respectively.



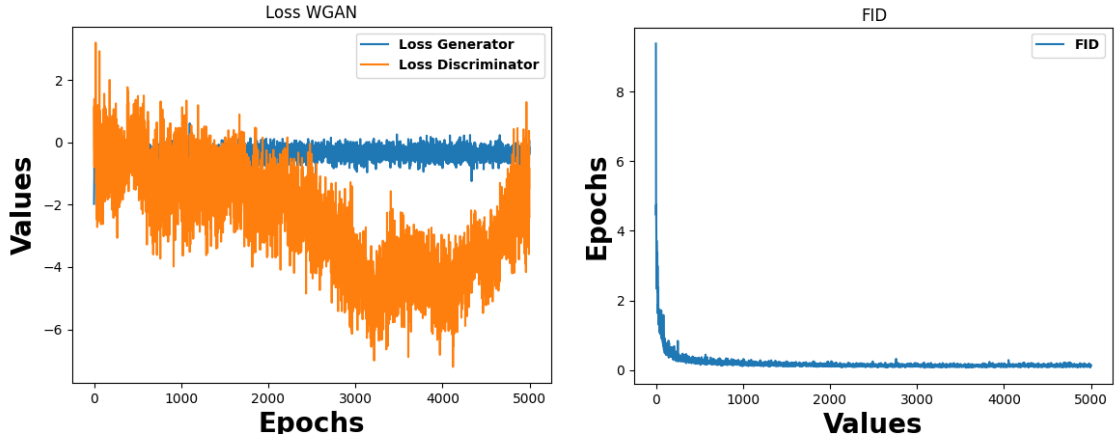
a) b)

Figure 3. 18. Training of the WGAN model on the HNMU2 dataset according to AWG-GC a) Loss value. b) FID value.



a) b)

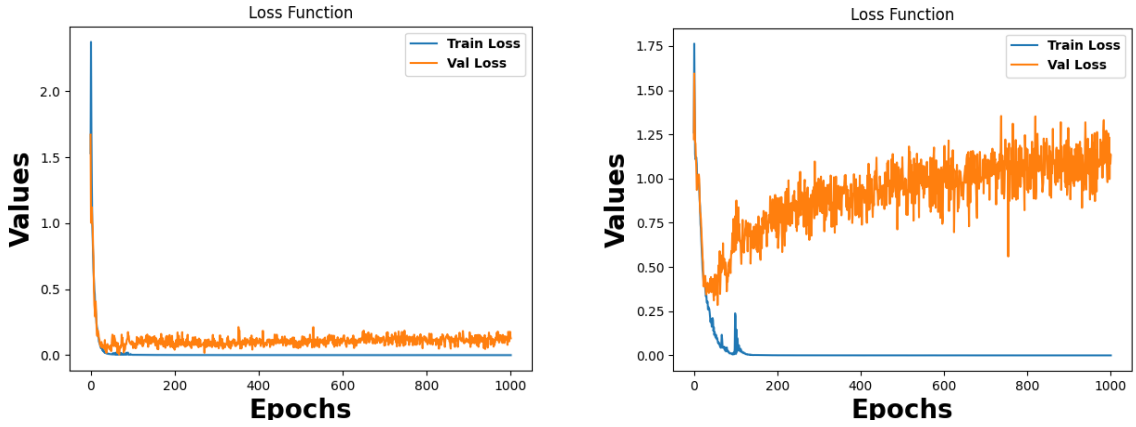
Figure 3. 19. Training of the WGAN model on the VNU dataset according to AWG-GC a) Loss value. b) FID value.



a) b)

Figure 3. 20. Training of the WGAN model on the SATDAP dataset according to AWG-GC a) Loss value. b) FID value.

In AWG-GC, we trained the Graphomer model on the three datasets (1000 epochs). The training graphs of these models are shown in Figure 3.21. The principle of selecting the best model is to take the average of the training and validation loss values. In which epoch gives the smallest value, the model is selected at that epoch..



a) b)

Figure 3. 21. Training of the Graphomer model according to AWG_GC a) on the HNMU2 dataset. b) on the VNU dataset.

3.3.4. Results and discussions

3.3.4.1. Results obtained on the HNMU2 dataset

Table 3. 9. Prediction results on the HNMU2 dataset.

Method	Accuracy	Precision	Recall	F1-Score
SVM	80.29	41.38	41.81	40.55
KNN	80.29	40.68	41.58	40.59
RF	95.62	47.79	48.31	48.34
Transformer	95.62	72.77	60.99	64.79
GAT	89.05	53.52	57.95	55.16
Graphomer	97.08	73.45	73.97	73.67
AutoGAT	93.43	59.84	59.74	59.74
AWG_GAT	**97.08**	**98.50**	**86.41**	**90.37**
AWG-GC	**98.54**	**99.25**	**99.25**	**99.25**

From Table 3.10, we can see that the accuracy of AWG-GC was the highest (98.54%), 18.25% higher than that of SVM, 18.25% higher than that of KNN, 2.92% higher than that of Transformer and RF, 9.49% higher than that of GAT, 1.46% higher than that of Graphomer, 5.11% higher than that of AutoGAT and 1.46% higher than that of AWG_GAT.

In addition, the prediction accuracy of AWG-GC was the highest (99.25%), 57.87% higher than that of SVM, 58.57% higher than that of KNN, with 51.46% higher than that of RF, with 26.48% higher than that of Transformer, 45.73% higher than that of GAT, 25.8% higher than that of Graphomer, 39.41% higher than that of AutoGAT and 0.75% higher than that of AWG_GAT. This resulted in a higher rate of correct positive predictions, minimizing false positives.

Table 3.10 also shows that the sensitivity of AWC-GC is the highest (99.25%), 57.44% higher than that of SVM, 57.67% higher than that of KNN, with 50.94% higher than that of RF, with 38.26% higher than that of Transformer, 41.3% higher than that of GAT, 25.34% higher than that of Graphomer, 39.51% higher than that of AutoGAT and 12.84% higher than that of AWG_GAT. This demonstrates that the AWG-GC model is capable of detecting more true positive samples and minimizing false negative cases.

In particular, the F1-Score of AWG-GC is the highest (99.25%), 58.7% higher than that of SVM, 58.66% higher than that of KNN, with 50.91% higher

than that of RF, with 34.46% higher than that of Transformer, 44.09% higher than that of GAT, 25.58% higher than that of Graphomer, 39.51% higher than that of AutoGAT, and 8.88% higher than that of AWG_GAT. The F1-score shows that AWG-GC achieves the best balance between accurately predicting positive samples and detecting more positive samples. These improvements indicate that AWG-GC outperforms the remaining methods.

3.2.3.2 Results obtained on VNU dataset

Table 3. 10. Prediction results obtained on the VNU dataset

Method	Accuracy	Precision	Recall	F1-Score
SVM	83.82	42.43	53.83	46.97
KNN	86.76	51.45	54.98	53.12
RF	82.35	54.03	46.91	49.39
Transformer	86.76	69.72	71.73	70.72
GAT	80.88	51.60	50.52	51.00
Graphomer	88.24	80.11	63.93	64.97
AutoGAT	85.29	74.50	58.59	53.96
AWG-GAT	**89.71**	**70.95**	**95.98**	**78.64**
AWG-GC	**94.12**	**81.67**	**97.70**	**88.17**

The results in Table 3.11 reveal that the AWG-GC model consistently outperforms all other methods across all evaluation metrics on the VNU dataset. It achieved the highest accuracy (94.12%), precision (81.67%), recall (97.70%), F1-score (88.17%), indicating both superior predictive accuracy and robust classification capability. Compared to the baseline models, AWG-GC improved accuracy by 5.88% over Graphormer, 7.36% over Transformer, and 10.30% over SVM. Notably, its recall increased dramatically by 33.77% compared to Graphormer and 25.97% over Transformer, suggesting an exceptional ability to correctly identify positive cases while minimizing false negatives.

Furthermore, the F1-score of AWG-GC (88.17%) represents a substantial improvement, outperforming Graphormer by 23.20% and Transformer by 17.45%, reflecting a well-balanced trade-off between precision and recall.

3.2.3.3 *Results obtained on SATDAP dataset*

Table 3. 11. Prediction results obtained on the SATDAP dataset

Method	Accuracy	Precision	Recall	F1-Score
SVM	77.59	71.89	68.17	68.85
KNN	66.67	57.66	54.73	55.35
RF	79.32	70.78	68.65	69.37
Transformer	80.34	71.87	70.99	71.34
Graphomer	80.79	74.08	70.30	71.67
AWG-GC	**81.81**	**74.74**	**73.89**	**74.21**
XGBoost (Martin et al., [63])	**73.00**			**65.00**

According to the results presented in Table 3.12, the AWG-GC model achieved the highest overall performance across all evaluation metrics. It reached an accuracy of 81.81%, surpassing SVM by 4.22%, KNN by 14.91%, RF by 2.49%, Transformer by 1.47%, and Graphormer by 1.02%. Its prediction accuracy (74.74%) similarly exceeded that of SVM by 2.85%, KNN by 17.08%, RF by 3.96%, Transformer by 2.87%, and Graphormer by 0.66%. Moreover, AWG-GC recorded the highest sensitivity (74.21%), outperforming SVM by 5.72%, KNN by 19.16%, RF by 5.24%, Transformer by 2.90%, and Graphormer by 3.59%. Its F1-score (74.21%) also led all models, with improvements of 5.36% over SVM, 18.86% over KNN, 4.84% over RF, 2.87% over Transformer, and 2.54% over Graphormer.

When compared to previous studies, the performance of AWG-GC is particularly notable. Martins et al. ([63]) reported that Extreme Gradient Boosting (XGBoost) achieved an accuracy of 73% and an F1-score of 65% in predicting student performance. In contrast, AWG-GC outperformed XGBoost by 8.81% in accuracy and 9.21% in F1-score. This comparison further underscores the superiority of AWG-GC in both predictive accuracy and balanced classification performance.

These results indicate that the integration of Autoencoder, WGAN, and Graphormer architectures enables the model to better capture the underlying structure of educational data and effectively address challenges such as small sample sizes and class imbalance. Overall, AWG-GC demonstrates a significant improvement over both traditional machine learning approaches and previously reported models in the literature ([63]).

Although AWG-GC achieves the highest performance on the SATDAP dataset, the model still has certain limitations. Overall accuracy does not exceed 82%, and the F1-Score remains below 80%, indicating that its ability to detect and classify correctly is still limited, especially in real-world contexts that demand high reliability. Furthermore, the performance improvement over Transformer or Graphormer is relatively small (only about 1–1.5%), while the computational cost is high due to the integration of multiple components (Autoencoder, WGAN, Graphormer). To address these issues, future work should incorporate additional features from learning behaviors and contextual factors, optimize the WGAN data generation strategy to produce more realistic synthetic samples, and develop a lightweight version of the model to reduce training costs and improve its practical applicability.

In AWG-GC, we demonstrated that for all three datasets HNMU2, VNU and SATDAP augmenting the training set with data generated by WGAN improved the model's predictive performance. Specifically, HNMU2 achieved an improvement of 1.46%, VNU improved by 5.88%, and SATDAP showed an increase of 1.02% compared to when no additional training data was used. This proves that a larger training dataset enhances the model's predictive capability.

For multi-class prediction tasks, it is essential to evaluate the performance for each class label individually to ensure balanced and reliable classification. Therefore, Tables 3.13, 3.14 and 3.15 have been included to provide a detailed breakdown of the model's performance across all class labels in each dataset.

Table 3. 12. Per-class performance evaluation table of the AWG-GC model on the HNMU2 dataset

	Precision	**Recall**	**F1-Score**
Medium	100	100	100
Good	98.65	98.65	98.65
Very good	98.33	98.33	98.33
Excellent	100	100	100

Table 3. 13. Per-Class Performance Evaluation Table of the AWG-GC Model on the VNU Dataset

	Precision	**Recall**	**F1-Score**
Good	75	100	85.71
Very good	100	93.10	96.43
Excellent	70	100	82.35

Table 3. 14. Per-Class Performance Evaluation Table of the AWG-GC Model on the SATDAP Dataset

	Precision	**Recall**	**F1-Score**
Graduate	86.33	90.77	88.49
Enrolled	47.69	49.60	48.63
Dropout	90.19	81.29	85.51

In addition to overall performance metrics, detailed error analysis helps clarify the model's predictive behavior. Figure 3.22 presents confusion matrices for the HNMU, VNU, and SATDAP datasets, highlighting misclassifications caused by unclear class boundaries and class imbalance.

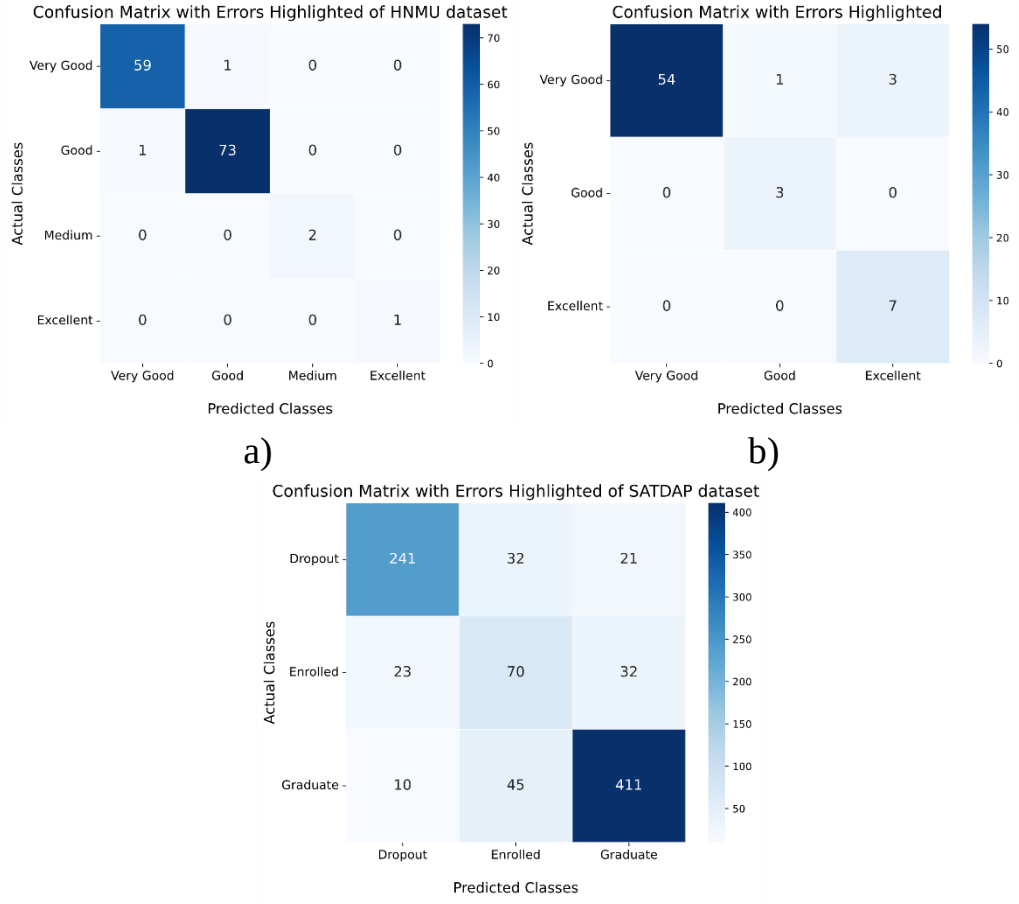


Figure 3. 22. Confusion Matrices (in the AWG-GC model)

a) on the HNMU2 dataset. b) on the VNU dataset. c) on the SATDAP dataset.

We conducted a detailed evaluation of the most commonly confused cases across the three datasets and identified the underlying causes of misclassification. For the HNMU2 dataset, although the AWG-GC model achieved near-perfect classification performance (F1-score = 99.25%), the

confusion matrix reveals some misclassifications between the "Good" and "Very Good" categories. This issue is largely due to the relatively close GPA ranges and overlapping behavioral features derived from survey responses. Additionally, the dataset's class imbalance, particularly the larger number of “Good” samples (338) compared to “Very Good” samples (190), may have biased the model toward favoring the majority class. These factors combined make it challenging for the model to draw a clear boundary between these two performance levels. In the VNU dataset, the classification task becomes more complex due to the small sample size and the large number of features, increasing the risk of overfitting. A common error here is the misclassification of "Good" students as "Very Good" or even "Excellent." There is significant overlap in course grades, especially in high-weight subjects. Although the "Excellent" group is identified with perfect recall (Recall = 100%), the boundary between "Good" and "Very Good" remains ambiguous in many cases, affecting overall classification accuracy. In the SATDAP dataset, the most challenging issue lies in distinguishing the "Enrolled" group from the other two groups: "Graduated" and "Dropout." This is understandable, as the "Enrolled" status is transitional, with students exhibiting characteristics similar to those who have either completed the program or are at risk of dropping out. Furthermore, the number of samples in this group is relatively small in the training data, resulting in class imbalance and affecting prediction accuracy. Nevertheless, the model still achieved an F1-score of 74.21%, demonstrating strong generalization capability even in the presence of structural class complexity and data imbalance. In summary, most classification errors stem from the inherent ambiguity between classes rather than limitations in the model’s capability.

In the experiments above, we observed the following differences between the datasets and their impact on model performance:

HNMU2 is the dataset with the most diverse features (including several surveys on soft skills, learning behaviors, and specialized course results), with a moderate sample size. The AWG-GC model achieved the highest performance here (Accuracy 98.54%, F1-score 99.25%), demonstrating its ability to effectively leverage complex relationships in graph data and the benefits of synthetic data generation.

VNU is the dataset with a small number of samples but a large number of features, increasing the risk of overfitting. However, the model still achieved an F1-score of 88.17%, proving the stability and high generalization capability of AWG-GC even in a context with limited data.

SATDAP represents an international context with imbalanced label distribution and distinct feature structures. Although this dataset contains many samples, the model maintained high performance (F1-score 74.21%), showcasing the wide applicability of the approach.

The differences between the three datasets in terms of sample size, feature types, and application contexts helped validate the adaptability and generalization ability of the proposed model. The AWG-GC model not only maintained high accuracy on each individual dataset but also demonstrated stable performance when facing different data characteristics, ranging from simple tabular structures to complex graph relationships, from small to large datasets, and from domestic to international data.

To validate the necessity of each component in AWG-GC, we conducted an ablation study by removing or replacing each individual component:

Removing WGAN: For the HNMU2 dataset, when trained only on real data without synthetic data from WGAN, the model experienced overfitting, and performance decreased by 1.46%. For the VNU dataset, when trained only on real data without synthetic data from WGAN, the model also suffered from overfitting, and performance decreased by 5.88%. For the SATDAP dataset, when trained only on real data without synthetic data from WGAN, the model exhibited overfitting, and performance decreased by 1.02%.

Replacing Graphormer with GAT: When replacing Graphormer with GAT, the model failed to effectively capture relationships in the data, resulting in an average decrease of 8.03% in accuracy for the HNMU2 dataset and 7.36% for the VNU dataset.

Although the AWG-GC model delivers superior performance compared to traditional methods, there are still several challenges and limitations that need to be addressed to ensure broader applicability. The use of synthetic data generated by WGAN may lead to overfitting if the generated samples lack sufficient diversity or do not accurately reflect the real distribution. This is particularly problematic when the original dataset is small, causing the model to overly rely on synthetic samples. The combination of Autoencoder, WGAN, and Graphormer increases the number of parameters compared to simpler methods like GCN or MLP. As a result, it requires substantial computational resources, making it challenging to deploy on systems with limited hardware.

3.4. Appendix to Chapter 3

3.4.1. Wasserstein GANs (WGAN)

The divergence that GANs typically minimize is probably discontinuous with respect to the Generator G parameters. This makes training difficult. The Wasserstein-1 (also called Earth-Mover) distance $W(p_1, p_2)$ is recommended. $W(p_1, p_2)$ is the minimum mass transport cost for converting distribution p_1 to p_2 (where cost is mass multiplied by transport distance). With loose assumptions, $W(p_1, p_2)$ is continuous everywhere and differentiable almost everywhere.

Use Kantorovich-Rubinstein duality ([78]) is used in building the WGAN value function to obtain:

$$\min_G \max_{D \in \widetilde{D}} E_{x \backsim \mathrm{P}_r}[logD(x)] - E_{\tilde{x} \backsim \mathrm{P}_g}[D(\tilde{x})], \tag{3.7}$$

In Equation (3.6), $\widetilde{D}$ is the set of 1-Lipschitz functions and P_g is the implicit distribution of the model determined by $\tilde{x} = G(z), z \backsim p(z)$. In this case, under optimal Discriminator D, minimizing the value function relative to the parameters of Generator G minimizes $W(p_r, p_g)$.

The WGAN objective function introduces a Critic function whose gradient concerning its input is more effective compared to a standard GAN. This enhancement facilitates the optimization of Generator G. To enforce a Lipschitz constraint on the Critic, (Arjovsky et al., 2017) suggested clipping the weights of the Critic to lie within a bounded range $[-m, m]$. The set of functions that adhere to this constraint forms a subset of k -Lipschitz functions, where the Lipschitz constant k depends on m and the architecture of the Critic.

The principle of model selection is also which WGAN model has the smallest FID value. The difference between the generated data and the original data.

3.4.2. The Transformer model for the task of predicting graduation classification

The architecture of the Transformer model in this chapter is fundamentally designed as described in Sections 1.2.2 and 2.4.3.

Assume that a data sample is represented by the pair (x, y), where $x_{cont} \in R^P$ is a vector of P continuous features representing a student's grades over four semesters, where each semester includes m_i subjects ($i = 1,2,3,4$), along with survey data that has been numerically encoded. The label y corresponds to the graduation classification, which falls into one of the following categories: excellent, very good, good, medium, poor and very poor.

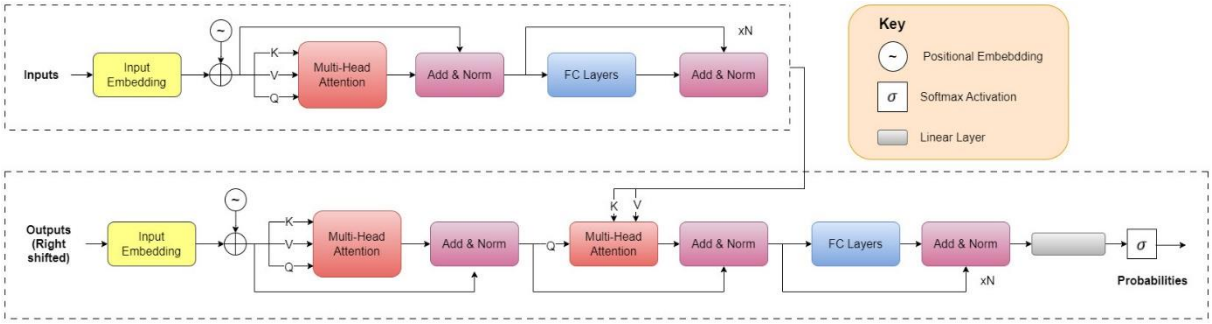


Figure 3. 23. Transformer model for the task of predicting graduation classification.

For the classification task, the loss function used is the cross-entropy loss.

$$L(x, y) = H(g_\psi(f_\theta(X_{emb})), y) \quad (3.8)$$

3.4.3. Graphormer

Graphormer ([80]) is an advanced deep learning architecture developed to extend the representational capabilities of the Transformer model from sequential (series) data to graph-structured data. While traditional graph models like GCN and GAT rely on local message passing through neighborhood layers, Graphormer leverages a global self-attention mechanism to capture long-range and diverse relationships between nodes in a graph.

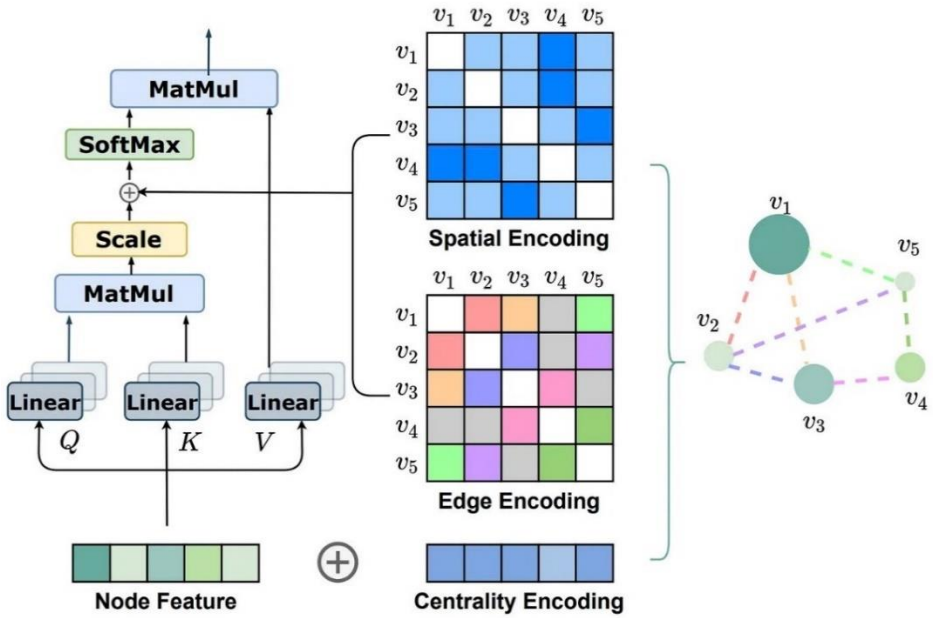


Figure 3. 24. The Graphormer model ([80])

Unlike traditional Transformers, Graphormer directly integrates three key structural aspects of graphs into the attention mechanism: centrality,

spatial position, and edge features. These encoding techniques are designed to preserve the non-sequential nature of graphs while providing structural context for each attention computation.

Centrality Encoding: To reflect the importance of each node within the graph, Graphormer incorporates in-degree and out-degree information into the node embeddings. The input vector of node *i* is defined as follows:

$$h_i^{(0)} = x_i + z_i^- + z_i^+. \tag{3.9}$$

where: x_i is the initial feature of node *i*, z_i^-: the embedding learned from the in-degree, and z_i^+: the embedding learned from the out-degree.

Spatial Encoding: The distance between nodes is calculated using the shortest path length $\phi(i, j)$, from which a spatial embedding $b_{\phi(i,j)}$ is generated. This embedding serves as a bias term in the attention mechanism:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}} + B\right)V \tag{3.10}$$

where B is the spatial bias matrix obtained from the embedding $b_{\phi(i,j)}$.

Edge Encoding: Graphormer utilizes information from the edges lying along the shortest path between two nodes *i* and *j*, computing the edge embedding component c_{ij} as follows:

$$c_{ij} = \frac{1}{N}\sum_{n=1}^{N} MLP(x_n^{(e)}) \tag{3.11}$$

where N is the number of edges on the shortest path between nodes *i* and *j*, $x_n^{(e)}$ is the feature of the *n*-th edge, and MLP is a deep neural network that generates an embedding from the edge feature. This edge embedding is then integrated into the attention score as follows:

$$A_{ij} = \frac{(h_i W_Q)(h_j W_K)^T}{\sqrt{d}} + b_{\phi(i,j)} + c_{ij}. \tag{3.12}$$

Virtual Node: Similar to the [CLS] token in BERT, Graphormer introduces a virtual node that connects to the entire graph. This node does not exist in the original structure but is capable of aggregating global information, effectively supporting tasks such as graph-level classification.

According to an analysis on Medium titled Graphormer on Medium, Graphormer demonstrates outstanding performance on several benchmark datasets, such as ZINC and PCQM4Mv2. Key improvements include:

- Eliminating the need for multi-layer message passing as seen in GCN/GAT.
- Mitigating the over-smoothing effect.
- Preserving the parallel computation capability of traditional Transformers.

In summary, Graphormer marks a significant advancement in generalizing the Transformer architecture to the domain of graph data. By directly incorporating graph structural information into the attention mechanism, Graphormer provides a powerful and flexible deep learning framework for complex graph-based machine learning tasks.

The conclusion of Chapter 3

The early prediction of graduation classification holds significant practical value in higher education. It enables institutions to make timely, data-informed decisions that support quality assurance, curriculum development, and strategic planning. For students, early insights into their likely graduation outcomes provide opportunities for academic adjustment, proactive learning, and career preparation.

This chapter demonstrated that graduation classification is influenced not only by academic performance but also by a variety of personal, family, social, and institutional factors. Therefore, predictive models must go beyond traditional approaches by integrating diverse data sources and addressing the uncertainty inherent in educational environments.

By proposing deep learning models that incorporate both academic and non-academic data - along with mechanisms for handling incomplete and uncertain information - this study contributes to the development of more accurate and realistic tools for early graduation classification prediction. These models lay the foundation for more personalized academic advising and improved educational management.

In this chapter, the dissertation presented integrated deep learning models aimed at improving performance in the task of early prediction of students' graduation classification, including the LATCGAd and AWG-GC models. LATCGAd and AWG-GC are notable for their tight integration of preprocessing, data generation, and classification learning within a unified architecture. Instead of handling data processing and model training as separate

stages, these models operate cohesively, allowing their components to support each other in optimizing overall system performance. This approach improves accuracy, enhances generalization capability, and offers better adaptability to complex, imbalanced, or low-label datasets. Experiments on the HNMU2 dataset demonstrated the effectiveness of the models: LATCGAd reached an accuracy of 96.97% and an F1 score of 73.66%, while AWG-GC outperformed the others with an accuracy of 98.54% and an impressive F1 score of 99.25%.

However, models with complex architectures like AWG-GC require significant computational resources and long training times, which may limit their practical deployment. Therefore, in scenarios with limited data or computational constraints, lighter models such as LATCGAd may be more suitable choices.

CONCLUSION AND FUTURE DEVELOPMENT

A. Key contributions of the dissertation

This dissertation has addressed the challenge of predicting student academic outcomes under the conditions of uncertainty, data scarcity, and imbalance that characterize real-world educational environments.

In the first stage, the research focused on short-term SGPA prediction, demonstrating that SGPA should be treated as a dynamic and uncertain indicator rather than a fixed value. To capture this complexity, two novel frameworks-NeutroDL and NeutroGNT-were proposed, integrating deep learning with neutrosophic theory to manage incomplete and uncertain data. Experimental results confirmed their superiority, with NeutroGNT achieving a minimum MSE of 0.018 and a maximum R^2 of 96.05%, significantly outperforming conventional approaches. These findings highlight the effectiveness of uncertainty-aware deep learning models in supporting timely academic monitoring, early intervention, and personalized learning pathways.

Building on this foundation, the research advanced to the long-term prediction of graduation classification, a task with broader strategic implications for educational policy and quality management. To this end, two hybrid deep learning models were developed: LATCGAd, which integrates Transformer, CGAN, and Adaptive Layer Normalization, achieving 96.97% accuracy and a 73.66% F1-score; and AWG-GC, which combines Autoencoder, Wasserstein GAN, and Graphormer, simultaneously addressing representation learning, data augmentation, and classification. The AWG-GC model achieved 98.54% accuracy and a 99.25% F1-score, markedly surpassing baseline models and demonstrating the benefits of unifying advanced generative and graph-based architectures.

Overall, the dissertation makes three major contributions: (i) the development of uncertainty-aware predictive frameworks (NeutroDL and NeutroGNT) for SGPA prediction, (ii) the design of advanced hybrid models (LATCGAd and AWG-GC) for robust graduation classification under imbalanced data conditions, and (iii) the creation of enriched educational datasets and analytical pipelines tailored for real-world application. Together, these results provide both methodological advances and practical tools to support data-driven, adaptive, and intelligent decision-making in higher education.

B. Future research directions

Based on the results achieved, the dissertation proposes several promising directions for future research:

1. Broaden prediction targets to include dropout risk, program completion, course satisfaction, and career orientation, thereby providing a more comprehensive view of students' learning trajectories.
2. Apply reinforcement learning and unsupervised learning, combined with explainable AI (XAI) techniques, to both personalize learning pathways and provide transparent, interpretable justifications that enhance trust in early intervention decisions by instructors and administrators.
3. Leverage federated learning and transfer learning to develop models that ensure predictive effectiveness and generalization capability while preserving data privacy across institutions.
4. Develop an online Learning Analytics (LA) system based on the proposed models, integrated with XAI, to deliver real-time monitoring, intuitive explanations, and actionable recommendations for both students and educators.

These directions not only extend the impact of the current research but also foster sustainable, data-driven digital transformation in higher education, toward a smart, adaptive, and transparent learning ecosystem.

LIST OF PUBLICATIONS

[**CT1**]**. Son, N. T. K.**, Bien, N. V., Quynh, N. H., and Tho, C. C. (2022). Machine learning-based admission data processing for early forecasting of students' learning outcomes. *International Journal of Data Warehousing and Mining*, *18*(1). (SCIE). https://www.igi-global.com/gateway/article/313585

[**CT2**]**. Son, N. T. K.,** Quynh, N. H., and Minh, B. T. (2024). Early prediction of students' graduation rank using LAGT: Enhancing accuracy with GCN and Transformer. *Journal of Computer Science and Cybernetics*, *40*(4), 299-314. https://doi.org/10.15625/1813-9663/21095

[**CT3**]**. Son, N. T. K.**, Hoa, N. H., Trang, H. T. T., and Ngan, T. Q. (2024). Introducing a cutting-edge dataset: Revealing key factors in student academic outcomes and learning processes. *VNU Journal of Science: Education Research*, *40*(4). https://js.vnu.edu.vn/ER/article/view/5157.

[**CT4**]**.** Vinh, T. D., **Son, N. T. K.,** and Diep, L. N. (2025). Dataset of factors affecting learning outcomes of students at the University of Education, Vietnam National University, Hanoi. *Data in Brief, 59*, 111438. (ESCI). https://doi.org/10.1016/j.dib.2025.111438

[**CT5**]**. Son, N. T. K.**, Dat, B. V., Hoa, N. H., and Trang, H. T. T. (2025). Hybrid artificial intelligence models incorporating neutrosophy for predicting student outcomes. *Lecture Notes in Networks and Systems.* (Accepted; Scopus).

[**CT6**]**. Son, N. T. K.**, Thong, N. T., and Quynh, N. H. (2025). Neutrosophy-driven deep learning for predicting student performance. *Neutrosophic Sets and Systems, Volume 87 (2025).* (Scopus).

[**CT7**]**. Son, N. T. K**., Quynh, N. H., and Minh, B. T. (2025). Refining graduation classification accuracy with synergistic deep learning models. *Cybernetics and Information Technologies, Volume **25**, No 2, 2025*. (Scopus, ESCI).

[**CT8**]**. Son, N. T. K.,** Vinh, T. D., Quynh, N. H., Tao, N. Q., Hop, D. T., and Minh, B. T. (2025). Enhancing student graduation prediction with integrated deep learning under data constraints. *Data Intelligence.* (Accepted; In proof, Scopus, ESCI).

REFERENCES

[1] D. Azcona and A. F. Smeaton, “Targeting at-risk students using engagement and effort predictors in an introductory computer programming course,” in *European Conf. on Technology Enhanced Learning*, Springer, Cham, 2017, pp. 361–366.

[2] O. Viberg, M. Hatakka, O. Bälter, and A. Mavroudi, “The current landscape of learning analytics in higher education,” *Computers in Human Behavior*, vol. 89, pp. 98–110, 2018.

[3] N. Capuano and D. Toti, “Experimentation of a smart learning system for law based on knowledge discovery and cognitive computing,” *Computers in Human Behavior*, vol. 92, pp. 459–467, 2019.

[4] X. Xu, J. Wang, H. Peng, and R. Wu, “Prediction of academic performance associated with internet usage behaviors using machine learning algorithms,” *Computers in Human Behavior*, vol. 98, pp. 166–173, 2019.

[5] P. J. Piety, D. T. Hickey, and M. J. Bishop, “Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems,” in *Proc. 4th Int. Conf. on Learning Analytics and Knowledge*, 2014, pp. 193–202.

[6] G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education,” *EDUCAUSE Review*, vol. 46, no. 5, pp. 31–40, 2011.

[7] H. Pallathadka, A. Wenda, E. Ramirez‐Asís, M. Asís‐López, J. Flores‐Albornoz, and K. Phasinam, “Classification and prediction of student performance data using various machine learning algorithms,” *Materials Today: Proceedings*, vol. 80, pp. 3782–3785, 2023.

[8] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, “Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses,” *Computers in Human Behavior*, vol. 73, pp. 247–256, 2017.

[9] L. Xu, M. Skoularidou, A. Cuesta-Infante, and K. Veeramachaneni, “Modeling tabular data using conditional GAN,” *arXiv preprint* arXiv:1907.00503, 2019.

[10] R. S. Baker and P. S. Inventado, “Educational data mining and learning analytics,” in *Learning Analytics*, Springer, 2014, pp. 61–75.

[11] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang , “Educational data mining techniques for student performance prediction: Method review and comparison analysis,” *Front. Psychol.*, vol. 12, Article 698490, 2021.

[12] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons, A. Dinu, and C. Stachl, “Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics,” *Sci. Rep.*, vol. 13, p. 5705, 2023.

[13] T. T. Khoei, G. Aissou, K. Al Shamaileh, V. K. Devabhaktuni, and N. Kaabouch, “Supervised deep learning models for detecting GPS spoofing attacks on unmanned aerial vehicles,” in *2023 IEEE Int. Conf. Electro Information Technology* (eIT), pp. 340–346, 2023.

[14] T. K. F. Chiu, Q. Xia, X. Zhou, C. S. Chai, and M. Cheng, “Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education,” *Computers and Education: Artificial Intelligence*, vol. 4, 100118, 2023.

[15] M. Bienkowski, M. Feng, and B. Means, “Enhancing teaching and learning through educational data mining and learning analytics,” *U.S. Dept. of Education*, Washington, D.C., 2012.

[16] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, “Student retention using educational data mining and predictive analytics: A systematic literature review,” *IEEE Access*, vol. 10, pp. 72480–72503, 2022.

[17] V. Simeunović, S. Milić, and S. Obradović-Ratković, “Educational data mining in higher education: Building a predictive model for retaining university graduates as master’s students,” *J. College Student Retention: Research, Theory and Practice*, 2024.

[18] E. Gedrimiene, I. Celik, K. Mäkitalo, and H. Muukkonen, “Transparency and trustworthiness in user intentions to follow career recommendations from a learning analytics tool,” *J. Learning Analytics*, vol. 10, no. 1, pp. 54–70, 2023.

[19] M. Mirza, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.

[20] T. T. Khoei, H. Ould Slimane, and N. Kaabouch, “Deep learning: Systematic review, models, challenges, and research directions,” *Neural Comput. Appl.*, vol. 35, pp. 23103–23124, 2023.

[21] M. Shorfuzzaman, M. S. Hossain, A. Nazir, G. Muhammad, and A. Alamri, “Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment,” *Computers in Human Behavior*, vol. 92, pp. 578–588, 2019.

[22] NSC Research Center, “Persistence and Retention–2023.” [Online]. Available: https://nscresearchcenter.org/persistenceretention2023/

[23] ACT, “National collegiate retention and persistence to degree rates,” 2012. [Online].
Available: http://www.act.org/research/policymakers/pdf/retain_2012.pdf

[24] J. Zeng, “Predicting student performance: Analyzing socio-economic and personal factors,” *Adv. Economics, Management and Political Sciences*, vol. 42, pp. 170–178, 2023.

[25] S. L. DesJardins, D. A. Ahlburg, and B. P. McCall, “An event history model of student departure,” *Economics of Education Review*, vol. 18, no. 3, pp. 375–390, 1999.

[26] E. T. Pascarella and P. T. Terenzini, *How College Affects Students: A Third Decade of Research*. San Francisco, CA: Jossey-Bass, 2005.

[27] P. A. Murtaugh, L. D. Burns, and J. Schuster, “Predicting the retention of university students,” *Research in Higher Education*, vol. 40, no. 3, pp. 355–371, 1999.

[28] Q. Jin, M. A. Altabtabaei, K. G. Gebre-Egziabher, and S. A. St. Peter, “A multi-outcome hybrid model for predicting student success in engineering,” in *Proc. ASEE Annu. Conf.*, Vancouver, BC, Canada, 2011.

[29] M. Ji, J. Le, B. Chen, and Z. Li, “A predictive model for classifying college students' academic performance based on visual-spatial skills,” *Frontiers in Psychology*, vol. 15, p. 1434015, 2024.

[30] W. Chango, R. Cerezo, and C. Romero, “Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses,” *arXiv preprint*, arXiv:2403.05552, 2024.

[31] R. N. T. Rincy and R. Gupta, “A survey on machine learning approaches and its techniques,” in *Proc. IEEE Int. Students’ Conf. Electr., Electron. Comput. Sci. (SCEECS)*, 2020.

[32] V. Sheth, U. Tripathi, and A. Sharma, “A comparative analysis of machine learning algorithms for classification purpose,” *Procedia Computer Science*, vol. 215, pp. 422–431, 2022.
[33] S. Dong, P. Wang, and K. Abbas, “A survey on deep learning and its applications,” *Computer Science Review*, vol. 40, p. 100379, 2021.
[34] F. Piccialli, V. Di Somma, F. Giampaolo, S. Cuomo, and G. Fortino, “A survey on deep learning models and applications,” *Journal of Big Data*, vol. 8, pp. 1–54, 2021.
[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
[36] R. Newbury, M. Gu, L. Chumbley, A. Mousavian, C. Eppner, J. Leitner, J. Bohg, A. Morales, T. Asfour, D. Kragic, D. Fox, and A. Cosgun, “Deep learning approaches to grasp synthesis: A review,” *IEEE Trans. Robotics*, 2023.
[Online]. Available: https://doi.org/10.1109/TRO.2023.3280597
[37] T. Anderson and R. Anderson, “Applications of machine learning to student grade prediction in quantitative business courses,” *Global J. Business Pedagogy*, vol. 1, no. 3, pp. 13–22, 2017.
[38] H. Waheed, S. U. Hassan, N. R. Aljohani, A. Rauf, and M. H. Saeed, “Predicting academic performance of students from VLE big data using deep learning models,” *Computers in Human Behavior*, vol. 104, p. 106189, 2020.
[39] A. Anggrawan, H. Hairani, and C. Satria, “Improving SVM classification performance on unbalanced student graduation time data using SMOTE,” *Int. J. Inf. Educ. Technol.*, vol. 13, no. 2, pp. 149–153, 2023.
[40] N. Thai-Nghe, T. Horváth, and L. Schmidt-Thieme, “Factorization models for forecasting student performance,” in *Proc. 4th Int. Conf. Educ. Data Mining*, 2011, pp. 11–20.
[41] N. T. Uyên and N. M. Tâm, “Dự đoán kết quả học tập của sinh viên bằng kỹ thuật khai phá dữ liệu,” Tạp chí Khoa học Trường Đại học Vinh, no. 48, pp. 68–73, 2019 (in Vietnamese).

[42] M. Wasif, H. Waheed, N. R. Aljohani, and S.-U. Hassan, "Understanding student learning behavior and predicting their performance," in *Cognitive Computing in Technology-Enhanced Learning*, IGI Global, 2019, pp. 1–28.

[43] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, and H. Rangwala, "Predicting student performance using personalized analytics," *Computer*, vol. 49, no. 4, pp. 61–69, 2016.

[44] M. Fei and D.-Y. Yeung, "Temporal models for predicting student dropout in massive open online courses," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, 2015, pp. 256–263.

[45] F. Okubo, T. Yamashita, A. Shimada, and S. I. Konomi, "Students' performance prediction using data of multiple courses by recurrent neural network," in *Proc. Int. Conf. Computers in Education (ICCE*), 2017.

[46] O. Corrigan and A. F. Smeaton, "A course agnostic approach to predicting student success from VLE log data using recurrent neural networks," in *Data Driven Approaches in Digital Education: Proc. 12th Eur. Conf. Technol. Enhanced Learning (EC-TEL)*, 2017.

[47] V. Christou, I. Tsoulos, V. Loupas, A. T. Tzallas, C. Gogos, P. S. Karvelis, N. Antoniadis, E. Glavas, and N. Giannakeas, "Performance and early drop prediction for higher education students using machine learning," *Expert Syst. Appl.*, vol. 225, p. 120079, 2023.

[48] A. Sapkota and S. Kaur, "Enhancing student success: Predictive modeling of graduation and dropout rates in university management using machine learning," *Springer Nature Singapore*, 2025.

[Online]. Available: https://doi.org/10.1007/s11126-025-09587-2

[49] D. Hammoudi Halat, A.‐S. G. Abdel‐Salam, A. Bensaid, A. Soltani, L. Alsarraj, R. Dalli, and A. Malki, "Use of machine learning to assess factors affecting progression, retention, and graduation in first-year health professions students in Qatar: A longitudinal study," *BMC Med. Educ.*, vol. 23, no. 1, p. 123, 2023.

[50] F. Politeknik Negeri Lampung, R. A. Anggraini, and Y. D. Putri, "Student graduation prediction using machine learning algorithms: Application of linear and logistic regression on educational factors," *Routers: J. Sistem dan Teknol. Inf.*, vol. 3, no. 1, pp. 1–9, 2025.

[51] B. Alnasyan, M. Basheri, and M. Alassafi, “The power of deep learning techniques for predicting student performance in virtual learning environments: a systematic literature review,” *Computers and Education: Artificial Intelligence*, vol. 6, art. no. 100231, 2024. [Online]. Available: https://doi.org/10.1016/j.caeai.2024.100231

[52] Z. Iqbal, M. Ahmad, M. A. Khan, A. Ali, and M. S. Khattak, “Early student grade prediction: An empirical study,” in *Proc. 2nd Int. Conf. Adv. Comput. Sci. (ICACS)*, 2019.

[53] S. Alturki, L. Cohausz, and H. Stuckenschmidt, “Predicting master’s students’ academic performance: An empirical study in Germany,” *Smart Learning Environments*, vol. 9, no. 1, pp. 1–22, 2022.

[54] N. T. T. An, N. V. Thành, Đ. T. K. Oanh, and N. T. N. Thứ, “Những nhân tố ảnh hưởng kết quả học tập của sinh viên năm I-II Trường Đại học Kỹ thuật - Công nghệ Cần Thơ,” *Tạp chí Khoa học Trường Đại học Cần Thơ*, pp. 46–82, 2016. (in Vietnamese)

[55] N. T. Nghe, “Hệ thống dự đoán kết quả học tập của sinh viên sử dụng thư viện hệ thống gợi ý mã nguồn mở Mymedialite,” in *Kỷ yếu Hội thảo Quốc gia về CNTT, Trường Đại học Cần Thơ*, 2013. (in Vietnamese)

[56] N. T. Nghe and T. Q. Định, “Hệ thống hỗ trợ tư vấn tuyển sinh đại học,” *Tạp chí Khoa học Trường Đại học Cần Thơ*, p. 152, 2015 . (in Vietnamese)

[57] L. H. Sang, T. T. Dien, N. T. Nghe, and N. T. Hai, “Dự đoán kết quả học tập bằng kỹ thuật học sâu với mạng nơ-ron đa tầng,” *Tạp chí Khoa học Trường Đại học Cần Thơ*, vol. 56, no. 3, pp. 20–28, 2020. (in Vietnamese)

[58] T. T. Dien, S. H. Luu, N. Thanh-Hai, and N. Thai-Nghe, “Deep learning with data transformation and factor analysis for student performance prediction,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 663–670, 2020. [Online]. Available: https://doi.org/10.14569/IJACSA.2020.0110886

[59] N. Iam-On and T. Boongoen, “Generating descriptive model for student dropout: A review of clustering approach,” *Hum.-Cent. Comput. Inf. Sci.*, vol. 7, no. 1, 2017.

[60] S. I. Popoola, A. A. Atayero, J. A. Badejo, T. M. John, J. A. Odukoya, and D. O. Omole, “Learning analytics for smart campus: Data on academic performances of engineering undergraduates in Nigerian private university,” *Data Brief*, vol. 17, pp. 76–94, 2018.

[61] T. Hasan, M. M. Hasan, and T. Manzoor, “Student performance metrics dataset,” *Mendeley Data*, 2024.
[62] M. V. Martins, D. Tolledo, J. Machado, L. M. Baptista, and V. Realinho, “Early prediction of student’s performance in higher education: A case study,” in *Trends Appl. Inf. Syst. Technol*., vol. 1, pp. 166–175, 2021.
[63] P. Cortez and A. Silva, “Using data mining to predict secondary school student performance,” in *Proc. 5th Future Business Technol. Conf. (FUBUTEC)*, pp. 5–12, 2008.
[64] D. Boud, Developing student autonomy in learning, 2nd ed. London, U.K.: Kogan, 1988.
[65] I. Saleh and S.-I. Kim, “A fuzzy system for evaluating students’ learning achievement,” *Expert Syst. Appl*., vol. 36, no. 3, pp. 6236–6243, 2009.
[66] I. A. Hameed, “Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems,” *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7135–7142, 2011.
[67] R. Biswas, “An application of fuzzy sets in students’ evaluation,” *Fuzzy Sets Syst*., vol. 74, no. 2, pp. 187–194, 1995.
[68] I. A. Hameed and C. G. Sørensen, “Fuzzy systems in education: A more reliable system for student evaluation,” *in Fuzzy Systems*, A. T. Azar, Ed. INTECH, 2010, pp. 45–67.
[69] M. Voskoglou and S. Broumi, “A hybrid method for the assessment of analogical reasoning skills,” *J. Fuzzy Ext. Appl*., vol. 3, no. 2, pp. 152–157, 2022.
[70] U. Vakkas, “Q-neutrosophic soft graphs in operations management and communication network,” *Soft Comput*., vol. 25, pp. 8441–8459, 2021.
[71] P. A. Ejegwa and D. Zuakwagh, “Fermatean fuzzy modified composite relation and its application in pattern recognition,” *J. Fuzzy Ext. Appl.*, vol. 3, no. 2, pp. 140–151, 2022.
[72] F. Smarandache, Neutrosophy: Neutrosophic probability, set, and logic. Gallup, NM, USA: American Research Press, 1998.
[73] F. Smarandache, Neutrosophic precalculus and neutrosophic calculus. Rehoboth, NM, USA: Europa-Nova, 2015.

[74] S. Mayukh, K. Ilanthenral, and W. B. Vasantha, “Comparison of neutrosophic approach to various deep learning models for sentiment analysis,” *Knowl.-Based Syst.*, vol. 223, p. 107058, 2021.

[75] C. Villani, “Optimal transport: Old and new”. Berlin, Germany: Springer, 2008.

[76] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” *arXiv preprint*, arXiv:1701.07875, 2017.

[77] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, vol. 27, 2014. [Online]. Available: https://arxiv.org/abs/1406.2661

[78] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, “Do transformers really perform badly for graph representation?” in *Advances in Neural Information Processing Systems*, vol. 34, pp. 28877–28888, 2021.

[79] M. Mirza, “Conditional generative adversarial nets,” *arXiv preprint*, arXiv:1411.1784, 2014.

[80] L. Xu, M. Skoularidou, A. Cuesta‐Infante, and K. Veeramachaneni, “Modeling tabular data using conditional GAN,” in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Số: 821/QĐ-HVKHCN

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 25 tháng 08 năm 2025

QUYẾT ĐỊNH
Về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện

GIÁM ĐỐC
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Căn cứ Quyết định số 364/QĐ-VHL ngày 01/03/2025 của Chủ tịch Viện Hàn lâm Khoa học và Công nghệ Việt Nam về việc ban hành Quy chế tổ chức và hoạt động của Học viện Khoa học và Công nghệ;

Căn cứ Thông tư số 18/2021/TT-BGDĐT ngày 28/06/2021 của Bộ Giáo dục và Đào tạo ban hành Quy chế tuyển sinh và đào tạo trình độ tiến sĩ;

Căn cứ Quyết định số 1968/QĐ-HVKHCN ngày 28/12/2021 của Giám đốc Học viện Khoa học và Công nghệ về việc ban hành Quy định đào tạo trình độ tiến sĩ;

Căn cứ Quyết định số 903/QĐ-HVKHCN ngày 25/05/2022 của Giám đốc Học viện Khoa học và Công nghệ về việc công nhận nghiên cứu sinh đợt 1 năm 2022;

Xét đề nghị của Trưởng phòng Đào tạo.

QUYẾT ĐỊNH:

Điều 1. Thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện cho nghiên cứu sinh Nguyễn Thị Kim Sơn với đề tài,

Tên tiếng Việt: **Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học**

Tên tiếng Anh: **Research on the application of deep learning models for predicting learners' academic performance**

Ngành: Hệ thống thông tin Mã số: 9 48 01 04

Danh sách thành viên Hội đồng đánh giá luận án kèm theo Quyết định này.

Điều 2. Hội đồng có trách nhiệm đánh giá luận án tiến sĩ theo đúng quy chế hiện hành của Bộ Giáo dục và Đào tạo, của Học viện Khoa học và Công nghệ.Quyết định có hiệu lực tối đa 90 ngày kể từ ngày ký. Hội đồng tự giải thể sau khi hoàn thành nhiệm vụ.

Điều 3. Trưởng phòng Tổ chức - Hành chính, Trưởng phòng Đào tạo, Trưởng phòng Kế toán, các thành viên có tên trong danh sách Hội đồng và nghiên cứu sinh có tên tại Điều 1 chịu trách nhiệm thi hành Quyết định này./.

***Nơi nhận*:**
- Như Điều 3;
- Lưu hồ sơ NCS;
- Lưu: VT, ĐT.PQ.15.

GIÁM ĐỐC

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

GS. TS. Vũ Đình Lãm

DANH SÁCH HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ
CẤP HỌC VIỆN

(Kèm theo quyết định số 821/QĐ-HVKHCN ngày 25/08/2025 của Giám đốc Học viện Khoa học và Công nghệ)

Cho luận án của nghiên cứu sinh: Nguyễn Thị Kim Sơn

Tên tiếng Việt: **Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học**

Tên tiếng Anh: **Research on the application of deep learning models for predicting learners' academic performance**

Ngành: Hệ thống thông tin Mã số: 9 48 01 04

Người hướng dẫn chính: PGS.TS. Nguyễn Hữu Quỳnh, Trường Đại học CMC

Người hướng dẫn phụ: PGS.TS. Ngô Quốc Tạo, Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TT	Họ và tên, học hàm, học vị	Ngành	Cơ quan công tác	Trách nhiệm trong Hội đồng
1	PGS.TS. Nguyễn Long Giang	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Chủ tịch
2	PGS.TS. Bùi Thu Lâm	Khoa học máy tính	Học viện Kỹ thuật Mật mã, Ban Cơ yếu Chính phủ	Phản biện 1
3	TS. Nguyễn Như Sơn	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Phản biện 2
4	TS. Trần Đức Nghĩa	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Ủy viên - Thư ký
5	PGS.TS. Nguyễn Văn Long	Hệ thống thông tin	Trường Đại học Giao thông Vận tải, Bộ Xây dựng	Ủy viên
6	PGS.TS. Đỗ Trung Tuấn	Hệ thống thông tin	Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội	Ủy viên
7	PGS.TS. Phạm Văn Hải	Hệ thống thông tin	Trường Công nghệ thông tin và Truyền thông, Đại học Bách khoa Hà Nội	Ủy viên

Hội đồng gồm 07 thành viên./.

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập - Tự do - Hạnh phúc

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài luận án: **Nghiên cứu ứng dụng một số mô hình sử dụng học sau trong dự đoán kết quả học tập của người học**

Ngành: Hệ thống thông tin

Mã số ngành: *9.48.01.04*

Nghiên cứu sinh: ***Nguyễn Thị Kim Sơn***

Người hướng dẫn: 1. PGS.TS Nguyễn Hữu Quỳnh, 2. PGS.TS Ngô Quốc

Tạo Người nhận xét: *PGS.TS Nguyễn Long Giang, Chủ tịch HD*

Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KHCNVN

NỘI DUNG NHẬN XÉT

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Trong bối cảnh chuyển đổi số và cách mạng công nghiệp 4.0, các trường đại học ngày càng ứng dụng mạnh mẽ hệ thống quản lý học tập và đào tạo trực tuyến, tạo ra khối lượng dữ liệu học tập rất lớn của sinh viên. Tuy nhiên, việc khai thác dữ liệu này để dự đoán kết quả học tập vẫn còn hạn chế. Việc áp dụng các mô hình trí tuệ nhân tạo, đặc biệt là học sâu, nhằm dự đoán chính xác kết quả học tập của sinh viên đang được các nhà khoa học trong và ngoài nước quan tâm nghiên cứu. Về mặt khoa học, đề tài góp phần bổ sung cơ sở lý luận và thực nghiệm cho việc ứng dụng học sâu trong khai phá dữ liệu giáo dục và phân tích học tập ở bậc đại học. Về mặt thực tiễn, kết quả nghiên cứu có thể hỗ trợ giảng viên và nhà quản lý trong việc theo dõi quá trình học tập, phát hiện sớm nhóm sinh viên có nguy cơ kết quả thấp để đưa ra giải pháp hỗ trợ phù hợp, góp phần nâng cao chất lượng đào tạo trong các trường đại học hiện nay. Do vậy, đề tài luận án với mục tiêu phát triển các mô hình dự đoán kết quả học tập dựa trên học sâu của *NCS Nguyễn Thị Kim Sơn* là có ý nghĩa khoa học và ý nghĩa thực tiễn cao.

2. Sự không trùng lặp của đề tài nghiên cứu; vấn đề trích dẫn tài liệu tham khảo

Các kết quả nghiên cứu của NCS không trùng lặp so với công trình, luận án đã công bố ở trong và ngoài nước. Luận án này được viết trên cơ sở các công trình khoa học của chính NCS và cộng sự, đã tham khảo 80 tài liệu khoa học chính thống. Các tài liệu tham khảo về cơ bản phù hợp, cập nhật và đã được trích dẫn đầy đủ trong luận án.

3. Sự phù hợp của tên đề tài với nội dung, giữa nội dung với chuyên ngành

Đề tài luận án phù hợp với nội dung nghiên cứu, hướng nghiên cứu phù hợp với ngành đào tạo "Hệ thống thông tin", mã số: 9480104. Các kết quả thu được của NCS có thể áp dụng trong việc xây dựng hệ thống thông tin dự đoán kết quả học tập của sinh viên.

4. Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu

NCS đã nghiên cứu lý thuyết thông qua khảo sát, phân tích và đánh giá các tài liệu khoa học đã công bố liên quan đến bài toán dự báo sớm kết quả học tập của sinh viên dựa trên học máy và học sâu. Sau đó, NCS đề xuất kết hợp mô hình học sâu với lý thuyết Neutrosophy và các kiến trúc như sinh dữ liệu, cấu trúc đồ thị, cấu trúc transformer để giải quyết bài toán dữ liệu không chắc chắn, nhỏ hay mất cân bằng. Các mô hình đề xuất trong luận án đã được NCS phân tích đánh giá và kiểm chứng bằng thực nghiệm trên các bộ dữ liệu tự thu thập hoặc công bố công khai, sử dụng các độ đo thông dụng trong lĩnh vực học máy. Phương pháp nghiên cứu của NCS là hợp lý và đáng tin cậy.

5. Kết quả nghiên cứu mới của tác giả

Kết quả của nghiên cứu đã góp phần bổ sung tri thức khoa học, thúc đẩy phát triển các hệ thống thông minh, khai phá dữ liệu, hệ chuyên gia hoặc ra quyết định mờ. Luận án của NCS có các kết quả chính như sau:

- Xây dựng được hai mô hình NeutroDL và NeutroGNT dựa trên học sâu kết hợp lý thuyết Neutrosophy để dự đoán sớm điểm trung bình học kỳ (*SGPA-Semester Grade Point Average*) của sinh viên, với khả năng xử lý dữ liệu thiếu và không chắc chắn.
- Xây dựng được hai mô hình lai LATCGAd và AWG-GC dựa trên sự kết hợp các kiến trúc như sinh dữ liệu, đồ thị và transformer để dự đoán phân loại tốt nghiệp dài hạn trong điều kiện dữ liệu nhỏ và mất cân bằng.

Ngoài hai đóng góp chính nói trên đây, NCS đã phát triển được ba bộ bộ dữ liệu từ các trường đại học Việt Nam để thực nghiệm các mô hình đề xuất. Các kết quả của luận án là mới và đã đáp ứng đầy đủ mục tiêu nghiên cứu đặt ra cho đề tài.

6. Về ưu điểm và nhược điểm của luận án

- Luận án có cấu trúc hợp lý, nội dung nghiên cứu và các kết quả thu được đã được NCS trình bày khá nghiêm túc, rõ ràng và lô gíc.
- Luận án được viết bằng tiếng Anh, ít lỗi in ấn, bảng biểu, hình vẽ rõ ràng.
- Mục CONTENTS (trang viii) của luận án cần được đưa lên trước SYMBOLS AND ABBREVIATIONS theo qui định của cơ sở đào tạo. Bảng SYMBOLS AND ABBREVIATIONS (trang iii) cần được sắp xếp theo vần abc để dễ tra cứu.

- Các thuật toán cần được đánh số theo chương như hình và bảng, ví dụ Algorithm 1. (trang 48) thành Algorithm 2.1,...
- Phân tích thêm Thông tư 42/2021 "Quy định về cơ sở dữ liệu giáo dục và đào tạo" của BGD-ĐT trước khi kết luận "However, in education, there is currently a lack of large, standardized, ..." (trang 2).
- Nên bổ sung vào Research Subjects (trang 3): Ngoài các bài toán dự báo kết quả học tập của sinh viên, đối tượng nghiên cứu còn là các mô hình học sâu.
- Mục 5. Key contributions of the dissertation (trang 5): Mặc dù đã có mô tả "From an information systems perspective,...", tuy nhiên nên bổ sung sơ đồ khối trực quan Hệ thống thông tin dự báo kết quả học tập của sinh viên, từ đó chỉ ra các kết quả nghiên cứu của luận án phục vụ cho các khối chức năng nào.
- Các mục 2.4. Appendix to Chapter 2 (trang 64), 3.4. Appendix to Chapter 3 (trang 114) cần chuyển về Chương 1 hoặc để cuối luận án.
- Phân tích sâu hơn độ phức tạp của các thuật toán đề xuất trong luận án. NCS mới quan tâm đến độ chính xác, chưa quan tâm đến thời gian và bộ nhớ cần thiết khi thực hiện các thuật toán.
- Nên có thêm thực nghiệm các thuật toán Chương 2 với các bộ dữ liệu khác nhau để đánh giá khách quan hơn.

7. Về các công trình đã công bố của NCS

NCS có 06 công trình khoa học đã công bố và 02 công trình khoa học đã chấp nhận đăng, viết cùng tập thể giáo viên hướng dẫn và các đồng nghiệp, trong đó, có 05 bài báo trên các tạp chí quốc tế và 03 bài báo trên các tạp chí trong nước đều có phản biện và được HĐCDGSNN tính điểm. Nội dung các công trình khoa học đã công bố là phù hợp, thống nhất với nội dung thực tế của luận án.

8. Kết luận

Bản tóm tắt luận án đã phản ánh trung thành nội dung cơ bản của luận án. Luận án của *NCS Nguyễn Thị Kim Sơn* đã đáp ứng đầy đủ các yêu cầu cả về nội dung và hình thức đối với một luận án Tiến sĩ theo các qui chế hiện hành. Tôi đồng ý cho NCS bảo vệ luận án của mình trước Hội đồng cấp Học viện để nhận bằng Tiến sĩ ngành Hệ thống thông tin.

Hà Nội, ngày 1 tháng 9 năm 2025

Người nhận xét

PGS.TS Nguyễn Long Giang

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ

Họ và tên người viết nhận xét luận án: Bùi Thu Lâm

Học hàm, học vị: PGS TS

Cơ quan công tác: Học viện Kỹ thuật Mật mã/Ban Cơ yếu chính phủ

Họ và tên nghiên cứu sinh: NGUYỄN THỊ KIM SƠN

Tên đề tài luận án: NGHIÊN CỨU ỨNG DỤNG MỘT SỐ MÔ HÌNH SỬ DỤNG HỌC SÂU TRONG DỰ ĐOÁN KẾT QUẢ HỌC TẬP CỦA NGƯỜI HỌC

Ý KIẾN NHẬN XÉT

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài:

Đề tài có tính cần thiết cao trong bối cảnh chuyển đổi số giáo dục tại Việt Nam và trên thế giới, nơi dữ liệu học tập ngày càng phong phú nhưng chưa được nghiên cứu ứng dụng và khai thác hiệu quả để dự đoán kết quả học tập. Việc ứng dụng học sâu (deep learning) để dự đoán điểm trung bình học kỳ và phân loại tốt nghiệp sớm là phù hợp với xu hướng chuyển đổi số hiện nay, giúp phát hiện sớm rủi ro thất bại học tập.

Ý nghĩa khoa học thể hiện ở việc đề xuất các mô hình lai như NeutroDL, NeutroGNT, LATCGAd và AWG-GC, kết hợp lý thuyết trung tính (neutrosophy) với các kiến trúc học sâu hiện đại (Transformer, Graphormer), cải thiện độ chính xác dự đoán lên đến 98.54% và R^2 lên 96.05% trên dữ liệu thực tế.

Đề tài hy vọng hỗ trợ sinh viên điều chỉnh kế hoạch học tập, giảng viên can thiệp kịp thời, và quản lý giáo dục tối ưu hóa chính sách, đặc biệt với dữ liệu từ các trường đại học Việt Nam như HNMU và VNU.

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận văn, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo.

Đề tài không trùng lặp với các công trình trước, vì tập trung vào việc tích hợp neutrosophy với học sâu để xử lý bất định trong dữ liệu giáo dục Việt Nam, khác biệt so với các nghiên cứu quốc tế như Waheed et al. (2020) hay Okubo et al. (2017) chỉ dùng RNN/LSTM cơ bản, hoặc trong nước như Sang et al. (2020) sử dụng MLP đơn giản.

Các mô hình lai như NeutroGNT và AWG-GC có tính mới; người đọc chưa thấy trong tài liệu tham khảo.

Dữ liệu thực từ HNMU và VNU được khai báo rõ ràng, kết quả thí nghiệm trung thực với lỗi chuẩn (standard deviation).

Trích dẫn rõ ràng, đầy đủ với hơn 70 tài liệu, bao quát từ kinh điển đến gần đây (2025), sử dụng định dạng chuẩn và không có dấu hiệu đạo văn.

3. Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với chuyên ngành và mã số chuyên ngành.

Tên đề tài phù hợp hoàn toàn với nội dung, tập trung vào "ứng dụng học sâu" trong dự đoán kết quả học tập, với các mô hình cụ thể như NeutroDL và LATCGAd.

Nội dung phù hợp với chuyên ngành Hệ thống thông tin (mã 9 48 01 04), vì nghiên cứu xây dựng hệ thống dự đoán dựa trên dữ liệu giáo dục, tích hợp AI vào hệ thống thông tin giáo dục, hỗ trợ phân tích và ra quyết định. Các chương từ tổng quan đến thí nghiệm đều liên kết chặt chẽ với lĩnh vực, nhấn mạnh Learning Analytics như một phần của Information Systems.

4. Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu.

Phương pháp có độ tin cậy cao nhờ sử dụng dữ liệu thực tế từ các trường đại học Việt Nam và quốc tế, kết hợp với xác thực chéo (10-fold cross-validation) và các chỉ số đánh giá chuẩn (Accuracy, Precision, Recall, F1-score, MSE, RMSE, R^2).

Thí nghiệm lặp lại 10 lần với lỗi chuẩn cho thấy kết quả ổn định.

Tính hiện đại nổi bật: tích hợp neutrosophy (Smarandache, 1998) với học sâu hiện đại như Transformer (Vaswani, 2017), Graphormer (Ying et al., 2021), và GAN biến thể (CGAN, WGAN), phù hợp với xu hướng hybrid models trong EDM.

Các phương pháp như AdaLN và tiêm nhiễu (noise-injection) tăng cường ổn định, phản ánh tiến bộ gần đây trong deep learning cho dữ liệu nhỏ và bất định.

5. **Kết quả nghiên cứu mới của tác giả; những đóng góp mới cho sự phát triển khoa học chuyên ngành; đóng góp mới phục vụ cho sản xuất, kinh tế, quốc phòng, xã hội và đời sống. Ý nghĩa khoa học, giá trị và độ tin cậy của những kết quả đó.**

Kết quả mới: Đề xuất 4 mô hình lai (NeutroDL, NeutroGNT, LATCGAd, AWG-GC) cho dự đoán SGPA và phân loại tốt nghiệp, đạt độ chính xác lên 98.54% và F1-score 99.25% trên dữ liệu thực.

Về đóng góp khoa học: Luận án mở rộng lý thuyết trung tính vào học sâu, cải thiện xử lý bất định trong dữ liệu giáo dục, bổ sung cho chuyên ngành Hệ thống thông tin bằng khung tích hợp cho phân tích học tập LA.

Về đóng góp thực tiễn: Luận án hỗ trợ dự đoán sớm rủi ro học tập, tối ưu hóa quản lý giáo dục tại Việt Nam, góp phần vào kinh tế - xã hội qua nâng cao chất lượng nguồn nhân lực.

Có ý nghĩa khoa học cao, giá trị thực tiễn rõ ràng với dữ liệu địa phương, độ tin cậy vững nhờ thí nghiệm lặp và so sánh với baseline.

6. **Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án.**

Ưu điểm:

Nội dung được trình bày chi tiết, tập trung vào vấn đề thực tiễn với dữ liệu Việt Nam, đề xuất mô hình mới hiệu quả. Kết cấu logic, giới thiệu rõ vấn đề, tổng quan toàn diện, phương pháp hiện đại, thí nghiệm chi tiết. Bảng biểu, đồ thị rõ ràng, ngôn ngữ khoa học, 142 trang cân đối.

Nhược điểm:

Về nội dung:

(1) Dữ liệu nhỏ (HNMU1: 932, HNMU2: 551, VNU: 271), dẫn đến R^2 thấp ở một số case, thiếu tổng quát hóa cho trường hợp lớn hơn.

(2) So sánh baseline chưa đầy đủ, thiếu mô hình SOTA cho dữ liệu giáo dục.

(3) Chưa phân tích sâu yếu tố đặc thù Việt Nam, dẫn đến mô hình chưa tối ưu hóa địa phương; Đặc biệt, mô hình dự báo ở Chương 2 dừng lại ở mức độ đơn biến, chưa tính tới yếu tố đa biến.

Về kết cấu: Chương 2 và 3 lặp lại một số phần (như GAN), thiếu liên kết chặt chẽ giữa dự đoán SGPA và loại tốt nghiệp.

Về hình thức:

Không nên phân biệt nghiên cứu trong và ngoài nước (trang 16)

Nên có các giải thích về sự hợp lý trong các lai ghép mô hình

Các ký hiệu toán học ở chương 2 nên chuẩn hóa cho thống nhất và dùng văn phong academic hơn

7. Nội dung của luận án đã được công bố trên tạp chí, kỷ yếu Hội nghị Khoa học nào và giá trị của các công trình đã công bố (cấp công bố WoS (SSCI, SCI/E, ESCI ...), Scopus, quốc tế có phản biện, tạp chí trong nước được tính điểm theo Hội đồng Giáo sư nhà nước ... và xếp hạng SCIMAGO).

Luận án có 08 công trình công bố, bao gồm [CT1] đến [CT8], đăng trên các hội nghị và tạp chí quốc tế/nội địa như: VNICT 2024, MCO 2025, và các tạp chí Scopus/Q2.

Các bài chứng minh mô hình mới, đạt chỉ số Scopus (Q2/Q3), và tính điểm theo HĐGSNN (0.5-1.0 điểm/bài), phản ánh đóng góp thực tiễn và khoa học.

8. Kết luận:

- Mức độ đáp ứng các yêu cầu đối với một luận án tiến sĩ chuyên ngành:

  Luận án đáp ứng tốt, với nội dung mới mẻ, phương pháp hiện đại, và ý nghĩa thực tiễn.

- Bản tóm tắt luận án có phản ánh trung thành nội dung cơ bản của luận án không:

  Có, tóm tắt phản ánh chính xác các mô hình, phương pháp và kết quả.

- Luận án có thể đưa ra bảo vệ cấp Học viện để nhận bằng Tiến sĩ được hay không:

  Có, luận án đủ chất lượng để bảo vệ.

Hà Nội, ngày 7 tháng 10 năm 2025

Người viết nhận xét

(Ký và ghi rõ họ và tên

Bùi Thu Lâm

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập - Tự do - Hạnh phúc

BẢN NHẬN XÉT (PHẢN BIỆN) LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Tên đề tài: **Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học**

Ngành: **Hệ thống thông tin** Mã số: **9.48.01.04**

Nghiên cứu sinh: **Nguyễn Thị Kim Sơn**

Người hướng dẫn: PGS.TS. Nguyễn Hữu Quỳnh, PGS.TS. Ngô Quốc Tạo

Người phản biện: TS. Nguyễn Như Sơn

Cơ quan công tác: Viện Công nghệ thông tin – Viện Hàn lâm KHCN Việt Nam

1, Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Luận án tập trung vào việc dự đoán kết quả học tập của sinh viên dựa trên dữ liệu thu thập trong quá trình học, giúp phát hiện sớm các nguy cơ thất bại và triển khai các biện pháp can thiệp kịp thời, trực tiếp hỗ trợ mục tiêu giáo dục hiện đại, bao gồm cá nhân hóa trải nghiệm học tập và nâng cao tỷ lệ tốt nghiệp. Để giải quyết vấn đề này, luận án lựa chọn các mô hình học sâu đang phổ biến làm nền tảng, kết hợp với các kỹ thuật như tăng cường dữ liệu, chọn lọc đặc trưng và tối ưu siêu tham số. Đồng thời, phát triển mô hình lai (kết hợp deep learning với machine learning truyền thống, hoặc kết hợp nhiều kiến trúc deep learning) là một hướng đi triển vọng, vừa tận dụng sức mạnh biểu diễn dữ liệu, vừa cải thiện tính giải thích của mô hình. Vì vậy, việc nghiên cứu của luận án là hết sức cần thiết, có tính thời sự, ý nghĩa khoa học và thực tiễn.

2, Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo

Theo hiểu biết của người nhận xét, kết quả của luận án không trùng lặp với các công trình, luận án đã công bố ở trong và ngoài nước. Luận án có trích dẫn rõ ràng 80 tài liệu tham khảo, các tài liệu có tính mới cập nhật trong những năm gần đây.

3, Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã số

Tên đề tài cơ bản phù hợp với nội dung luận án. Nội dung luận án phù hợp với chuyên ngành và mã số chuyên ngành Hệ thống thông tin.

4, Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu

Phương pháp nghiên cứu của luận án tập trung vào nghiên cứu lý thuyết và thực nghiệm. Các đề xuất, cải tiến trong luận án đều được thử nghiệm nhằm minh chứng cho tính hiệu quả của phương pháp. Dữ liệu thực nghiệm được thu thập từ các đơn vị đào tạo thực tế. Do đó, phương pháp nghiên cứu sử dụng trong luận án là hợp lý, bảo đảm tính khoa học và độ tin cậy.

5, Kết quả nghiên cứu mới của tác giả

Luận án này đã giải quyết bài toán dự đoán kết quả học tập của sinh viên trong điều kiện không chắc chắn, thiếu hụt dữ liệu và mất cân bằng lớp, vốn đặc trưng cho các môi trường giáo dục thực tiễn. Đóng góp mới của tác giả tập trung vào các nội dung chính sau đây:

1. Đề xuất hai mô hình NeutroDL và NeutroGNT, tích hợp học sâu với lý thuyết neutrosophic nhằm xử lý dữ liệu thiếu và không chắc chắn để dự đoán kết quả học tập của sinh viên.

2. Đề xuất hai mô hình lai kết hợp kiến trúc sinh dữ liệu và mạng dựa trên đồ thị trong bối cảnh dữ liệu giáo dục mất cân bằng và quy mô nhỏ: LATCGAd và AWG-GC để dự đoán phân loại tốt nghiệp đài hạn có độ chính xác cao hơn so với các mô hình đối sánh.

3. Phát triển bộ dữ liệu mở rộng và quy trình phân tích phục vụ ứng dụng trong giáo dục.

6, Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án

Luận án bố cục làm 3 chương gồm 120 trang nội dung và các phần liên quan về công trình công bố và tài liệu tham khảo, phù hợp với cấu trúc của luận án tiến sĩ. các chương được phân chia hợp lý các nội dung liên quan, trình bày chi tiết rõ ràng các mô hình, kỹ thuật đề xuất trong phạm vi nghiên cứu của luận án. Hệ thống 80 tài liệu tham khảo được trích dẫn đúng chuẩn, các tài liệu có tính cập nhật mới trong những năm gần đây.

Tuy nhiên cần xem xét một số ý kiến sau đây để chỉnh sửa luận án cho hợp lý hơn nếu có thể được:

- Về quy mô dữ liệu, do việc thu thập thực tế sẽ mất nhiều thời gian nên hiện tại có quy mô còn chưa lớn, sẽ ảnh hưởng chất lượng mô hình?.

- Các thuật toán đề xuất hiện chưa có đánh giá sơ bộ về độ phức tạp, thời gian tính toán, có thể xem như phụ thuộc vào độ phức tạp của các mô hình cơ bản được sử dụng?.

- Các kết quả hiện không đánh giá thời gian chạy, tuy nhiên thực nghiệm nên có các thông số về cấu hình phần cứng để dùng cho việc tham khảo khác sau này?

- Trình bày thuật toán nên theo format truyền thống: Input, Output (không đánh số dòng lệnh). Begin – End mới bắt đầu đánh số.

- Chương ba cần nói thêm nội dung công bố ở CT2.

- Rà soát về các chú thích Hình vẽ, Bảng biểu cố gắng không nên ngắt trang ở cả 2 bản luận án và tóm tắt.

7, Nội dung luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố

Các kết quả nghiên cứu này đã được công bố trên các tạp chí và được trao đổi trong các báo cáo tại những hội thảo khoa học chuyên ngành bao gồm 08 công trình khoa học, trong đó có 01 bài báo tạp chí quốc tế thuộc danh mục SCIE, 03 bài báo thuộc danh mục Scopus, 02 bài báo tạp chí khác và 01 bài báo kỷ yếu hội thảo quốc tế, các tạp chí và hội thảo có chất lượng cao.

NCS là tác giả chính đầu tiên trong 07 công trình. Nội dung của các công trình phù hợp với nội dung nghiên cứu của luận án và là các phần kết quả chính của luận án.

8, Kết luận

Luận án của NCS Nguyễn Thị Kim Sơn đáp ứng các yêu cầu đối với một luận án tiến sĩ chuyên ngành Hệ thống thông tin sau khi chỉnh sửa theo các góp ý ở trên. Bản tóm tắt luận án phản ánh trung thành nội dung cơ bản của luận án. Luận án có thể đưa ra bảo vệ cấp Học viện và thực hiện các bước tiếp theo để nhận học vị tiến sĩ.

Hà Nội, ngày 4 tháng 9 năm 2025

Người nhận xét

TS. Nguyễn Như Sơn

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ
(Cấp Học viện)

Tên đề tài luận án: Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học/ Reseach on the application of deep learning models for predicting learners'academic performance

Chuyên ngành: Hệ thống thông tin Mã số: 9 48 01 04

Nghiên cứu sinh: **Nguyễn Thị Kim Sơn**

Người hướng dẫn 1: PGS.TS. Nguyễn Hữu Quỳnh

Người hướng dẫn 2: PGS.TS Ngô Quốc Tạo

Người nhận xét: **TS. Trần Đức Nghĩa**

Đơn vị công tác: Viện Công nghệ thông tin, Viện Hàn lâm KH & CN Việt Nam

Chức trách trong hội đồng: Ủy viên thư ký

I. NỘI DUNG NHẬN XÉT

1. Tính cấp thiết, thời sự ý nghĩa khoa học của đề tài luận án:

Đề tài LA nghiên cứu việc ứng dụng mô hình học sâu trong dự đoán kết quả học tập của người học có ý nghĩa khoa học ở việc khai thác khả năng mô hình hóa quan hệ phi tuyến giữa dữ liệu học tập và kết quả đầu ra. Kết quả nghiên cứu cung cấp cơ sở cho việc xây dựng các hệ thống cảnh báo sớm và hỗ trợ cá nhân hóa hoạt động học tập. Ngoài ra, đề tài góp phần bổ sung cho tiếp cận sử dụng trí tuệ nhân tạo trong phân tích dữ liệu giáo dục, đồng thời làm nền tảng cho các ứng dụng trong quản lý và ra quyết định trong môi trường học trực tuyến.

2. Sự phù hợp của đề tài luận án với chuyên ngành đào tạo:

Đề tài luận án có tính khoa học, phù hợp với chuyên ngành đào tạo tiến sĩ Hệ thống thông tin.

3. Sự trùng lặp của đề tài so với công trình khoa học đã công bố:

Nội dung luận án không trùng lặp với các luận án đã bảo vệ và các kết quả nghiên cứu đã công bố trong và ngoài nước.

4. Sự phù hợp của các phương pháp nghiên cứu, độ tin cậy của các kết quả đã đạt được:

Các lập luận, chứng minh và các kết quả đạt được là đáng tin cậy.

5. Những đóng góp mới của đề tài

Luận án nghiên cứu và phát triển các mô hình học máy và học sâu để phân tích dữ liệu giáo dục, nhằm mục đích nâng cao khả năng dự đoán sớm kết quả học tập của sinh viên. Luận án có ý nghĩa khoa học và thực tiễn, những đóng góp mới của luận án cụ thể như sau:

(1) Đề xuất 02 mô hình mới là NeutroDL và NeutroGNT, tích hợp quy trình neutrosophic vào các mô hình học sâu để nâng cao hiệu suất dự đoán GPA sớm.

(2) Đề xuất 02 mô hình lai mới là LATCGAd và AWG-GC, để dự đoán phân loại tốt nghiệp cho sinh viên.

(3) Phát triển 03 tập dữ liệu đa thuộc tính từ nhiều nguồn khác nhau và đề xuất các khuôn khổ phân tích phù hợp với dữ liệu giáo dục.

6. Về các công trình khoa học đã công bố của nghiên cứu sinh liên quan đến nội dung của luận án

Các kết quả của luận án đã được công bố tại 8 hội thảo/tạp chí chuyên ngành có uy tín. Các kết quả công bố có nội dung khoa học và là kết quả chính của luận án.

7. Tính trung thực, minh bạch trong trích dẫn tài liệu.

Đảm bảo.

8. Góp ý các thiếu sót về hình thức, nội dung của luận án mà nghiên cứu sinh cần chỉnh sửa, bổ sung.

Ưu điểm:

Luận án có ý nghĩa lý thuyết và ứng dụng thực tiễn, phương pháp nghiên cứu đáng tin cậy.

Nhược điểm/góp ý:

Trong các nghiên cứu tương lai, nên bổ sung thêm các yếu tố ảnh hưởng đến kết quả học tập khác, ví dụ như thái độ/sự tích cực trong học tập, như vậy việc dự đoán sẽ tốt hơn.

Nên thử nghiệm thêm trên các bộ dữ liệu khác nhau để đảm bảo tính khách quan và sự hiệu quả của đề xuất.

Lưu ý rà soát LA đảm bảo theo mẫu/cấu trúc của Học viện.

II. KẾT LUẬN

Đánh giá về mức độ đạt yêu cầu của luận án:

Luận án "*Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học/ Reseach on the application of deep learning models for predicting learners 'academic performance*" đáp ứng các yêu cầu của Bộ Giáo dục và Đào tạo và cơ sở đào tạo đề ra cho một luận án tiến sĩ chuyên ngành Hệ thống thông tin, mã số: 9 48 01 04.
Đồng ý cho NCS đưa luận án ra bảo vệ tại Hội đồng cấp Học viện để nhận học vị Tiến sĩ.

Hà Nội, ngày 1 tháng 09 năm 2025

Người nhận xét

TS. Trần Đức Nghĩa

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc
****

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ

Đề tài: **Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học**

Ngành: **Hệ thống thông tin;** Mã số: **9.48.01.04**

Họ và tên nghiên cứu sinh: **Nguyễn Thị Kim Sơn**

Người hướng dẫn: PGS.TS. Nguyễn Hữu Quỳnh, trường Đại học CMC; PGS.TS. Ngô Quốc Tạo, Viện Công nghệ thông tin

Cơ sở đào tạo NCS: **Học viện Khoa học và Công nghệ, Viện hàn lâm khoa học và công nghệ Việt Nam**

Người nhận xét: **PGS.TS. Nguyễn Văn Long**

Đơn vị công tác của người nhận xét: **Bộ môn Khoa học Máy tính trường Đại học Giao thông Vận tải.**

Số điện thoại: **0933819869**, Email: **nvlongdt@utc.edu.vn**

1. Ý nghĩa khoa học và thực tiễn của luận án:

Đề tài nghiên cứu có ý nghĩa khoa học góp phần mở rộng lĩnh vực khai phá dữ liệu giáo dục (EDM) và phân tích học tập (Learning Analytics), thông qua việc ứng dụng các mô hình học sâu hiện đại để dự đoán kết quả học tập của người học; đã làm rõ những hạn chế của các mô hình học máy truyền thống vốn chỉ khai thác dữ liệu tĩnh và tuyến tính, từ đó khẳng định tính ưu việt của các mô hình học sâu trong việc nhận diện các mối quan hệ phi tuyến và đặc thù theo chuỗi thời gian trong dữ liệu giáo dục.

Luận án đóng góp cơ sở lý luận mới về việc tích hợp các kiến trúc học sâu (LSTM, Transformer, Graph Neural Network...) với các kỹ thuật xử lý dữ liệu (neutrosophic, GAN, data augmentation), giúp khắc phục những thách thức do dữ liệu giáo dục thường nhỏ, phân tán và mất cân bằng. Đây là hướng nghiên cứu mới, làm cơ sở cho việc thiết kế hệ thống phân tích thông minh trong giáo dục đại học.

Về mặt thực tiễn, kết quả nghiên cứu có giá trị trực tiếp trong công tác quản lý đào tạo và hỗ trợ người học. Các mô hình dự đoán GPA theo học kỳ và phân loại tốt nghiệp sớm cho phép nhà trường sớm nhận diện sinh viên có nguy cơ bất lợi, từ đó triển khai các biện pháp can thiệp kịp thời như tư vấn học tập, điều chỉnh kế hoạch đào tạo, hay hỗ trợ tâm lý – xã hội. Đồng thời, việc dự đoán chính xác lộ trình học tập giúp nâng cao tỷ lệ duy trì và tốt nghiệp đúng hạn, qua đó tiết kiệm chi phí và nguồn lực cho cả người học và cơ sở đào tạo.

Đề tài còn có ý nghĩa trong việc xây dựng hệ thống *ra quyết định dựa trên dữ liệu* cho các nhà quản lý giáo dục. Những bằng chứng thực nghiệm thu được có thể được vận dụng vào hoạch định chính sách, đánh giá chất lượng chương trình, cũng như cải tiến phương pháp giảng dạy. Nghiên cứu góp phần phát triển hệ thống phân tích học tập thông minh, tích hợp dữ liệu – phần mềm – phần cứng – con người – quy trình, tạo tiền đề cho một môi trường giáo dục số hóa, thông minh và thích ứng trong bối cảnh chuyển đổi số quốc gia.

Đề tài không chỉ mang giá trị khoa học trong việc mở rộng tri thức về ứng dụng trí tuệ nhân tạo trong giáo dục, mà còn mang giá trị thực tiễn to lớn, phục vụ trực tiếp cho người học, giảng viên và nhà quản lý, đồng thời đáp ứng xu thế phát triển của giáo dục hiện đại.

2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình, luận án đã công bố ở trong và ngoài nước; tính trung thực, rõ rang và đầy đủ trong trích dẫn tài liệu tham khảo.

Nội dung luận án không trùng lặp với các công trình khoa học đã công bố trong và ngoài nước.

Luận án đã tuân thủ các nguyên tắc về đạo đức học thuật, đảm bảo tính trung thực trong nghiên cứu khoa học. Các nguồn tài liệu, số liệu, công trình nghiên cứu trong và ngoài nước được trích dẫn đầy đủ, chính xác và đúng quy cách tài liệu tham khảo của luận án tiến sĩ. Tác giả đã phân biệt rõ ràng giữa kết quả nghiên cứu của mình với các kết quả kế thừa, đồng thời ghi rõ xuất xứ các luận điểm, số liệu, công thức và mô hình tham khảo.

Danh mục tài liệu tham khảo trong luận án bao gồm các công trình quốc tế ISI/Scopus, tạp chí và hội thảo khoa học trong nước, phản ánh tính đa dạng, cập nhật và độ tin cậy. Không có hiện tượng sao chép hay sử dụng tài liệu mà không ghi nguồn.

3. Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành và mã ngành.

Nội dung phản ánh tên đề tài luận án, phù hợp với ngành Hệ thống thông tin và mã ngành.

4. Độ tin cậy và tính hiện đại của phương pháp đã sử dụng để nghiên cứu.

Trong bối cảnh chuyển đổi số giáo dục, việc dự đoán sớm kết quả học tập của sinh viên là nhu cầu cấp thiết nhằm nâng cao chất lượng đào tạo và hỗ trợ người học. Tuy nhiên, các nghiên cứu trước đây chủ yếu dựa vào mô hình học máy truyền thống, vốn khó nắm bắt quan hệ phi tuyến và tính chuỗi thời gian trong dữ liệu học tập. Từ hạn chế đó, luận án đặt vấn đề cần ứng dụng các mô hình học sâu hiện đại, kết hợp kỹ thuật xử lý dữ liệu tiên tiến, để cải thiện độ chính xác và khả năng khái quát trong dự đoán học tập.

Luận án triển khai cách tiếp cận tổng hợp: nghiên cứu cơ sở lý thuyết và tổng quan công trình liên quan; thiết kế và thử nghiệm nhiều mô hình học sâu (DNN, CNN, LSTM, Transformer, GNN); tích hợp kỹ thuật neutrosophic, tăng cường dữ liệu (GAN, CGAN) nhằm khắc phục tình trạng dữ liệu nhỏ và mất cân bằng; phát triển các mô hình lai mới (LATCGAd, AWG-GC) để nâng cao hiệu quả phân loại. Các mô hình được kiểm chứng qua dữ liệu thực tiễn từ HNMU, VNU và đối sánh với bộ dữ liệu quốc tế, bảo đảm tính khoa học, sáng tạo và ứng dụng.

Phương pháp nghiên cứu của luận án kết hợp nghiên cứu lý thuyết, khảo sát thực tiễn và thực nghiệm mô hình. Việc sử dụng đồng thời các mô hình học sâu, mô hình lai và kỹ thuật xử lý dữ liệu bảo đảm tính khoa học, phù hợp với đặc thù dữ liệu giáo dục. Các bộ dữ liệu được thu thập từ nhiều nguồn (HNMU, VNU, quốc tế), trải qua bước tiền xử lý và chuẩn hóa. Quá trình kiểm định bằng nhiều chỉ số (MAE, RMSE, F1-score...) và so sánh với các phương pháp truyền thống. Kết quả nghiên cứu có giá trị học thuật, có tính ứng dụng. Các công trình công bố liên quan đến luận án đều có giá trị trong các tạp chí có uy tín nên độ tin cậy cao.

5. Kết quả nghiên cứu mới của tác giả

Luận án có nội dung mới sau:

- Xây dựng các khung mô hình học sâu tích hợp yếu tố bất định cho dự đoán SGPA

- Thiết kế các mô hình lai cho bài toán phân loại tốt nghiệp trong điều kiện dữ liệu nhỏ và mất cân bằng

- Phát triển bộ dữ liệu mở rộng và quy trình phân tích phục vụ ứng dụng trong giáo dục

6. Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức luận án

6.1. Ưu điểm

+ Về hình thức: Luận án được trình bày đúng theo quy định hiện hành đối với luận án tiến sĩ, bảo đảm cấu trúc chặt chẽ gồm: Mở đầu, các chương nội dung, kết luận – kiến nghị, danh mục công trình đã công bố và tài liệu tham khảo. Văn phong khoa học, mạch lạc; bảng biểu, hình vẽ minh họa rõ ràng, có chú thích đầy đủ, thuận tiện cho việc theo dõi và đối chiếu. Hệ thống tài liệu tham khảo được trích dẫn đúng chuẩn, phản ánh tính nghiêm túc và độ tin cậy của nghiên cứu.

+ Về nội dung: Luận án tập trung vào một vấn đề có tính cấp thiết và mới trong lĩnh vực Hệ thống thông tin, đó là ứng dụng mô hình học sâu trong dự đoán kết quả học tập của người học. Nghiên cứu đã kế thừa, tổng hợp và phân tích có chọn lọc các công trình trong và ngoài nước, xác định rõ khoảng trống nghiên cứu, từ đó đề xuất các mô hình mới và giải pháp khả thi. Kết quả nghiên cứu vừa có giá trị khoa học, mở rộng tri thức trong lĩnh vực Educational Data Mining và Learning Analytics, vừa có giá trị ứng dụng trong quản lý đào tạo đại học, phù hợp với yêu cầu của một công trình tiến sĩ.

6.2. Về một số ý kiến khuyến nghị với luận án:

+ Về nguyên tắc dữ liệu cho việc học máy là tính "đồng nhất" dữ liệu, phản ánh giá trị cốt lõi đặc trưng của bộ dữ liệu, đủ lớn về số lượng. Trong luận án, sử dụng 4 loại bộ dữ liệu, mỗi bộ dữ liệu phụ thuộc vào cơ sở đào tạo (quy mô - thương hiệu - chất lượng trường, cơ sở vật chất - chương trình dạy học - thực hành thực tập - trải nghiệm - giảng viên của các ngành nghề, sự đồng đều trong hoạt động các ngành trong trường ...), sẽ ảnh hưởng đến tính đồng nhất. Ngay cả trong một bộ dữ liệu, giá trị cốt lõi đặc trưng của dữ liệu còn chịu ảnh hưởng bởi chính dữ liệu cá nhân của mỗi sinh viên (hoàn cảnh gia đình, sở thích, mức độ chăm chỉ...). Cả 2 yếu tố nêu trên đều tác động đến độ tin cậy của bộ dữ liệu sau khi tăng cường dữ liệu, ảnh hưởng đến tính chụm của đặc trưng dữ liệu. Mặt khác, số lượng dữ liệu cho việc học máy còn khiêm tốn ảnh hưởng đến chất lượng mô hình. Tác giả cho ý kiến về vấn đề này.

+ Bộ dữ liệu do luận án thu thập, chưa phải là bộ dữ liệu chuẩn đã từng được sử dụng trong các công trình đã nghiên cứu, độ tin cậy của bộ dữ liệu này thế nào? Dữ liệu tại một cơ sở đào tạo đưa vào học, ngoài dữ liệu chung như luận án đã đề cập, có đưa vào khai thác đặc trưng của đặc thù cơ sở đó không (cơ sở vật chất, chương trình đào tạo ngành...) và đưa như nào?

Đánh giá chung: Luận án có hình thức, nội dung, đóng góp mới, đáp ứng đủ tiêu chuẩn của một luận án tiến sĩ ngành Hệ thống thông tin.

7. Nội dung của luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học nào và giá trị khoa học của các công trình đã công bố

- Bài báo đăng trên tạp chí quốc tế uy tín (ISI/Scopus, Q2–Q3): 02 bài; nội dung tập trung vào ứng dụng mô hình học sâu (CNN, LSTM, Transformer) trong dự đoán kết quả học tập và xử lý dữ liệu không chắc chắn bằng neutrosophic.

- Bài báo đăng trên tạp chí khoa học trong nước có uy tín: 03 bài; đăng tải tại các tạp chí chuyên ngành Công nghệ thông tin, Hệ thống thông tin và Giáo dục, được Hội đồng chức danh Giáo sư Nhà nước tính điểm.

- Bài báo tại hội thảo khoa học quốc tế: 02 bài; trình bày tại các hội thảo VNICT (2024) và MCO (2025), phản ánh kết quả mô hình lai LATCGAd và AWG-GC trong dự đoán phân loại tốt nghiệp.

- Bài báo tại hội thảo khoa học trong nước: 01 bài; trình bày tại Hội thảo FS&IS, Trường Công nghệ Thông tin & Truyền thông – Đại học Công nghiệp Hà Nội, tập trung vào ứng dụng học sâu trong phân tích dữ liệu giáo dục.

Các công trình được công bố có số lượng đủ, chất lượng, trong đó có công bố quốc tế ISI/Scopus, đảm bảo yêu cầu bắt buộc đối với luận án tiến sĩ. Nội dung các công trình phản ánh đúng hướng nghiên cứu, có tính mới và đóng góp rõ rệt cho lĩnh vực Educational Data Mining (EDM) và Learning Analytics (LA).

8. Kết luận:

- Luận án có nội dung và hình thức đáp ứng đầy đủ các điều kiện của một luận án tiến sĩ.
- Bản tóm tắt đã phản ánh trung thực được nội dung của luận án.
- Đề nghị cho NCS đưa luận án ra bảo vệ tại Hội đồng cấp Học viện để nhận học vị Tiến sĩ.

Trường ĐH GTVT xác nhận PGS.TS. Nguyễn Văn Long là giảng viên của Nhà trường.

TL. HIỆU TRƯỞNG
KT. TRƯỞNG PHÒNG TCCB
PHÓ TRƯỞNG PHÒNG

TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI

PGS.TS. Hồ Xuân Nam

Hà nội, ngày 5 tháng 9 năm 2025
Người nhận xét

Nguyễn Văn Long

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc lập - Tự do - Hạnh phúc

BẢN NHẬN XÉT LUẬN ÁN TIẾN SĨ

Về đề tài : Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự báo kết quả học tập của người học

Ngành: Hệ thống thông tin Mã số: 948.01.04

Nghiên cứu sinh: Nguyễn Thị Kim Sơn

Người nhận xét luận án: Đỗ Trung Tuấn,

Cơ quan công tác: Trường đại học Khoa học tự nhiên, Đại học Quốc gia Hà Nội

NỘI DUNG NHẬN XÉT

1. Tính cần thiết, thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận án

Luận án của nghiên cứu sinh Nguyễn Thị Kim Sơn đề cập ứng dụng của học sâu, một nhánh trong học máy, thuộc trí tuệ nhân tạo. Học sâu là một phầntrong một nhánh rộng hơn các phương pháp học máy dựa trên mạng thần kinhnhân tạo kết hợp với việc học biểu diễn đặc trưng. Việc học này có thể có giám sát, nửa giám sát hoặc không giám sát. một biểu diễn có tổ chức của các thực thể trong thế giới thực và các mối quan hệ của chúng.

Nghiên cứu sinh công tác trong lĩnh vực giáo dục nên đề tài luận án hướngtới ứng dụng trong ngành, đề cập (i) khai phá thông tin về học; (ii) cá nhân hóa trải nghiệm khi học. Trong bản viết luận án, nghiên cứu sinh đã liên kết hai mục đích cần đạt về lí thuyết và ứng dụng, để dẫn ra các kĩ thuật như (i) tăng cường dữ liệu; (ii) chọn đặc trưng; (iii) tối ưu các tham số mô hình. Vậy nên đề tài và luận án của nghiên cứu sinh có ý nghĩa.

Tuy tiếp cận theo cách truyền thống trong ngành giáo dục và đào tạo, kết quả luận án có thể được phát triển cho các mô hình tương tự.

2. Sự phù hợp giữa tên đề tài với nội dung, giữa nội dung với ngành đào tạo và mã số

Đề tài luận án và kết quả trình bày trong luận án của nghiên cứu sinh phù hợp với yêu cầu của chuyên ngành Hệ thống thông tin, mã số : 9.48.01.04.

Đề tài không trùng lặp với các đề tài nghiên cứu hay luận án khác đã bảo vệ tại cơ sở đào tạo. Luận án của nghiên cứu sinh Nguyễn Thị Kim Sơn được tập thể PGS. TS. Nguyễn Hữu Quỳnh và PGS. TS. Ngô Quốc Tạo hướng dẫn. Các thầy là chuyên trong lĩnh vực nghiên cứu này.

3. Độ tin cậy của phương pháp nghiên cứu

Căn cứ vào động cơ nghiên cứu đã trình bày trong chương đầu của luận án, dựa trên công nghệ học sâu, áp dụng một số kĩ thuật học máy và xử lí dữ liệu, nghiên cứu sinh đã (i) dự báo; (ii) phân loại dữ liệu.

Những dữ liệu nghiên cứu sinh sử dụng thuộc cơ sở đào tạo, có tính thực tế.

Loại dữ liệu đa dạng, cho phép luận án thực hiện các thử nghiệm để lựa chọn được tiếp cận tốt hơn.

So sánh với luận án liên quan đến giáo dục học, tâm lí học, luận án này hướng công nghệ. Do vậy các kiểm định giả thuyết thống kê như các luận án về giáo dục được thay thế bằng các phân tích thống kê và đánh giá mức độ chính xác khi huấn luyện các mô hình đề xuất.

4. Kết quả nghiên cứu mới của nghiên cứu sinh

Nghiên cứu sinh sử dụng thuật ngữ về logic thần kinh, phát triển logic cổ điển và logic mờ để tích hợp vào mô hình học máy.

Người nhận xét nhất trí với nghiên cứu sinh về hai đóng góp chính về (i) mô hình học sử dụng logic thần kinh; (ii) mô hình dự báo cải tiến. Các kết quả này được trình bày trong (i) phần 2.3, trang 53 của bản viết luận án, với các thuật toán; (ii) phần 3.2 trang 74, với các thuật toán. Dữ liệu đóng góp cho cộng đồng xử lí dữ liệu giáo dục được nghiên cứu sinh xem như công sức mới là bộ dữ liệu, kèm theo đề xuất khung phân tích dữ liệu.

5. Ưu điểm và nhược điểm về nội dung, kết cấu và hình thức của luận án

Nghiên cứu sinh trình bày vấn đề sáng sủa, người nhận xét dễ theo dõi. Luận án được trình bày cẩn thận, bằng tiếng Anh. Một số phục lục nên đặt trong phần phụ lục của luận án, thay vì đặt cuối chương.

Về cấu trúc, luận án trình bày theo các chương, ứng với các kết quả chính của luận án. Tỉ lệ số trang giữa các chương là hợp lí.

6. Nội dung luận án đã được công bố trên tạp chí, kỷ yếu hội nghị khoa học

Nghiên cứu sinh đã công bố cùng tập thể hướng dẫn và đồng nghiệp các kết quả nghiên cứu liên quan đến đề tài luận án trên các địa chỉ có uy tín và có phản

biện. Các địa chỉ này bảo đảm tính mới của của các kết quả luận án. Các phản biện sẽ đánh giá chất lượng chi tiết của các công bố và các thuật toán đã được trình bày trong bản viết luận án. Nghiên cứu sinh cùng tập thể hướng dẫn đã công bố 8 công trình liên quan đến kết quả luận án. Trong đó có 4 công trình năm 2025 đã được chấp nhận đăng.

- Liên quan đến dự báo với kĩ thuật học máy , nghiên cứu sinh có công bố năm 2022 trên Tạp chí quốc tế, ứng với CT1, năm 2024 trên Tạp chí Tính toán và điều khiển, ứng với CT2.
- Liên quan đến các mô hình học sâu và kĩ thuật học máy , có các công bố năm 2025, ứng với CT5, CT6, CT7, CT8.
- Liên quan đến bộ dữ liệu thử nghiệm cho quá trình học máy , nghiên cứu sinh có công bố năm 2024, 2025 ứng với CT3, CT4.

So sánh với một số luận án đã bảo vệ cơ sở đào tạo, số lượng chất lượng công bố là đáng khích lệ.

7. Kết luận

Căn cứ vào yêu cầu của luận án tiến sĩ chuyên ngành, dựa vào kết quả về lí thuyết và thực nghiệm mà nghiên cứu sinh đã trình bày trong các chương luận án, người nhận xét nhất trí đề nghị bản luận án được bảo vệ trước Hội đồng chấm luận án cấp Học viện Khoa học và Công nghệ.

Hà Nội, ngày 4 tháng 9 năm 2025

Người nhận xét

Đỗ Trung Tuấn

CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Hà Nội, 6 tháng 10 năm 2025

BẢN NHẬN XÉT PHẢN BIỆN LUẬN ÁN TIẾN SĨ

Tên đề tài luận án: **Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học**

Chuyên ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Người hướng dẫn: PGS.TS. Nguyễn Hữu Quỳnh, PGS.TS Ngô Quốc Tạo

Người nhận xét: PGS.TS Phạm Văn Hải, Ủy viên

Đơn vị công tác: Trường Công nghệ Thông tin- Truyền thông, Đại học Bách Khoa Hà Nội, B1, Tạ Quang Bửu, Hà Nội

1- Nội dung nhận xét

- Đề tài có ý nghĩa khoa học thực tiễn ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học.
- Giá trị khoa học, số liệu, kết quả nghiên cứu và kết luận của luận án đáng tin cậy.
- Đề tài luận án, các kết quả nghiên cứu, các nhận xét, kết luận không trùng lặp với các tài liệu, công trình công bố ở trong nước và quốc tế.
- Các phương pháp phân tích dữ liệu, mô hình học sâu đề xuất phù hợp với hệ thống đề xuất.
- Trích dẫn tài liệu rõ ràng, đầy đủ và trung thực.

2- Văn phong, kết cấu và cách trình bày của luận án

- Luận án được trình bày văn phong và cách trình bày bảng biểu, hình vẽ rõ ràng với cách trích dẫn tài liệu chính xác, minh bạch.
- Cách trình bày luận án phù hợp, rõ ràng.

3- Kết quả và đóng góp mới nghiên cứu của tác giả

Các kết quả này đã trình bày trong mục 5 trang 5 của luận án phù hợp.

4- Các công trình khoa học

- Tác giả công bố 01 bài báo tạp chí quốc tế SCIE, ISI, 02 bài báo tạp chí công bố chỉ số scopus, 02 bài báo uy tín công bố tạp chí trong nước, hội thảo quốc tế Spinger 01 bài. Các bài báo đủ các công bố khoa học cho điều kiện bảo vệ luận án tiến sĩ. Bài báo chưa được chấp thuận, không nhất thiết để trong danh mục công bố khoa học.
- Các bài báo thể hiện nội dung chính trong luận án của tác giả và thể hiện cách thức làm việc chủ động của tác giả trong quá trình nghiên cứu.

5- Hình thức trình bày và những thiếu xót trong bản luận án

- Nội dung luận án trung thực, bố cục và các nội dung cơ bản của luận án một cách hợp lý.

6- Các góp ý cần cập nhật, sửa chữa luận án

- Luận án trình bày cấu trúc logic, các nội dung trình bày hợp lý. Luận án viết bằng ngôn ngữ tiếng Anh, có hàm lượng khoa học và ứng dụng thực tiễn. Tuy nhiên các sửa đổi góp ý như sau:
- Mục Motivation of the dissertation cần tóm lược thông tin, nội dung viết chưa cô đọng. Cần viết điểm chính tóm tắt thay vì mô tả dài dòng.
- Rà soát một số chỗ, cách dùng từ và hiệu đính tiếng Anh trong luận án.

7- Kết luận

Luận án đã đạt đầy đủ yêu cầu của một luận án tiến sĩ. Kiến nghị luận án đưa ra bảo vệ cấp Học viện để nhận học vị tiến sĩ.

Người nhận xét

PGS.TS Phạm Văn Hải

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 08 tháng 10 năm 2025

BIÊN BẢN CỦA
HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Căn cứ quyết định số 821/QĐ-HVKHCN ngày 25 tháng 08 năm 2025 của Giám đốc Học viện Khoa học và Công nghệ về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện, Hội đồng đã họp vào hồi 09 giờ ngày 08 tháng 10 năm 2025 tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, số 18 đường Hoàng Quốc Việt, Cầu Giấy, Hà Nội để đánh giá luận án tiến sĩ.

Họ và tên NCS: **Nguyễn Thị Kim Sơn**

Tên đề tài luận án: Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học / Reseach on the application of deep learning models for predicting learners'academic performance

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Người hướng dẫn: PGS.TS. Nguyễn Hữu Quỳnh, PGS.TS. Ngô Quốc Tạo

THAM DỰ BUỔI BẢO VỆ GỒM CÓ

- Đại diện cơ sở đào tạo:
  1. GS.TS. Vũ Đình Lãm - Giám đốc Học viện KH&CN
- Đại diện Viện Công nghệ thông tin:
  1. PGS.TS. Nguyễn Trường Thắng - Viện trưởng Viện Công nghệ thông tin
- Đại diện Cơ quan chủ quản của NCS:

  TS. Nguyễn Văn Thiện, Đại học Công nghiệp Hà Nội
- Thành viên Hội đồng có mặt: 7/7 thành viên
  1. PGS.TS. Nguyễn Long Giang, Chủ tịch Hội đồng
  2. PGS.TS. Bùi Thu Lâm, Phản biện 1
  3. TS. Nguyễn Như Sơn, Phản biện 2

4. TS. Trần Đức Nghĩa, Thư ký Hội đồng
5. PGS.TS. Nguyễn Văn Long, Ủy viên
6. PGS.TS. Đỗ Trung Tuấn, Ủy viên
7. PGS.TS. Phạm Văn Hải, Ủy viên

- Thành viên Hội đồng vắng mặt: PGS.TS. Nguyễn Văn Long
- Đại diện tập thể cán bộ hướng dẫn: PGS.TS. Nguyễn Hữu Quỳnh, PGS.TS. Ngô Quốc Tạo
- Cùng tham dự buổi bảo vệ còn có nhiều cán bộ nghiên cứu khoa học trong và ngoài Học viện.

TIẾN TRÌNH BUỔI BẢO VỆ

1. *Đại diện cơ sở đào tạo,* cô Phạm Thị Như Quỳnh, tuyên bố lý do, giới thiệu đại biểu và đọc quyết định số 821/QĐ-HVKHCN ngày 25 tháng 08 năm 2025 của Giám đốc HVKHCN về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện cho NCS Nguyễn Thị Kim Sơn và đề nghị Chủ tịch Hội đồng điều khiển phiên họp.
2. *Chủ tịch Hội đồng*, PGS.TS. Nguyễn Long Giang, công bố danh sách thành viên có mặt là 06, thông qua chương trình buổi bảo vệ, đề nghị Thư ký thông báo các điều kiện chuẩn bị cho buổi bảo vệ và đọc lý lịch khoa học của NCS.
3. *Thư ký Hội đồng*, TS. Trần Đức Nghĩa thông báo các điều kiện cho buổi bảo vệ
   - Đọc lý lịch khoa học của NCS Nguyễn Thị Kim Sơn.
   - Đã nhận đủ 07 nhận xét của các phản biện và các thành viên HĐ.
   - Lịch bảo vệ của NCS đã được đăng trên Cổng thông tin điện tử Học viện Khoa học và Công nghệ ngày 15/09/2025.
   - Các giấy tờ cần thiết khác.

   NCS Nguyễn Thị Kim Sơn có đủ các điều kiện về thủ tục để bảo vệ luận án trước Hội đồng đánh giá luận án cấp Học viện.
4. Các thành viên hội đồng và những người tham dự thông qua về lý lịch khoa học và quá trình đào tạo của nghiên cứu sinh.
5. Nghiên cứu sinh Nguyễn Thị Kim Sơn trình bày nội dung luận án trong 30 phút trước Hội đồng. Báo cáo của NCS bao gồm các nội dung chính như sau:

   Giới thiệu

Tổng quan về dự đoán kết quả học tập từ các phương pháp học máy và học sâu.
Các nghiên cứu liên quan.
Động lực và thách thức.
Các kết quả chính của luận án:

(1) Đề xuất 02 mô hình mới là NeutroDL và NeutroGNT, tích hợp quy trình neutrosophic vào các mô hình học sâu để nâng cao hiệu suất dự đoán GPA sớm.

(2) Đề xuất 02 mô hình lai mới là LATCGAd và AWG-GC, để dự đoán phân loại tốt nghiệp cho sinh viên.

(3) Phát triển 03 tập dữ liệu đa thuộc tính từ nhiều nguồn khác nhau và đề xuất các khuôn khổ phân tích phù hợp với dữ liệu giáo dục.

Kết luận và phát triển trong tương lai

6. Các phản biện đọc bản nhận xét và đặt câu hỏi đánh giá luận án của NCS Nguyễn Thị Kim Sơn

1) **Phản biện 1, PGS.TS. Bùi Thu Lâm**, đọc nhận xét đánh giá luận án và kết luận (có văn bản kèm theo).

Ưu điểm:

Nội dung được trình bày chi tiết, tập trung vào vấn đề thực tiễn với dữ liệu Việt Nam, đề xuất mô hình mới hiệu quả. Kết cấu logic, giới thiệu rõ vấn đề, tổng quan toàn diện, phương pháp hiện đại, thí nghiệm chi tiết. Bảng biểu, đồ thị rõ ràng, ngôn ngữ khoa học, 142 trang cân đối.

Nhược điểm:

Về nội dung:

(1) Dữ liệu nhỏ (HNMU1: 932, HNMU2: 551, VNU: 271), dẫn đến R^2 thấp ở một số case, thiếu tổng quát hóa cho trường hợp lớn hơn.

(2) So sánh baseline chưa đầy đủ, thiếu mô hình SOTA cho dữ liệu giáo dục.

(3) Chưa phân tích sâu yếu tố đặc thù Việt Nam, dẫn đến mô hình chưa tối ưu hóa địa phương; Đặc biệt, mô hình dự báo ở Chương 2 dừng lại ở mức độ đơn biến, chưa tính tới yếu tố đa biến.

Về kết cấu:

Chương 2 và 3 lặp lại một số phần (như GAN), thiếu liên kết chặt chẽ giữa dự đoán SGPA và loại tốt nghiệp.

Về hình thức:

Không nên phân biệt nghiên cứu trong và ngoài nước (trang 16)

Nên có các giải thích về sự hợp lý trong các lai ghép mô hình

Các ký hiệu toán học ở chương 2 nên chuẩn hóa cho thống nhất và dùng văn phong academic hơn.

2) **Phản biện 2, TS. Nguyễn Như Sơn**, đọc nhận xét đánh giá luận án và kết luận (có văn bản kèm theo).

Ưu điểm:

Luận án bố cục làm 3 chương gồm 120 trang nội dung và các phần liên quan về công trình công bố và tài liệu tham khảo, phù hợp với cấu trúc của luận án tiến sĩ, các chương được phân chia hợp lý các nội dung liên quan, trình bày chi tiết rõ ràng các mô hình, kỹ thuật đề xuất trong phạm vi nghiên cứu của luận án. Hệ thống 80 tài liệu tham khảo được trích dẫn đúng chuẩn, các tài liệu có tính cập nhật mới trong những năm gần đây.

Nhược điểm:

Tuy nhiên cần xem xét một số ý kiến sau đây để chính sửa luận án cho hợp lý hơn nếu có thể được:

- Về quy mô dữ liệu, do việc thu thập thực tế sẽ mất nhiều thời gian nên hiện tại có quy mô còn chưa lớn, sẽ ảnh hưởng chất lượng mô hình?.

- Các thuật toán đề xuất hiện chưa có đánh giá sơ bộ về độ phức tạp, thời gian tính toán, có thể xem như phụ thuộc vào độ phức tạp của các mô hình cơ bản được sử dụng?.

- Các kết quả hiện không đánh giá thời gian chạy, tuy nhiên thực nghiệm nên có các thông số về cấu hình phần cứng để dùng cho việc tham khảo khác sau này?

- Trình bày thuật toán nên theo format truyền thống: Input, Output (không đánh số dòng lệnh), Begin – End mới bắt đầu đánh số.

- Chương ba cần nói thêm nội dung công bố ở CT2,

- Rà soát về các chú thích Hình vẽ, Bảng biểu cố gắng không nên ngắt trang ở cả 2 bản luận án và tóm tắt.

7. NCS Nguyễn Thị Kim Sơn tiếp thu ý kiến nhận xét của các phản biện và trả lời đầy đủ câu hỏi của các uỷ viên phản biện.
8. Các thành viên khác trong Hội đồng đưa ra ý kiến nhận xét; Hội đồng và những người tham dự đặt câu hỏi

- **PGS.TS. Nguyễn Long Giang**

• Luận án có cấu trúc hợp lý, nội dung nghiên cứu và các kết quả thu được đã được NCS trình bày khá nghiêm túc, rõ ràng và lô gíc.

• Mục CONTENTS (trang viii) của luận án cần được đưa lên trước SYMBOLS AND ABBREVIATIONS theo qui định của cơ sở đào tạo. Bảng SYMBOLS AND ABBREVIATIONS (trang iii) cần được sắp xếp theo vần abc để dễ tra cứu.

• Các thuật toán cần được đánh số theo chương như hình và bảng, ví dụ Algorithm 1. (trang 48) thành Algorithm 2.1,...

• Phân tích thêm Thông tư 42/2021 "Quy định về cơ sở dữ liệu giáo dục và đào tạo" của BGD-ĐT trước khi kết luận "However, in education, there is currently a lack of large, standardized, ..." (trang 2).

• Nên bổ sung vào Research Subjects (trang 3): Ngoài các bài toán dự báo kết quả học tập của sinh viên, đối tượng nghiên cứu còn là các mô hình học sâu.

• Mục 5. Key contributions of the dissertation (trang 5): Mặc dù đã có mô tả "From an information systems perspective,...", tuy nhiên nên bổ sung sơ đồ khối trực quan Hệ thống thông tin dự báo kết quả học tập của sinh viên, từ đó chỉ ra các kết quả nghiên cứu của luận án phục vụ cho các khối chức năng nào.

• Các mục 2.4. Appendix to Chapter 2 (trang 64), 3.4. Appendix to Chapter 3 (trang 114) cần chuyển về Chương 1 hoặc để cuối luận án.

• Phân tích sâu hơn độ phức tạp của các thuật toán đề xuất trong luận án. NCS mới quan tâm đến độ chính xác, chưa quan tâm đến thời gian và bộ nhớ cần thiết khi thực hiện các thuật toán.

• Nên có thêm thực nghiệm các thuật toán Chương 2 với các bộ dữ liệu khác nhau để đánh giá khách quan hơn.

- **PGS.TS. Nguyễn Văn Long**

+ Về nguyên tắc dữ liệu cho việc học máy là tính "đồng nhất" dữ liệu, phản ánh giá trị cốt lõi đặc trưng của bộ dữ liệu, đủ lớn về số lượng. Trong luận án,

sử dụng 4 loại bộ dữ liệu, mỗi bộ dữ liệu phụ thuộc vào cơ sở đào tạo (quy mô - thương hiệu - chất lượng trường, cơ sở vật chất - chương trình dạy học - thực hành thực tập - trải nghiệm - giảng viên của các ngành nghề, sự đồng đều trong hoạt động các ngành trong trường …), sẽ ảnh hưởng đến tính đồng nhất. Ngay cả trong một bộ dữ liệu, giá trị cốt lõi đặc trưng của dữ liệu còn chịu ảnh hưởng bởi chính dữ liệu cá nhân của mỗi sinh viên (hoàn cảnh gia đình, sở thích, mức độ chăm chỉ…). Cả 2 yếu tố nêu trên đều tác động đến độ tin cậy của bộ dữ liệu sau khi tăng cường dữ liệu, ảnh hưởng đến tính chụm của đặc trưng dữ liệu. Mặt khác, số lượng dữ liệu cho việc học máy còn khiêm tốn ảnh hưởng đến chất lượng mô hình. Tác giả cho ý kiến về vấn đề này.

+ Bộ dữ liệu do luận án thu thập, chưa phải là bộ dữ liệu chuẩn đã từng được sử dụng trong các công trình đã nghiên cứu, độ tin cậy của bộ dữ liệu này thế nào? Dữ liệu tại một cơ sở đào tạo đưa vào học, ngoài dữ liệu chung như luận án đã đề cập, có đưa vào khai thác đặc trưng của đặc thù cơ sở đó không (cơ sở vật chất, chương trình đào tạo ngành…) và đưa như nào?

- **PGS.TS. Đỗ Trung Tuấn**

Nghiên cứu sinh trình bày vấn đề sáng sủa, người nhận xét dễ theo dõi.

Luận án được trình bày cẩn thận, bằng tiếng Anh.

Một số phụ lục nên đặt trong phần phụ lục của luận án, thay vì đặt cuối chương.

- **PGS.TS. Phạm Văn Hải**

Các bài báo đủ các công bố khoa học cho điều kiện bảo vệ luận án tiến sĩ. Bài báo chưa được chấp nhận đăng, không nhất thiết để trong danh mục công bố khoa học.

Luận án trình bày cấu trúc logic, các nội dung trình bày hợp lý. Luận án viết bằng ngôn ngữ tiếng Anh, có hàm lượng khoa học và ứng dụng thực tiễn. Tuy nhiên các sửa đổi góp ý như sau:

- Mục Motivation of the dissertation cần tóm lược thông tin, nội dung viết chưa cô đọng. Cần viết điểm chính tóm tắt thay vì mô tả dài dòng.
- Rà soát một số chỗ, cách dùng từ và hiệu đính tiếng Anh trong luận án.

- **TS. Trần Đức Nghĩa**

Trong các nghiên cứu tương lai, nên bổ sung thêm các yếu tố ảnh hưởng đến kết quả học tập khác, ví dụ như thái độ/sự tích cực trong học tập, như vậy việc dự đoán sẽ tốt hơn.

Nên thử nghiệm thêm trên các bộ dữ liệu khác nhau để đảm bảo tính khách quan và sự hiệu quả của đề xuất.

Lưu ý rà soát LA đảm bảo theo mẫu/cấu trúc của Học viện.

9. Thư ký đọc nhận xét của 02 chuyên gia phản biện độc lập.
10. NCS Nguyễn Thị Kim Sơn tiếp thu ý kiến nhận xét của các thành viên của Hội đồng. NCS trả lời các câu hỏi của các thành viên Hội đồng và trình bày giải trình nhận xét của 02 chuyên gia phản biện độc lập.
11. Đại diện tập thể hướng dẫn phát biểu ý kiến bằng văn bản.
12. Hội đồng tiến hành họp riêng để bầu ban kiểm phiếu, bỏ phiếu kín và thảo luận thông qua quyết nghị của Hội đồng.

1) Bầu ban kiểm phiếu gồm:

- Trưởng ban: PGS.TS. Đỗ Trung Tuấn
- Uỷ viên: PGS.TS. Phạm Văn Hải
- Uỷ viên: TS. Trần Đức Nghĩa

2) Bỏ phiếu kín và thảo luận thông qua Quyết nghị của Hội đồng.

- Trưởng ban kiểm phiếu, PGS.TS. Đỗ Trung Tuấn công bố kết quả kiểm phiếu (có biên bản kiểm phiếu).
- Chủ tịch Hội đồng, PGS.TS. Nguyễn Long Giang, thông qua Quyết nghị (có văn bản kèm theo).

3) Tóm tắt nghị quyết của Hội đồng

3.1. Tính phù hợp của tên đề tài và sự không trùng lặp về nội dung luận án

- Tên đề tài, nội dung và kết quả nghiên cứu của luận án phù hợp với Ngành đào tạo "Hệ thống thông tin", mã số "9 48 01 04".
- Nội dung của luận án không trùng lặp với các luận án đã bảo vệ và các kết quả nghiên cứu đã công bố trong và ngoài nước.
- Các tài liệu tham khảo của luận án có nội dung phù hợp và đã được trích dẫn trong luận án.

3.2. Các kết quả chính của luận án

Luận án có một số kết quả nghiên cứu mới trong bài toán dự đoán sớm kết quả học tập của sinh viên:

(1) Đề xuất 02 mô hình mới là NeutroDL và NeutroGNT, tích hợp quy trình neutrosophic vào các mô hình học sâu để nâng cao hiệu suất dự đoán

GPA sớm.

(2) Đề xuất 02 mô hình lai mới là LATCGAd và AWG-GC, để dự đoán phân loại tốt nghiệp cho sinh viên.

(3) Phát triển 03 tập dữ liệu đa thuộc tính từ nhiều nguồn khác nhau và đề xuất các khuôn khổ phân tích phù hợp với dữ liệu giáo dục.

3.3. Các điểm cần bổ sung chỉnh sửa trước khi nộp luận án cho Thư viện Quốc gia Việt Nam

NCS cần tiếp thu, rà soát, chỉnh sửa, bổ sung nội dung luận án theo ý kiến đóng góp trong bản nhận xét của các thành viên Hội đồng và Biên bản của Hội đồng đánh giá luận án tiến sĩ cấp Học viện trước khi nộp luận án cho Thư viện Quốc gia Việt Nam.

3.4. Mức độ đáp ứng yêu cầu của luận án tiến sĩ về cả nội dung và hình thức

- Luận án của NCS Nguyễn Thị Kim Sơn đáp ứng yêu cầu của luận án tiến sĩ ngành "Hệ thống thông tin", mã số "9 48 01 04" về nội dung và hình thức theo các qui chế hiện hành về đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo, của Học viện Khoa học và Công nghệ.
- Đề nghị Học viện Khoa học và Công nghệ công nhận kết quả bảo vệ và cấp bằng tiến sĩ cho NCS Nguyễn Thị Kim Sơn sau khi chỉnh sửa, bổ sung luận án theo các góp ý của Hội đồng.

13. Tổng kết

- Trưởng ban kiểm phiếu, PGS.TS. Đỗ Trung Tuấn, công bố kết quả bỏ phiếu đánh giá luận án.
- Chủ tịch Hội đồng, PGS.TS. Nguyễn Long Giang, đọc Quyết nghị của Hội đồng.
- Chủ tịch Hội đồng tuyên bố Hội đồng đã hoàn thành nhiệm vụ và trao lại quyền điều khiển cho Cơ sở đào tạo.
- Các đại biểu và NCS phát biểu ý kiến.
- Đại diện cơ sở đào tạo tuyên bố kết thúc buổi bảo vệ luận án tiến sĩ.

Biên bản họp Hội đồng này được / ủy viên Hội đồng biểu quyết công khai thông qua.

Buổi họp Hội đồng đánh giá luận án tiến sĩ cấp Học viện kết thúc vào lúc 11 giờ 30 phút, ngày 08 tháng 10 năm 2025.

Thư ký Hội đồng	Chủ tịch Hội đồng
TS. Trần Đức Nghĩa	PGS.TS. Nguyễn Long Giang

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Nguyễn Thị Trung

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

Hà Nội, ngày 08 tháng 10 năm 2025

QUYẾT NGHỊ CỦA
HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Căn cứ quyết định số 821/QĐ-HVKHCN ngày 25 tháng 08 năm 2025 của Giám đốc Học viện Khoa học và Công nghệ về việc thành lập Hội đồng đánh giá luận án tiến sĩ cấp Học viện, Hội đồng đã họp vào hồi 09 giờ 00 ngày 08 tháng 10 năm 2025 tại Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, số 18 đường Hoàng Quốc Việt, Cầu Giấy, Hà Nội để đánh giá luận án tiến sĩ.

Họ và tên NCS: **Nguyễn Thị Kim Sơn**

Tên đề tài luận án: Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học / Reseach on the application of deep learning models for predicting learners'academic performance

Ngành: Hệ thống thông tin

Mã số: 9 48 01 04

Người hướng dẫn: PGS.TS. Nguyễn Hữu Quỳnh và PGS.TS. Ngô Quốc Tạo

HỘI ĐỒNG ĐÁNH GIÁ LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN CỦA NCS NGUYỄN THỊ KIM SƠN KẾT LUẬN

1. Tính phù hợp của tên đề tài và sự không trùng lặp về nội dung luận án

- Tên đề tài, nội dung và kết quả nghiên cứu của luận án phù hợp với ngành đào tạo Hệ thống thông tin, mã số 9 48 01 04.

- Nội dung của luận án không trùng lặp với các luận án đã bảo vệ và các kết quả nghiên cứu đã công bố trong và ngoài nước.

- Các tài liệu tham khảo của luận án có nội dung phù hợp và đã được trích dẫn trong luận án.

2. Kết quả, ý nghĩa khoa học, thực tiễn của đề tài

Luận án có một số kết quả nghiên cứu mới trong bài toán dự đoán sớm kết quả học tập của sinh viên:

(1) Đề xuất 02 mô hình lai NeutroDL và NeutroGNT nhằm tích hợp quy trình neutrosophic vào các mô hình học sâu để nâng cao hiệu suất dự đoán GPA sớm.

(2) Đề xuất 02 mô hình lai LATCGAd và AWG-GC để dự đoán phân loại tốt nghiệp cho sinh viên.

Ngoài ra, luận án xây dựng 03 tập dữ liệu đa thuộc tính từ nhiều nguồn khác nhau và đề xuất các khuôn khổ phân tích phù hợp với dữ liệu giáo dục.

3. Những thiếu sót của luận án, vấn đề cần bổ sung, sửa chữa

NCS cần tiếp thu, rà soát, chỉnh sửa, bổ sung nội dung luận án theo ý kiến đóng góp trong bản nhận xét của các thành viên Hội đồng và Biên bản của Hội đồng đánh giá luận án tiến sĩ cấp Học viện trước khi nộp luận án cho Thư viện Quốc gia Việt Nam.

4. Mức độ đáp ứng yêu cầu của luận án tiến sĩ về cả nội dung và hình thức theo các quy chế hiện hành về đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo

Luận án của NCS Nguyễn Thị Kim Sơn đáp ứng yêu cầu của một luận án tiến sĩ ngành "Hệ thống thông tin", mã số 9 48 01 04 về nội dung và hình thức theo các quy chế hiện hành về đào tạo tiến sĩ của Bộ Giáo dục và Đào tạo và của Học viện Khoa học và Công nghệ.

Kết luận:

Kết quả bỏ phiếu đánh giá luận án của Hội đồng: 6/6 thành viên tán thành.

Hội đồng kết luận thông qua luận án, đề nghị Học viện Khoa học và Công nghệ công nhận kết quả bảo vệ và cấp bằng tiến sĩ cho NCS Nguyễn Thị Kim Sơn.

Quyết nghị này được 6/6 thành viên Hội đồng biểu quyết công khai thông qua.

THƯ KÝ	**CHỦ TỊCH**
TS. Trần Đức Nghĩa	**PGS.TS. Nguyễn Long Giang**

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Nguyễn Thị Trung

(Mẫu 6-HV-Bản giải trình chỉnh sửa, bổ sung cấp HV)

VIỆN HÀN LÂM
KHOA HỌC VÀ CÔNG NGHỆ VN
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc

BẢN GIẢI TRÌNH CHỈNH SỬA, BỔ SUNG LUẬN ÁN TIẾN SĨ CẤP HỌC VIỆN

Ngày 8 tháng 10 năm 2025, Học viện Khoa học và Công nghệ đã tổ chức đánh giá luận án tiến sĩ cấp Học viện cho nghiên cứu sinh Nguyễn Thị Kim Sơn theo Quyết định số 821/QĐ-HVKHCN ngày 25 tháng 8 năm 2025 của Giám đốc Học viện.

Đề tài: Nghiên cứu ứng dụng một số mô hình sử dụng học sâu trong dự đoán kết quả học tập của người học (Research on the application of deep learning models for predicting learners' academic performance)

Ngành: Hệ thống thông tin; Mã số : 9 48 01 04

Tập thể hướng dẫn khoa học: PGS.TS Nguyễn Hữu Quỳnh; PGS.TS Ngô Quốc Tạo

Theo Biên bản của Hội đồng, NCS phải bổ sung và chỉnh sửa luận án các điểm sau đây:

STT	Nội dung đề nghị chỉnh sửa, bổ sung	Nội dung đã được chỉnh sửa, bổ sung *(Ghi rõ số trang/chương/mục... đã được chỉnh sửa)*
1	Về quy mô dữ liệu, do việc thu thập thực tế sẽ mất nhiều thời gian nên hiện tại có quy mô còn chưa lớn. Dữ liệu nhỏ (HNMU1: 932, HNMU2: 551, VNU: 271). Ảnh hưởng chất lượng mô hình, thiếu tổng quát hóa cho trường hợp lớn hơn.	NCS tiếp thu. Đúng là trong bối cảnh giáo dục đại học ở Việt Nam, việc thu thập dữ liệu quy mô lớn, có nhãn, đồng bộ giữa nhiều khóa học và nhiều hệ thống là một thách thức thực tế. Đây là thực tế chung trong nhiều công trình cả trong nước và quốc tế do đặc thù dữ liệu giáo dục thường liên quan đến quyền riêng tư, phân tán ở nhiều hệ thống và khó chuẩn hóa. Hạn chế về quy mô dữ liệu là đặc thù khách quan của lĩnh vực giáo dục, nhưng luận án đã **biến hạn chế đó thành bài toán nghiên cứu trung tâm**. Các mô hình đề xuất (NeutroGNT, LATCGAd, AWG-GC) không những giải quyết được vấn đề dữ liệu nhỏ mà còn cho thấy hiệu quả vượt trội so với baseline, chứng minh giá trị khoa học và thực tiễn của hướng tiếp cận. Trong tương lai, hướng mở rộng là thu thập thêm dữ liệu liên cơ sở hoặc ứng dụng federated learning để vừa tăng kích thước mẫu vừa bảo đảm quyền riêng tư.
2	Trong luận án, dữ liệu từ nhiều cơ	Dữ liệu giáo dục có sự khác biệt giữa các cơ sở đào tạo và giữa từng cá nhân sinh viên. Các yếu tố "không đồng nhất" và "ít dữ liệu" đã

	sở đào tạo khác nhau và cả yếu tố cá nhân sinh viên có thể làm giảm tính đồng nhất, trong khi số lượng dữ liệu còn hạn chế. Nêu rõ hơn về độ tin cậy và khả năng tổng quát hóa.	được xem xét kỹ trong thiết kế mô hình để đảm bảo độ tin cậy và khả năng tổng quát hóa. Cụ thể: • Chuẩn hoá dữ liệu (thang điểm, cấu trúc học tập) nhằm giảm khác biệt giữa các đặc trưng và giữa các trường. • Tăng cường dữ liệu có kiểm soát (GAN, CGAN, WGAN) và luôn đánh giá riêng biệt để tránh nhiễu.
3	Dữ liệu tại một cơ sở đào tạo đưa vào học, ngoài dữ liệu chung như luận án đã đề cập, có đưa vào khai thác đặc trưng của đặc thù cơ sở đó không (cơ sở vật chất, chương trình đào tạo ngành…) và đưa như nào?	Ngoài các thông tin chung về cá nhân và kết quả học tập, bộ dữ liệu có tích hợp thêm các đặc trưng đặc thù của cơ sở đào tạo như: Cơ sở vật chất: điều kiện cơ sở vật chất, hệ thống hỗ trợ học tập. Chương trình đào tạo và phương pháp giảng dạy: mức độ phù hợp của chương trình đào tạo, chất lượng giảng viên, điều kiện học tập. Môi trường học tập: mức độ cạnh tranh trong học tập, sự hỗ trợ từ giảng viên và nhà trường, sự thích nghi của sinh viên với môi trường, năng lực cạnh tranh trong học tập. Các yếu tố này được mã hóa thành biến số (feature) trong dữ liệu, ví dụ: "Facility conditions", "Quality of instructors", "Suitability of the training program". Nhờ đó, mô hình không chỉ phân tích đặc điểm cá nhân của sinh viên mà còn phản ánh đúng ảnh hưởng từ điều kiện, chính sách và chất lượng đào tạo của từng cơ sở.
4	So sánh baseline chưa đầy đủ, thiếu mô hình SOTA cho dữ liệu giáo dục.	Tiếp thu. Baseline hiện mới dừng ở mô hình ML/DL phổ biến. Hiện tại luận án nhấn mạnh giá trị học thuật ở việc đề xuất mô hình lai ghép (hybrid DL), chưa đặt trọng tâm vào so sánh toàn diện. Trong nghiên cứu tiếp theo, NCS sẽ nghiên cứu thêm một số baseline SOTA trong EDM để kết quả có tính thuyết phục cao hơn.
5	Chưa phân tích sâu yếu tố đặc thù Việt Nam, dẫn đến mô hình chưa tối ưu hóa địa phương; Đặc biệt, mô hình dự báo ở Chương 2 dừng lại ở mức độ đơn biến, chưa tính tới yếu tố đa biến.	Tiếp thu. Ở Chương 3, nghiên cứu đã mở rộng sang mô hình đa biến (tích hợp yếu tố học tập, môi trường, xã hội). Trong tương lai, NCS sẽ khai thác thêm các đặc trưng đặc thù của Việt Nam (chính sách, chương trình đào tạo tổng thể) để tối ưu hóa mô hình cho bối cảnh địa phương.
6	Chương 2 và 3 lặp lại một số phần	Tiếp thu, sự trùng lặp khi sử dụng lại GAN cũng thể hiện mối liên kết học thuật giữa mô hình của chương 2 và chương 3.

	(như GAN). Bổ sung pineline cho từng mô hình. Bổ sung nhấn mạnh liên kết chặt chẽ giữa dự đoán SGPA và loại tốt nghiệp.	NCS đã bổ sung các pineline cho từng mô hình đề xuất trong luận án (trang 47, 58, 80, 95) Luận án đã nhấn mạnh sự chuyển tiếp từ dự báo ngắn hạn (SGPA - hồi quy) sang dự báo dài hạn (tốt nghiệp - phân loại).
7	Về hình thức: Không nên phân biệt nghiên cứu trong và ngoài nước (trang 16)	Tiếp thu. NCS đã chỉnh sửa luận án. "Nghiên cứu trong nước/ngoài nước" sẽ được diễn đạt lại theo hướng học thuật, tránh phân biệt. Các tiểu mục 1.3.1 và 1.3.2. được viết gộp lại thành 1.3.1. Related works (trang 16-18)
8	Nên có các giải thích về sự hợp lý trong các lai ghép mô hình	Tiếp thu. Với các mô hình lai ghép, luận án đã nêu ngắn gọn về lý do kết hợp. NCS chỉnh sửa luận án, trình bày chi tiết cụ thể hơn về lý do lựa chọn mô hình (ở các tiểu mục: 2.2.1, 2.3.1, 3.3.1 và 3.3.2).
9	Trình bày thuật toán nên theo format truyền thống: Input, Output (không đánh số dòng lệnh), Begin – End mới bắt đầu đánh số.	Luận án đánh số dòng của thuật toán theo khuôn dạng của IEEE.
10	Các thuật toán đề xuất hiện chưa có đánh giá sơ bộ về độ phức tạp, thời gian tính toán, có thể xem như phụ thuộc vào độ phức tạp của các mô hình cơ bản được sử dụng không?.	Tiếp thu. Các mô hình đề xuất phụ thuộc vào độ phức tạp của các mô hình cơ bản được sử dụng. NCS đã bổ sung vào luận án, trang 48, 60, 81, 96, 97.
11	Trong phần thực nghiệm nên có các thông số về cấu hình phần cứng để dùng cho việc tham khảo	Tiếp thu. NCS đã bổ sung cấu hình phần cứng đang dùng vào luận án, các phần mô tả thực nghiệm, trang 50, 51 (mục 2.2.3), trang 60 (mục 2.3.3), trang 82 (mục 3.2.3) và trang 99 (mục 3.3.3)..

	khác sau này.	
12	Chương ba cần nói thêm nội dung công bố ở CT2	Tiếp thu. NCS đã chỉnh sửa và thêm nội dung trong chương 3, mục 3.1.2, trang 75.
13	Mục Motivation of the dissertation cần tóm lược thông tin, nội dung cô đọng.	Tiếp thu. Mục 1. Motivation of the dissertation đã được chỉnh sửa ngắn gọn thành 1. General Introduction trang 1 của luận án.
14	Nên thêm thực nghiệm chương 2 với các bộ dữ liệu khác nhau	Tiếp thu. Trong chương 2, thực nghiệm đã được triển khai với 06 bộ dữ liệu (trong nước và quốc tế).
15	Phân tích thêm thông tư 42/2021 về bộ dữ liệu của cơ sở giáo dục của BGD&ĐT trước khi kết luận "However, in education, there is currently a lack…."	Tiếp thu. NCS đã bổ sung thông tin phân tích trong Trang 1 của Luận án.
16	Trong nghiên cứu tương lai, nên bổ sung thêm các yếu tố ảnh hưởng đến kết quả học tập khác như thái độ, sự tích cực học tập	Tiếp thu.
17	Bổ sung vào phần Research Subjects về các mô hình học sâu	Tiếp thu. NCS bổ sung trong trang 3 của luận án.

	Bổ sung sơ đồ khối trực quan Hệ thống thông tin, từ đó chỉ ra kết quả của luận án phục vụ khối chức năng nào	Tiếp thu. NCS đã bổ sung sơ đồ, trang 5 của Luận án.
18	Mục Contents nên để trước mục Symbols… Mục Symbols nên xếp theo a,b,c	Tiếp thu. NCS đã rà soát, chỉnh sửa (trang iii, trang vi của luận án).
19	Thuật toán cần đánh số theo chương	Tiếp thu. NCS đã chỉnh sửa (Thuật toán 2.1, 2.2, 3.1 và 3.2).
20	Các phần phụ lục chương nên để phần phụ lục chung của luận án, hoặc cuối chương 1.	Ý kiến góp ý rất xác đáng về bố cục luận án. Tuy nhiên, việc bố trí phụ lục ngay sau mỗi chương đã được NCS và tập thể hướng dẫn cân nhắc kỹ, với mục tiêu tạo sự thuận tiện cho người đọc trong việc đối chiếu trực tiếp với nội dung thực nghiệm, thay vì phải tra cứu ở cuối luận án. Cách trình bày này cũng phản ánh đặc thù của từng bài toán: tuy cùng khai thác mô hình Transformer nhưng được áp dụng theo các cách tiếp cận khác nhau.
21	Rà soát về các chú thích Hình vẽ, Bảng biểu cố gắng không nên ngắt trang ở cả 2 bản luận án và tóm tắt.	Tiếp thu. NCS đã rà soát chỉnh sửa bản luận án và tóm tắt.
22	Rà soát một số chỗ, cách dùng từ và hiệu đính tiếng Anh trong luận án.	Tiếp thu. NCS đã rà soát chỉnh sửa cách dùng từ và tiếng Anh trong luận án.

	NCS cần kiểm tra toàn bộ luận án để sửa các lỗi hình thức/trình bày.	NCS đã kiểm tra cẩn thận toàn bộ luận án và chỉnh sửa các lỗi liên quan đến hình thức/trình bày.

Nghiên cứu sinh chân thành cảm ơn Quí thầy, cô trong Hội đồng đánh giá luận án tiến sĩ cấp Học viện đã góp ý và tạo cơ hội cho NCS hoàn thiện luận án của mình.

Xin trân trọng cảm ơn./.

Hà Nội, ngày 20 tháng 10 năm 2025

TẬP THỂ HƯỚNG DẪN
(Trường hợp có 02 người hướng dẫn xin chữ ký cả 02 người, ký và ghi rõ họ tên)

Nguyễn Hữu Quỳnh **Ngô Quốc Tạo**

NGHIÊN CỨU SINH

Nguyễn Thị Kim Sơn

CHỦ TỊCH HỘI ĐỒNG

Nguyễn Long Giang

THƯ KÝ HỘI ĐỒNG

Trần Đức Nghĩa

XÁC NHẬN CỦA HỌC VIỆN
KHOA HỌC VÀ CÔNG NGHỆ

KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Nguyễn Thị Trung