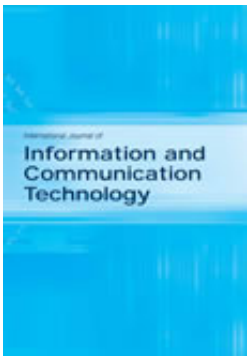


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Educational resource allocation optimisation driven by multimodal feature fusion

Junhua Hao

DOI: 10.1504/IJICT.2025.10072769

Article History:

Received:	17 June 2025
Last revised:	08 July 2025
Accepted:	08 July 2025
Published online:	26 August 2025

Educational resource allocation optimisation driven by multimodal feature fusion

Junhua Hao

School of Business,
Anyang Institute of Technology,
Anyang 455000, China
Email: ayhjh0624@163.com

Abstract: As information technology and artificial intelligence grow quickly, intelligent transformation in education has become a major trend. This study suggests an intelligent educational resource allocation model based on multimodal feature fusion-driven (MERA) to fix the flaws with the current system, which are that it is static, not very responsive, and not very personalised. MERA combines the transformer structure, self-attention mechanism, and graph neural network (GNN) with multi-objective optimisation strategies to provide a detailed model and dynamic resource allocation for complicated, varied educational data. To fully test the model's performance, three related experiments are planned and carried out. The results reveal that the MERA model is far better at using resources efficiently. In general, this study gives intelligent educational resource management a new technical path and a theoretical base.

Keywords: multimodal feature fusion; intelligent educational resource allocation; transformer; graph neural network; GNN; multi-objective optimisation.

Reference to this paper should be made as follows: Hao, J. (2025) 'Educational resource allocation optimisation driven by multimodal feature fusion', *Int. J. Information and Communication Technology*, Vol. 26, No. 31, pp.105–125.

Biographical notes: Junhua Hao received his Master degree from Dongbei University of Finance and Economics in June 2008. He is currently a Lecturer at Anyang Institute of Technology in China. His research interests include management, educational principles and multi-objective optimisation.

1 Introduction

1.1 Background of study

The education sector is going through a huge digital and intelligent transition because of the constant development of artificial intelligence, data mining, and big data processing technologies (Cantú-Ortiz et al., 2020). The standard approach for allocating educational resources mainly uses administrative regulations, empirical judgements, or static indicators, and it does not give accurate answers to the needs of each student or the needs of different educational situations. This method is hard to use in the present education

ecosystem because it is so diverse and changes so quickly. This is especially true when resources are limited and there are conflicts in educational justice.

There are now many new types of education, such as online learning platforms, smart classrooms, and personalised learning systems. These new types of education give us a lot of technical tools and situations to collect and analyse education data. While they are running, these systems create a lot of different and large amounts of data, such as students' behaviour patterns, learning progress, exam scores, interactive voice, teaching videos, courseware text, and other types of information. Multimodal data has more semantic layers and is more like the true process of teaching and learning than single-modal data. This makes it much more useful for making resource decisions (Gandhi et al., 2023).

But it is still hard to figure out how to get the most important information out of these complicated and varied datasets and combine information from multiple modes in a useful way. Multimodal feature fusion is a key topic in cross-modal learning research. In the last few years, it has made great progress in computer vision, natural language processing, and recommender systems. In intelligent education situations, this technical framework is expected to help solve the problem of data fragmentation and help us better understand and model the needs of learners, which will lead to better support for allocating educational resources.

By building a deep model architecture that combines GNN, transformer, self-attention mechanism, and other tools, it can effectively model the higher-order relationship and dynamic dependence between multimodal data. This will improve the model's ability to represent students' portraits and the generalisation ability of resource scheduling strategy. Also, using both optimisation algorithms and scheduling techniques together can make resource allocation even more real-time, fair, and efficient, which is a technical assurance for the smart administration of educational resources.

So, the topic of optimising the allocation of educational resources based on multimodal feature fusion is at the cutting edge of the intersection of educational technology and artificial intelligence. This research not only helps move educational resources from rough allocation to intelligent optimisation but also gives us a new way to think about how to use computer technology in intelligent education.

1.2 Objective and significance of study

The study wants to create a multimodal feature fusion-driven intelligent educational resource allocation model, or MERA for short. MERA will fully explore and combine different types of data from the educational environment, such as students' behavioural records, course text content, teaching images, interactive speech, and more. It will then create a single multimodal expression representation to accurately perceive, dynamically schedule, and intelligently match educational resources. The study looks at how to make educational resources more useful, distributed, and responsive in a multimodal information environment. It also looks at how to use deep fusion technology and graph modelling methods to find the best way to go from data perception to feature expression to decision optimisation.

As part of the model design process, the transformer structure, GNN, and self-attention mechanism from deep learning are combined to make the system better at modelling complicated, different kinds of interactions. At the same time, the resource scheduling technique uses multi-objective optimisation ideas to find a balance between

fairness, personalisation, and timeliness in the allocation results (Alkurd et al., 2020). This study not only wants to fix the problems with existing educational resource allocation strategies that make them static and unresponsive, but it also wants to create an intelligent decision-making framework that can be expanded and moved to meet the different needs of educational management systems.

This research is essential for both theory and practice:

1 Theoretical level: this study combines multimodal learning, graph computation, and resource optimal scheduling techniques to suggest a cross-modal feature modelling and driving mechanism for educational situations. This adds to the technical depth and research scope of using AI in smart education. The MERA model looks at a closed-loop path from bottom-layer perception to top-layer reasoning, which is especially useful for multimodal embedded representation and allocation mechanism linkage modelling. It solves the problem of fragmentation between data processing and resource decision making in traditional methods and helps task-oriented AI system construction methods become more common in education.

2 Application level: the MERA model can help school administrators better allocate resources so they can better respond to the rising variations in demand for education and pressure on resources. With the help of multimodal data, the system can see the status of learners and the demand for instruction in real time, allocate resources on demand, and make changes as needed. This helps to cut down on waste and bias in resource allocation. MERA can also help with algorithms and technology in a variety of common situations, such as recommending personalised lessons, managing regional education in a balanced way, and improving the services offered by online education platforms.

3 Methodological level: at the feature expression level, a fusion strategy for the different types of education is suggested. This breaks the current mainstream methods' reliance on a single modality. At the allocation strategy level, graph structure learning and dynamic optimisation techniques are used to make resource scheduling understandable and flexible, and to make the system more stable and adaptable during deployment.

In short, this study looks at the deep connection between multimodal feature fusion and intelligent resource scheduling. By building the MERA model, it not only adds to the body of research on intelligent education, but it also gives a methodological basis and a way to build personalised and intelligent education service systems in the future. This is very useful for research and promotion and has a lot of social value.

2 Theoretical foundation and related research

2.1 Theoretical foundation of multimodal feature fusion

As information technology grows quickly, data sources are becoming more diverse and varied. This is especially true when different types of data, like images, text, speech, video, sensor signals, and so on, are all present at the same time. Multimodal feature fusion has become an important technology for improving the system's ability to think and make decisions (Zhang et al., 2020). Multimodal fusion is more than just putting

together data from diverse sources; it is also the process of getting useful information from different sources and combining it into one body of knowledge. Fusion combines the best parts of several modalities to make up for the shortcomings of single-modal information expression which results in a more expressive and generalised joint representation.

In theory, multimodal fusion has two fundamental problems: aligning the different modes and combining the features. The idea behind multimodal fusion is that diverse modalities can be aligned in terms of time, space, meaning, and even structure. This way, features from different modalities can match up at the same semantic level. Because the data types and representations of different modalities are so different, like how text sequences and image pixel matrices are structured differently and how speech signals are continuous over time while text is made up of separate word sequences, these differences make it hard for direct fusion to get a good joint representation.

Modal alignment is commonly done with modality-specific encoders to fix this problem. There are M various types of modalities, and the input data for each one is represented as x_i. Each modality's deep coding function f_i extracts and transforms the features of x_i.

$$h_i = f_i(x_i), \qquad i = 1, 2, \ldots, M \tag{1}$$

where h_i is the semantic embedding vector for the i^{th} modality. The encoder usually uses deep learning structures like convolutional neural network (CNN), recurrent neural network (RNN), transformer, and others to create the right network structure for the data of each modality (Zhao and Ji, 2022). This lets the different modalities be mapped into the same potential representation space.

The quality of fusion is directly affected by modal alignment. After coding and mapping, the modal features should be close to each other in the semantic space. This shows how relevant distinct modal information is to each other. For instance, in a visual question-and-answer system, the image and text descriptions should be able to show the same scene or meaning after they have been aligned. Many studies have used cross-modal comparison learning approaches to improve the alignment effect even further (Zheng and Zhang, 2020). These methods encourage inter-modal semantic consistency by making positive sample pairs as similar as possible and negative sample pairs as distinct as possible.

The next step after modal alignment is to combine the features. The way multimodal features are combined influences how well the final fusion representation works and how well the job that comes after it works. There are three main types of fusion strategies: early fusion, middle fusion, and late fusion. Early fusion is when you directly splice or fuse at the original feature level (He and Liu, 2021). This is simple and intuitive, but it is hard to deal with the differences in distribution and time between heterogeneous modalities. Mid-term fusion takes into account the expressive ability and the capture of interaction information by fusing at the deep feature representation level. This is thought to be an effective way to do multimodal fusion. Late fusion is when the decision-making results of each modality are merged after they have been processed by an independent model. This is good for model integration and multitasking, but it may lose the fine-grained interaction information across modalities.

The attention mechanism is often utilised in the mid-term fusion session of deep learning frameworks which can give distinct modalities different weights and change

how much each modal feature contributes based on the input samples' contextual information. This makes the fusion representation more flexible and focused. Its main way of doing math is commonly shown as:

$$z = \sum_{i=1}^{M} \alpha_i h_i \tag{2}$$

$$\sum_{i=1}^{M} \alpha_i = 1 \tag{3}$$

In this case, w is the parameter vector that needs to be trained, and α_i is the weight that shows how important the i^{th} modality is. The model can automatically change the weight distribution based on the input samples after training. This makes the information expression more accurate. The attention mechanism can also effectively block out noise and improve the fusion quality when the information between modalities is unbalanced or some modalities are noisy.

GNN has also been used in multimodal fusion, along with the weight-weighting fusion method. GNN provides feature interaction and integration through message transmission and node update mechanisms by building a heterogeneous graph between multimodal features (Xu et al., 2022). In this graph, nodes represent modal features or entities and edges show how they are related. The graph structure may model complicated higher-order dependencies, find the hidden connections between modalities, and make the fusion representation more expressive and stable. This structured integration gives a deeper semantic meaning to multimodal information, especially in areas like scene data and social networks.

To sum up, the theory of multimodal feature fusion deals with two main problems: how to align modes and how to combine features. Modal alignment fixes the differences between distinct modalities so that they can all be shown in the same potential space. Feature combination, on the other hand, uses a number of methods to effectively combine information from different modalities. The ongoing development of ideas and methods has led to widespread use of multimodal technology in many areas. This technology is the foundation for building intelligent systems that can perceive and reason well.

2.2 Current research status of intelligent educational resource allocation

Smartly allocating educational resources is a key aspect of making personalised education a reality. The goal is to meet the learning needs of each student and raise the overall quality of teaching by making the best use of limited resources. As the idea of intelligent education becomes more popular and technology improves, the problem of resource allocation has gotten more difficult, and the data dimensions and interaction interactions involved have become more and more complicated. To deal with this complexity, academics have come up with a number of algorithms and models from different points of view, all with the goal of making resource allocation as optimal and adaptable as possible.

In the beginning, the field of intelligent resource allocation was mostly based on traditional mathematical planning approaches including linear programming, integer programming, and dynamic programming (Forootani et al., 2021). These methods use clear objective functions and limitations to mathematically describe the resource

allocation problem and find the best or worst solutions. As an example, linear programming can be used to schedule resources at the system level by creating an objective function that maximises the use of learning resources or student happiness, along with limitations like the number of resources and time schedules. But in real educational settings, there are challenges with multiple objectives, multiple constraints, and uncertainty that make standard planning methods less effective and less scalable. This is especially true when dealing with large-scale dynamic change scenarios.

Population intelligence algorithms have become a popular area of research to fix these problems. Genetic algorithms (GA) are often used to allocate resources because they replicate the process of biological evolution and find the best answer for everyone by using selection, crossover, and mutation procedures. It does not need to use gradient information, which is a plus, and it works well for nonlinear and non-convex optimisation problems. On the other hand, particle swarm optimisation (PSO) uses the way bird flocks look for food to efficiently move through the search space by having groups work together. It can also swiftly find the best solution. The ant colony algorithm (ACO) works effectively for path planning and combinatorial optimisation issues by mimicking how ants move pheromones (Tutuko et al., 2018). These algorithms all have strong global search capabilities and are robust enough to handle the complex constraints and multi-objective optimisation needs of allocating educational resources. However, they also have the problems of being hard to adjust parameters and having a slow convergence speed.

As machine learning technology has improved, intelligent recommended systems have become one of the most important tools for allocating educational resources. Collaborative filtering (CF) algorithms look at how similar users' past behaviours are to offer learning resources that fit students' requirements and interests. The model-based CF method uses matrix decomposition and deep learning to make recommendations more accurate and useful in more situations. But CF is sensitive to the cold-start problem and is hard to deal with the changing needs of each learner. Content-based recommendation algorithms make up for this by looking at the feature information of resources and students' learning profiles to find better matches for each student. Deep learning approaches have been utilised a lot in the last few years for things like extracting resource features and modelling student behaviour. Structures like CNN and RNN can automatically find complex feature patterns to help make recommendations and schedules more accurate.

Reinforcement learning (RL) is especially good at allocating resources in changing contexts because it learns by interacting with them. People think of resource scheduling as a challenge of making decisions in a sequence, where people choose actions based on the status of the environment to get the most long-term benefits. Deep reinforcement learning (DRL) uses the representation power of deep neural networks to process high-dimensional state spaces effectively. For example, with deep Q networks (DQN), intelligence can enhance the allocation strategy by continuously trying and failing to get closer to the state-action value function (Talaat, 2022). But its training process is complicated and depends heavily on modelling the environment, so it needs to find ways to make its samples more stable and efficient in real-world use.

Hybrid algorithms, which have become more common in the last several years, combine the best parts of other methodologies. Also, hybrid methods that combine group intelligence algorithms with deep learning improve the optimisation even more by helping the search process through deep feature extraction. These kinds of strategies are

more flexible and robust when it comes to allocating educational resources, and they work well for projects that are multimodal and have several goals.

Also, intelligent resource allocation strategies based on multimodal fusion are becoming a hot topic in research as multimodal data collection becomes more prevalent. Multimodal fusion uses the multidimensional aspects of students' behavioural data, cognitive states, sentiment analysis, and teaching resources to create a rich feature space that gives the resource allocation model correct information support. GNN models the relationship graph between students and resources to capture the complex structure of their interactions and match resources to students in a way that takes the context into account. This strategy not only makes personalised recommendations more accurate, but it also makes the model easier to understand and more adaptable.

Overall, research in the area of intelligent educational resource allocation has changed throughout time, with algorithms changing from classical optimisation to intelligence and dynamics. To move intelligent education in the direction of higher personalisation and accuracy, researchers need to keep looking into algorithm design and system implementation.

3 Model design and optimisation methods

3.1 Model construction

The MFRE model is a system that uses deep learning and graph computing technologies to sense features across different modes and schedule resources dynamically for smart educational situations. See Figure 1. The model structure is set up in a hierarchical way, with four key parts: the feature expression layer, the multimodal fusion layer, the graph structure modelling layer, and the resource allocation optimisation layer.

3.1.1 Feature representation layer

The feature representation layer of the MFRE model has a feature encoding mechanism that works well and is specific to the different types of input data from different sources in an educational setting. The goal is to get a detailed picture of student behaviour, teaching resources, and environmental information all in one place. First, the layer cleans up the raw data and makes it more stable by removing noise and outliers. Next, different types of deep learning structures are utilised to find the most important attributes in each modality. This makes the expression more interesting and useful.

For the time series class of student behaviour data, the MFRE model adds the transformer encoder based on the self-attention mechanism. This encoder can pick up on long-distance dependencies and changes that happen over time in the learning process (Fu et al., 2023). Assuming the input sequence is X, we can write X as:

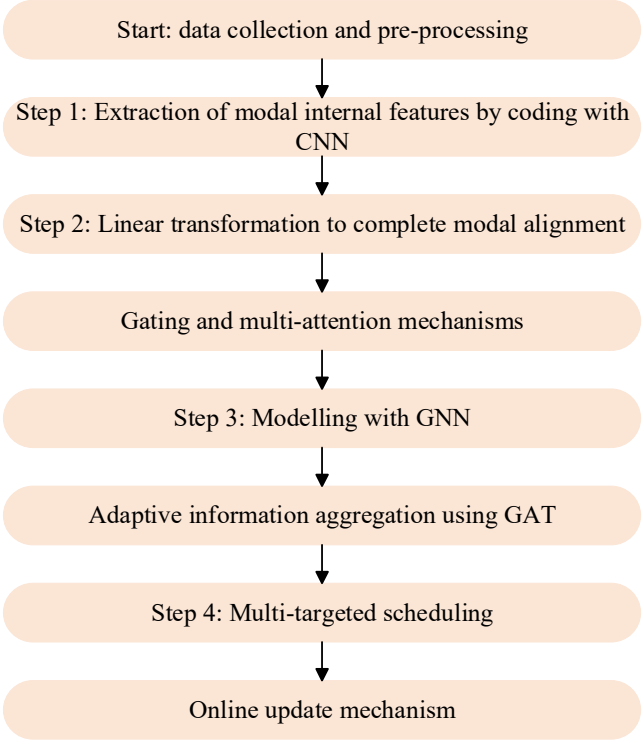
$$X = \{x_1, x_2, \ldots, x_T\} \tag{4}$$

where x_t is the feature vector at moment t. The self-attention mechanism gives more weight to the significant time nodes by figuring out how the query matrix Q, the key matrix K, and the value matrix V are related. This is how the formula looks:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{5}$$

where $\sqrt{d_k}$ is a scaling factor that keeps the dot product from getting too big, which would cause the gradients to disappear or explode. This method lets the model automatically focus on important learning behaviour segments, as when students keep trying to solve hard problems or review knowledge points a lot, to pick up on little changes in how motivated and effective they are at learning.

Figure 1 Structure of MFRE model (see online version for colours)



The methodology uses CNN to encode multi-dimensional information about structured or semi-structured data like instructional materials and environmental information. This includes things like content attributes, difficulty level, frequency of use, and instructor feedback on the resources. For example, the feature mapping method for the input feature vector x_r is:

$$h_r = \sigma(Wx_r + b) \tag{6}$$

where W is the weight matrix and b is the bias vector. The activation function σ is chosen to be leaky ReLU to make the nonlinear expression better. The encoder slowly improves the deep semantics of resource and environmental features through multi-layer stacking and regularisation approaches. It also makes features better at telling the difference between things and generalising.

The feature expression layer also looks at how to unify the scale and align the distribution of different modal features. It uses batch normalisation and layer normalisation to help with the problem of gradients disappearing during training, which makes the model more stable and speeds up the training process (Garbin et al., 2020). In the end, this layer maps the coding results of each modality to a single high-dimensional vector space. This gives multimodal fusion structured and fully articulated inputs, making sure that the information from diverse modalities may interact and function together efficiently.

In short, the layer builds a multidimensional feature representation framework by combining the self-attention mechanism and deep coding technology. This framework can not only deeply explore the internal structure of a single modality, but it can also lay a solid foundation for cross-modal fusion. This not only helps the model understand complicated, mixed-up educational data better, but it also gives it a strong foundation for later intelligent resource scheduling.

3.1.2 Multimodal fusion layer

The multimodal fusion layer is the main part of the MFRE model that connects the feature expression layer and the resource scheduling optimisation layer. Its main jobs are to combine data from different sources, make sure that information from different modes is aligned, and allow for deep interaction. Data in the education field includes things like student behaviour, teaching materials, and environmental factors. These things have structural and semantic differences and merging them directly could cause information loss or conflict. Therefore, it is important to achieve modal alignment first, so that the different modal features can be mapped to a single potential space. Let h_i be the feature representation of the i^{th} modality. To finish the alignment, use linear transformation:

$$\tilde{h}_i = W_i h_i + b_i \tag{7}$$

where W_i and b_i are trainable parameters that make the modal features constant in size and distribution, get rid of scale disparities and bias, and make it easier to combine them later.

After the alignment is done, the model adds a gating mechanism that changes the relevance of each modal characteristic on the fly (Tan et al., 2021). This stops the information from being redundant or conflicting, which can happen with simple superposition. In particular, the following form is used:

$$g_i = \sigma\left(W_g \tilde{h}_i + b_g\right) \tag{8}$$

$$z = \phi\left(\sum_{i=1}^{M} g_i \odot \tilde{h}_i\right) \tag{9}$$

where g_i is the sigmoid activation function that makes the gating vector, ϕ is the nonlinear activation, and $\odot$ is the product at the element level. This gating method can change the weights based on the task requirements and modal properties to make the fusion expression more relevant and useful.

The MFRE fusion layer adds a multi-head attention mechanism to better capture the complex and multi-level interactions between modalities. This lets the model focus on

different modal feature interactions in different subspaces by parallelising multiple attention heads. Set the query, key, and value matrices as Q, K and V. The steps in the multi-head attention calculation are:

$$\text{head}_i = \text{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \qquad i = 1, \ldots, h \tag{10}$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}\left(head_1, \ldots, head_h\right)W^O \tag{11}$$

where W_i^Q, W_i^K, W_i^V and W^O are trainable parameter matrices. The multi-head attention not only helps the model find more detailed modal relationships, but it also makes the fusion results more diverse and robust, which helps it deal with the problems of data heterogeneity and changing conditions in educational situations.

The fusion layer also makes sure that deep network training is stable and that information flows smoothly by using residual connection and normalisation methods. The multimodal fusion output has a lot of features (Kong et al., 2023). It shows the deep semantic synergy between modalities and the relative contribution of each modality to the current task. This gives an accurate and discriminative input basis for the next steps of graph structure modelling and resource allocation optimisation.

In short, the MFRE model's multimodal fusion layer combines modal alignment, gating fusion, and a multi-attention mechanism to get a deep understanding of and dynamic trade-offs of different types of educational data. This greatly improves the model's ability to perceive and make decisions about intelligent educational resource allocation, and it provides a solid technological guarantee for building a highly efficient and intelligent educational resource management system.

3.1.3 Graph structure modelling layer

In the MFRE model, the graph structure modelling layer takes over the multimodal features output from the fusion layer, aiming at mining the deep semantic connections between many types of entities and interactions in the education system. MFRE creates a GNN-based modelling scheme to better show these interactions. This scheme makes full use of the expressive capability of graph structure to improve the model's ability to understand and make inferences about systems. Data items in real educational contexts exhibit clear graph structure aspects, for example, the relationships between students and courses, teachers and resources, and social interactions among students collectively constitute a dynamic, diversified and heterogeneous network structure.

The model first makes multimodal things like learners, courses, and resources into node sets V and relationships like behaviours, teaching, and cooperation into edge sets E. This gives us the graph structure G:

$$G = (V, E) \tag{12}$$

The output of the multimodal fusion layer gives each node an embedding vector. The strength of the edges can change depending on how often certain behaviours happened in the past, how strong the relationships between nodes are, or how similar the content is. The different types of nodes and edges in the graph constitute a heterogeneous network, giving a rich structural priority for the model.

MFRE uses graph attention networks (GAT) to provide adaptive information aggregation in this graph structure so that it can learn context-aware node representations

(Wei et al., 2022). For each node v in the graph, its representation is updated by aggregating the features of nearby nodes u and weighting the sum:

$$h_v' = \sigma\left(\sum_{u \in N(v)} \alpha_{vu} W h_u\right) \tag{13}$$

where h_u is the neighbour node feature, W is the shared weight matrix, α_{vu} is the attention weight of node u to node v and σ is the activation function. This method enables nodes to focus on their more semantically important neighbours during representation update, effectively increasing the discriminative and contextual adaptability capabilities of feature representation.

To better reflect the diverse interaction structure in educational settings, MFRE adds a relationship type encoding and meta-path aggregation technique based on GAT to help share information between different types of nodes. For example, student-course-resource connections illustrate indirect correlations of resource requirements, while teacher-course-student paths reflect chains of pedagogical influence. The model can create a more hierarchical and explanatory embedded representation by controlling how certain paths spread through the graph.

The final output of the graph structural modelling layer is a global context-aware, structural relationship-driven set of node embeddings that give high-quality inputs to help further resource scheduling optimisation. This layer makes the model more adaptable to complex network structures and connects individual behaviours to system hierarchies in intelligent educational systems. This makes the model stronger and better able to work in changing educational settings.

3.1.4 Resource allocation optimisation layer

The resource allocation optimisation layer is the MFRE model's decision-making terminal. Its main goal is to use the integrated multimodal feature expression and graph structure modelling results to create an intelligent schedule and the best use of educational resources. In real life, educational resources like courses, teachers, and learning tools are often limited and structured. At the same time, users' needs are varied and changed over time. This makes it hard for traditional static allocation or rule-based systems to meet the need for accurate services in large-scale intelligent educational systems (Sophia, 2025). So, MFRE adds a multi-objective optimisation technique at this level to find a compromise between the fairness, personalisation, and timeliness of the resource allocation process.

Resource allocation optimisation is modelled as a multi-objective scheduling problem, where each objective function measures the system performance from a different optimisation point of view. For example, improving customer pleasure while preserving overall load balance, or guaranteeing optimal resource utilisation while enhancing services for disadvantaged groups. To do this, the model looks at a number of different metrics to create an optimisation objective vector and make scheduling choices that work within limits. In formal terms, it can be said as:

$$\min_{Y} L = \lambda_1 \cdot L_{fair} + \lambda_2 \cdot L_{personal} + \lambda_3 \cdot L_{timely} \tag{14}$$

where Y is the resource allocation matrix; the three loss functions are the fairness loss of allocation, the personalisation bias loss, and the response timeliness loss; λ_i is the

adjustable weight coefficient, which lets you switch strategies based on the needs of different educational situations. During training, the gradient optimisation approach can be used to change this objective function repeatedly. It can also be used with reinforcement learning or a dynamic feedback system to make the performance better when it is deployed online.

In real life, MFRE first uses the graph embedding output from the previous layer to figure out how well students and resources match up. Then it uses the multi-objective optimisation function to come up with the final allocation method. To improve practicality and scalability, the system additionally constructs a priority queue-based scheduler, which makes secondary corrections and filters the scheduling results to assure the overall optimal efficiency in resource-limited settings.

MFRE also has an online updating feature that lets you change the scheduling approach while tasks are running. The system can recognise changes in user behaviour and resource status in real time, combine past allocation records with expected trends, update and optimise goal weights, and create a closed-loop resource management process (Righi et al., 2019). This technique is particularly useful for complicated educational activities such as customised learning path suggestion, cross-regional resource deployment and multi-terminal resource collaboration.

The resource allocation optimisation layer is designed so that MFRE can not only work well with static data but also schedule tasks quickly and easily in changing situations. This layer effectively closes the gap between understanding and action by combining in-depth expression, structural perception, and decision feedback. It does this by showing the full path from multimodal perception to educational intervention and providing algorithmic guarantees and technical support for the long-term growth of intelligent educational systems.

3.2 Evaluation indicators

This research builds a set of normalised evaluation index systems from four different angles to fully evaluate the real-world consequences of the MFRE model on the intelligent educational resource scheduling job. These signs show that the model is better at the algorithmic level and that it can really be used in educational settings.

First, resource utilisation is used to see if the system is being used enough for the number of resources it has (Kjaer et al., 2019). This indicator looks at whether the system can make the best use of resources for users who require them without adding to the total quantity of resources. The overall service capacity of the system and the rationality of resource allocation both depend on how well resources are allocated. This is a crucial factor in deciding whether the resource scheduling method is effective.

Second, personalised matching shows if the resource distribution genuinely meets the demands and characteristics of each user. In smart education settings, various users have very varying needs and wants for resources. Because of this, the system's ability to effectively match users based on their profiles is the most important way to determine how smart it is. A higher level of personalised matching means that the system can better recognise what users need and offer them varied and accurate teaching resource push services. This makes learning more successful and satisfying.

Third, the fairness indicator checks to see if resources are being shared fairly across different groups of users. As a social issue, the distribution of educational resources, in addition to the goal of efficiency, needs to consider fairness, to ensure that all types of

user groups have access to the right services. This indicator gives a numerical score to the level of balance in the distribution of resources. It also checks to see if the system has major problems with resource concentration or skewed distribution. The system works better at protecting vulnerable users and keeping the balance across regions when the allocation is more even.

Finally, the response efficiency indicator looks at how quickly the system can respond to a request and finish allocating resources. In large-scale education systems, real-time is a key part of making sure that the service is good and that scheduling may change as needed (Canizo et al., 2019). Higher response efficiency means that the system can process information and give feedback in real time better. It can also change based on the needs of users and the state of resources, which makes it suitable for highly dynamic situations like online education.

Overall, these four assessment indicators create a performance evaluation framework that looks at things from many angles and is easy to use. It can fully show the MFRE model's main skills in allocating intelligent educational resources. It not only looks at how accurate and efficient the algorithm is, but also how fair and adaptable it is to change educational situations, which is very useful in real life and for guiding future use.

4 Experimental design and analysis of results

4.1 Experimental setup and dataset description

This research does an experimental design using real multimodal educational datasets to see if the suggested MFRE model works in real-life educational situations. It also sets up the assessment procedure by combining the normal task needs and model capabilities. To make sure the assessment results are accurate and consistent, the experiment includes data pretreatment, model training, parameter optimisation, and performance comparison.

Table 1 Dataset overview (EdNet)

Module category	*Data content*	*Description*
User features	Answer accuracy, active time, video watch time, click records	Describes user behaviour patterns and study habits; serves as input for user embeddings
Resource features	Question knowledge points, resource type, course difficulty, recommendation frequency	Describes static and dynamic features of educational resources
Interaction logs	User responses, question IDs, timestamps, feedback states	Represents dynamic user-resource relationships, used for heterogeneous graph construction

The EdNet dataset is used as the basis for the experiments in this study. It was released by Riiid from its e-learning platform Santa. The EdNet dataset is used as the basis for the experiments in this study. It was released by Riiid from its e-learning platform Santa and includes the interaction behaviours of more than 780,000 students. These behaviours include questions, explanation videos, time information, answer results, course structure, and more. The data is rather complex in terms of structure and usefulness. After filtering

and cleaning, this study isolates the three basic modules of user characteristics, resource attributes and interaction records, and produces a subset of experiments appropriate to multimodal modelling and resource allocation.

Table 1 shows the main information in the dataset.

To make sure that the model's training is effective and can be repeated, this paper does all the experiments in a local high-performance computing environment. Table 2 shows the specific configurations and hyper-parameter settings used.

Table 2 not only gives the experimental procedure a clear framework and genuine data, but it also makes sure that the MFRE model's ability to describe, schedule, and generalise in complicated educational situations can be rigorously examined.

Table 2 Experimental settings

Item	*Description*
Hardware platform	NVIDIA RTX 3090 GPU, Intel i9 CPU, 64 GB RAM
Software environment	Python 3.9, PyTorch 2.1, CUDA 11.7, DGL 1.1
Batch size	128
Learning rate	0.0005 (with Adam optimiser)
Maximum epochs	100 (with early stopping after 5 stagnant epochs)
Data split strategy	Time-based user-level split: 70% train, 15% validation, 15% test
Cross-validation method	5-fold cross-validation
Initialisation strategy	Xavier initialisation
Evaluation frequency	Once per epoch on validation set, for dynamic learning rate adjustment and model selection

4.2 Experiment 1: analysis of the resource allocation structure

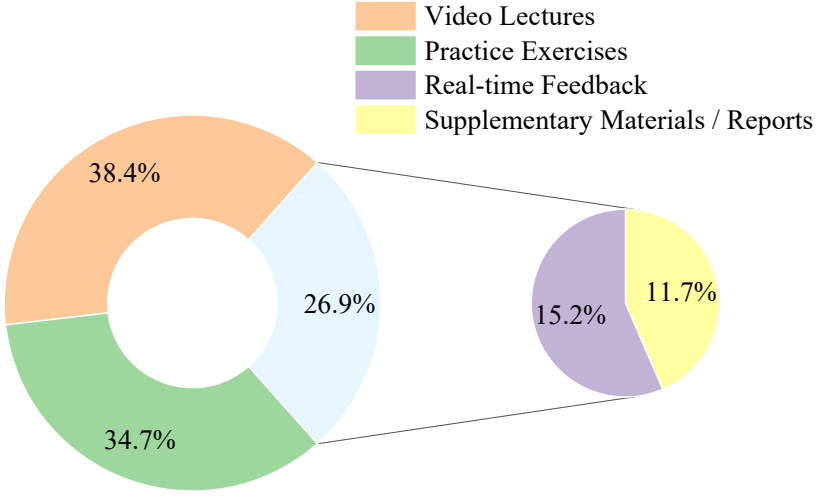
This paper first does resource allocation structure analysis experiments to see how the MFRE model responds when it has to deal with different types of resources. This is to check its adaptability and ability to schedule resources intelligently in real educational resource allocation scenarios. The experiments focus on common educational resources like video lectures, practice questions, extra materials, and real-time feedback. By looking at the proportion of resources in the model's final allocation decision, we can see how well the model understands and prefers the role weights of different resources. This lets us evaluate the system's scheduling equilibrium and strategy rationality.

A representative sample of 5,000 users is taken from the EdNet dataset and merged with their learning goals, behavioural traits, and current stage of learning. This information is then fed into the trained MFRE model to create the resource allocation plan. Next, the model's tendency to schedule resources under multimodal feature fusion is analysed by counting the percentage of each type of resource in the final recommended total. Figure 2 shows how the different types of resources are divided up in the model output:

Video lectures and practice problems take up the most space in the overall resource allocation findings, with 38.4% and 34.7%, respectively. This shows that the MFRE model, after adding multimodal characteristics, focuses on meeting learners' direct demands for knowledge explanation and practical training. This is in keeping with the present teaching strategy of 'understanding + training' in the intelligent education system.

The model automatically shows how important video is for transferring knowledge and creating concepts. It also stresses how important memory improvement and skill transfer through practice are, which shows that it has high content perception and scheduling skills.

Figure 2 Resource allocation proportions by MFRE model (see online version for colours)



Further analysis reveals that despite the relatively low allocation of real-time feedback and extension materials (15.2% and 11.7%, respectively), they are still moderately allocated by the system, suggesting that the model does not only focus on mainstream resources, but also considers the necessary role of ancillary support in the personalised learning process. The allocation ratio of real-time feedback shows that the model can change dynamically, which means it can give timely responses and guidance during the learning process. The expansion materials, on the other hand, show that the system has a differentiated supply strategy for some high-ability or independent learning users, which helps to create a multi-level and gradient learning resource system.

The MFRE model does a good job of following the allocation strategy of giving priority to core resources and balancing auxiliary resources in resource scheduling. This shows that it has a deep awareness of structure and can adapt to tasks. The proportion of different resources is not just an average; it is a smart and tailored distribution based on user profiles, resource attributes, and interaction history. This finding not only shows that the multimodal fusion technique works to represent features, but it also strongly supports the next steps of precise matching and dynamic optimisation of instructional resources.

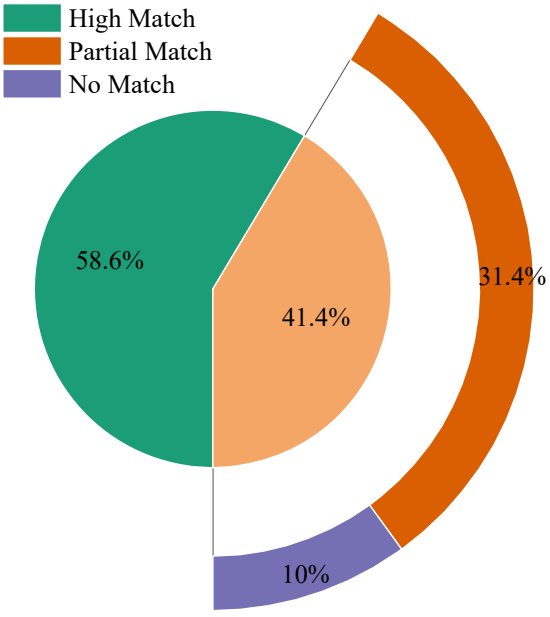
4.3 Experiment 2: overall assessment of personalised recommendation matching degree

This paper designs a second set of experiments to statistically measure how well the resource recommendation results match users' real learning needs from a global perspective. This is to check the overall adaptation performance of the MFRE model in personalised recommendation. The tests no longer look at different types of users;

instead, they look at how accurate and reasonable the recommended content is. The results are then sorted into three groups: highly matched, partially matched, and obviously mismatched. This shows how well the model can model complex groups of users as a whole.

The experiment randomly picks 3,000 EdNet users who are always learning and records the recommended content for each round after inputting the trained MFRE model. It then uses both manual annotation and behavioural validation methods to figure out how well the recommendations fit. Figure 3 shows the findings of the statistics:

Figure 3 Overall match distribution of MFRE recommendations (see online version for colours)



First, looking at the overall distribution, the MFRE model produces 58.6% of highly matched recommendation results. This shows that the model has strong reasoning and expression skills after considering user behaviour data, semantic features of the content, and scenario information. It can also accurately respond to the current learning needs of most users. This ratio reveals that the system has a good quality of content recognition and recommendation under multimodal situations, which meets the basic objectives of intelligent educational resource scheduling for accuracy and personalisation.

Second, 31.4% of the matches are partial, making them a bigger secondary element of the recommendation results. The analysis shows that this segment is largely about people who are in the knowledge migration or stage transition period, and their learning interests and behaviour labels are not very clear. The recommended contents are mostly going in the same direction, however some are deeper or harder than others. This shows that there is still opportunity for development in the model to cope with dynamic learning state changes, especially in capturing prospective learning motivations and intention prediction, which can be further increased.

Last but not least, the percentage of obvious mismatch outcomes is 10.0%. This is not a lot, but it still needs to be looked at from the point of view of model optimisation. This

kind of recommendation mistake happens most often when there are not many users, when there are problems with labels, or when a system is just starting off. MFRE uses GNN and multimodal fusion algorithms to help with this kind of difficulty, however the results show that there are still some areas that need work.

4.4 Experiment 3: contribution of key modules

This research describes ablation experiments that will help us better understand how each important module in the MFRE model affects the performance of intelligent educational resource allocation. By taking away the multimodal fusion layer, the graph structure modelling layer, and the resource allocation optimisation layer one at a time, we can see how missing modules affect the model as a whole and check that the design of each section is necessary and works. The tests are conducted using the same EdNet dataset and unified assessment indexes as the previous ones to ensure comparable findings.

The exact settings are as follows:

- complete model (MFRE): has all the design modules as a baseline for performance
- without fusion layer: takes off the multimodal fusion portion and solely uses single modal information to allocate resources
- without graph layer: takes off the GNN module and does not pay attention to the complicated relationship between users and resources in a graph
- without optimisation, there is no multi-objective optimisation approach, and the resource allocation strategy is limited to static rules.

Figure 4 displays the performance of each model version on the comprehensive assessment index.

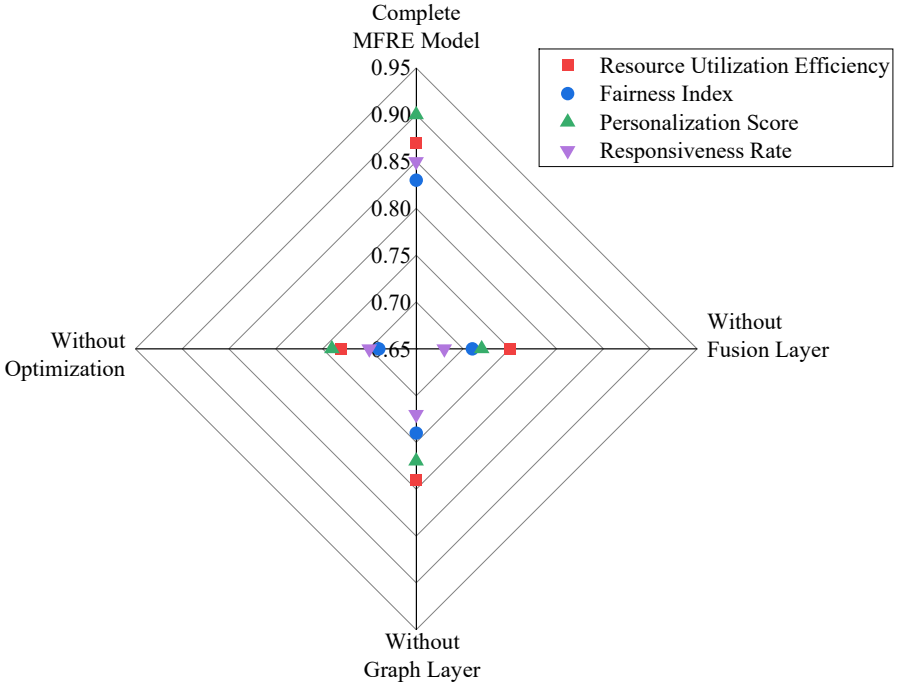
The full MFRE model has the highest scores on all the indicators. This shows how the multimodal fusion, graph structure modelling, and resource optimisation strategies used in this study work together to make intelligent educational resource allocation better. The model provides efficient and intelligent resource scheduling while taking into account numerous objectives. The resource utilisation efficiency is 0.87, the fairness index and personalisation score are both over 0.8, and the response rate is also good.

Second, the model's performance drops a lot when the multimodal fusion layer is left out. For example, the personalisation score drops from 0.90 to 0.72. This shows that multimodal fusion is very important for getting the rich information between users and resources and making personalised allocation possible. The reduction in resource utilisation efficiency and reaction rate also indicates the assisting influence of the fusion layer on overall decision accuracy and system agility.

Again, the version without the graph structure modelling layer shows a small drop in fairness and response rate metrics. This shows that GNN is necessary for modelling user-resource interaction networks and showing how different types of relationships work, which helps the model allocate resources fairly and respond quickly.

Finally, the experimental version without the resource allocation optimisation layer performs equally poorly, suggesting that the multi-objective optimisation strategy plays a significant role in balancing justice, customisation and timeliness. Without this layer, the resource allocation approach is less precise, and the system doesn't respond well to complicated educational needs.

Figure 4 Ablation experiment results of MFRE model (see online version for colours)



In short, the ablation experiment shows that each module of the MFRE model is necessary and useful. It also shows the benefits of combining multimodal fusion, graph structure modelling, and resource optimisation strategies in intelligent educational resource allocation. Finally, it gives theoretical and practical support for improving and promoting the next model.

5 Study summary and future prospects

5.1 Summary of study

This research looks at the issue of optimising the allocation of intelligent educational resources via multimodal feature fusion. It suggests an MFRE model that combines a transformer structure, a graph neural network, and a multi-objective optimisation technique. The approach gets a full picture of the learner's status, instructional materials, and environmental elements by deeply mining and effectively combining multimodal heterogeneous data. This makes educational resources much more useful.

Experiment 1 shows that the MFRE model can adapt and schedule educational resources intelligently in real-life situations. Experiment 2 shows that the MFRE model can adapt to personalised recommendations. The ablation experiment, which is the main part of experiment 3, shows that the three key modules of the MFRE model are irreplaceable. Each module plays a significant role in enhancing the indexes of fairness and response speed, which supports the scientificity and necessity of module co-design. In general, the work in this paper not only makes a big step forward in theory and

methodology, but it also gives strong support for the intelligence, dynamics, and personalisation of the smart education resource allocation system. This sets the stage for more optimisation and promotion of its use.

5.2 Limitations of study

The MFRE model suggested in this study has made a lot of progress in several areas, but there are still some problems that need to be solved:

1 The heterogeneity and quality issues of multimodal data are still the bottleneck for model performance improvement: different data sources have considerable variances in format, distribution and information expression, and missing data and noise will certainly occur in the actual collecting process. This not only impairs the model's successful fusing of multimodal data but also restricts its stability and generalisation capacity in complex and changing educational situations.

2 The diversity and dynamic changes of educational scenarios put forward higher requirements on the migration ability of the model: there are big disparities between areas, schools, and ways of teaching. The model is not yet flexible enough to cover all real-world situations, so it needs to be made more versatile and adaptable to be used in a wide range of situations.

3 The objective design of resource allocation optimisation is not comprehensive enough: the evaluation right now mostly looks at fairness, personalisation, and timeliness. However, the distribution of educational resources also involves more complicated factors, like keeping an eye on the quality of teaching, the level of professionalism of teachers, changes in subject demand, and other multi-dimensional information that have not yet been added to the model's optimisation framework.

4 Lack of long-term application practice in large-scale real education system: this paper's experimental validation is based on public datasets, which are not always complete or representative and cannot properly capture how complicated the real educational environment is.

5.3 Suggestions for improvement

To address the limitations in the above research, future work should be improved and optimised in the following areas:

1 Improve the quality and fusion capability of multimodal data: to make sure the data is accurate and consistent, you should add more data sources, make the dataset bigger, and make the process of cleaning and completing the data better in the future. At the same time, make the model more stable and able to generalise in complicated settings by making the multimodal fusion methods stronger, making the model more tolerant of missing data and noise, and making the model more stable.

2 Reduce the computational complexity of the model and improve computing efficiency: in the future, model compression methods like weight pruning, low-rank decomposition, and knowledge distillation will be used to improve the architecture of the network structure so that it is lighter. At the same time, heterogeneous

computing resources and distributed training are used to speed up model training and reasoning, decrease the hardware threshold, and make it easier for intelligent education systems to be used in real-world settings.

3 Enhance the adaptability and migration ability of models: in the future, use both online learning and meta-learning methods together. This will allow the model to swiftly adapt to varied educational situations and teaching environments that change over time (Zhou et al., 2021). It will also allow for smart resource allocation across regions and platforms. Set up a way for the model to update itself and improve itself over time to make sure it works steadily for a long time.

4 Construct a multi-dimensional resource allocation optimisation target system: include things like the quality of teaching, the level of professionalism of the teachers, the mental state of the pupils, and other factors in the optimisation framework. Then, create a multi-objective optimisation algorithm that achieves the goals of science, accuracy, and fairness in resource allocation. Make personalised allocation plans with changeable criteria to match the specific demands of each topic in school.

5 Strengthen experimental verification and application promotion in real environment: work more closely with schools and colleges and do big, long-term studies in real-life situations. Improve the system's usefulness and dependability by getting input from teachers and students and combining operational data with ongoing iterative optimisation of model performance and user experience (Chen, 2023). This will help the widespread use of intelligent educational resource allocation technology.

Declarations

The author declares that he has no conflicts of interest.

References

Alkurd, R., Abualhaol, I.Y. and Yanikomeroglu, H. (2020) 'Personalized resource allocation in wireless networks: an AI-enabled and big data-driven multi-objective optimization', *IEEE Access*, Vol. 8, pp.144592–144609.

Canizo, M., Conde, A., Charramendieta, S., Minon, R., Cid-Fuentes, R.G. and Onieva, E. (2019) 'Implementation of a large-scale platform for cyber-physical system real-time monitoring', *IEEE Access*, Vol. 7, pp.52455–52466.

Cantú-Ortiz, F.J., Sánchez, N.G., Garrido, L., Terashima-Marin, H. and Brena, R.F. (2020) 'An artificial intelligence educational strategy for the digital transformation', *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol. 14, pp.1195–1209.

Chen, Y. (2023) 'Analyzing the design of intelligent English translation and teaching model in colleges using data mining', *Soft Computing*, Vol. 27, No. 19, pp.14497–14513.

Forootani, A., Tipaldi, M., Zarch, M.G., Liuzza, D. and Glielmo, L. (2021) 'Modelling and solving resource allocation problems via a dynamic programming approach', *International Journal of Control*, Vol. 94, No. 6, pp.1544–1555.

Fu, Z., Li, J., Ren, L. and Chen, Z. (2023) 'SLDDNet: stagewise short and long distance dependency network for remote sensing change detection', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, pp.1–19.

Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E. and Hussain, A. (2023) 'Multimodal sentiment analysis: a systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions', *Information Fusion*, Vol. 91, pp.424–444.

Garbin, C., Zhu, X. and Marques, O. (2020) 'Dropout vs. batch normalization: an empirical study of their impact to deep learning', *Multimedia Tools and Applications*, Vol. 79, No. 19, pp.12777–12815.

He, Y. and Liu, Z. (2021) 'A feature fusion method to improve the driving obstacle detection under foggy weather', *IEEE Transactions on Transportation Electrification*, Vol. 7, No. 4, pp.2505–2515.

Kjaer, L.L., Pigosso, D.C., Niero, M., Bech, N.M. and McAloone, T.C. (2019) 'Product/service-systems for a circular economy: the route to decoupling economic growth from resource consumption?', *Journal of Industrial Ecology*, Vol. 23, No. 1, pp.22–35.

Kong, W., You, Z. and Lv, X. (2023) '3D face recognition algorithm based on deep Laplacian pyramid under the normalization of epidemic control', *Computer Communications*, Vol. 199, pp.30–41.

Righi, R.d.R., Lehmann, M., Gomes, M.M., Nobre, J.C., da Costa, C.A., Rigo, S.J., Lena, M., Mohr, R.F. and de Oliveira, L.R.B. (2019) 'A survey on global management view: toward combining system monitoring, resource management, and load prediction', *Journal of Grid Computing*, Vol. 17, pp.473–502.

Sophia, L. (2025) 'Optimization of intelligent education systems based on reinforcement learning', *Artificial Intelligence Education Studies*, Vol. 1, No. 1, pp.82–108.

Talaat, F.M. (2022) 'Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm', *Multimedia Tools and Applications*, Vol. 81, No. 28, pp.39945–39961.

Tan, Q., Zhang, X., Liu, H., Jiao, S., Zhou, M. and Li, J. (2021) 'Multimodal dynamics analysis and control for amphibious fly-drive vehicle', *IEEE/ASME Transactions on Mechatronics*, Vol. 26, No. 2, pp.621–632.

Tutuko, B., Nurmaini, S. and Ogi, G. (2018) 'Fuzzy logic-ant colony optimization for explorer-follower robot with global optimal path planning', *Computer Engineering and Applications Journal*, Vol. 7, No. 1, pp.61–74.

Wei, P., Zeng, B. and Liao, W. (2022) 'Joint intent detection and slot filling with wheel-graph attention networks', *Journal of Intelligent & Fuzzy Systems*, Vol. 42, No. 3, pp.2409–2420.

Xu, A., Zhong, P., Kang, Y., Duan, J., Wang, A., Lu, M. and Shi, C. (2022) 'THAN: multimodal transportation recommendation with heterogeneous graph attention networks', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 24, No. 2, pp.1533–1543.

Zhang, C., Yang, Z., He, X. and Deng, L. (2020) 'Multimodal intelligence: representation learning, information fusion, and applications', *IEEE Journal of Selected Topics in Signal Processing*, Vol. 14, No. 3, pp.478–493.

Zhao, L. and Ji, S. (2022) 'CNN, RNN, or ViT? An evaluation of different deep learning architectures for spatio-temporal representation of sentinel time series', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, pp.44–56.

Zheng, H. and Zhang, X-M. (2020) 'A cross-modal learning approach for recognizing human actions', *IEEE Systems Journal*, Vol. 15, No. 2, pp.2322–2330.

Zhou, Q., Qu, Z., Guo, S., Luo, B., Guo, J., Xu, Z. and Akerkar, R. (2021) 'On-device learning systems for edge intelligence: a software and hardware synergy perspective', *IEEE Internet of Things Journal*, Vol. 8, No. 15, pp.11916–11934.