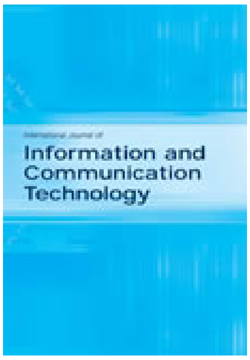


International Journal of Information and Communication Technology

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

Career planning and pathway generation based on multimodal learning analytics

Wenqing Liu

DOI: 10.1504/IJICT.2025.10072409

Article History:

Received:	16 May 2025
Last revised:	29 May 2025
Accepted:	29 May 2025
Published online:	05 August 2025

Career planning and pathway generation based on multimodal learning analytics

Wenqing Liu

Logistics Management Office,
Wuxi Institute of Technology,
Wuxi 214121, China
Email: 15251657837@163.com

Abstract: Concerning the growing complexity of the business environment and career growth, conventional career planning strategies have become challenging to satisfy personal needs. This paper suggests a new career planning and path generating model based on multimodal learning technology, which integrates data from many sources (e.g., educational background, work experience, social networks, emotional data, etc.) so giving individuals more personalised and accurate advice. First, by preprocessing the data with feature fusion, the study creates tailored career path recommendations by building an adaptive model architecture combining deep learning and data mining approaches. In the experimental phase, the results reveal that the model has major benefits in terms of recommendation accuracy, personalisation and flexibility, and offers a fresh technical path for the field of career planning by means of the validation of the parameter tuning of the model and the effect of career path recommendation.

Keywords: multimodal learning; career planning; path generation; personalised recommendation; data fusion.

Reference to this paper should be made as follows: Liu, W. (2025) 'Career planning and pathway generation based on multimodal learning analytics', *Int. J. Information and Communication Technology*, Vol. 26, No. 29, pp.39–58.

Biographical notes: Wenqing Liu received his Master's degree from Zhe Jiang Normal University in June 2011. Currently, he works in Logistics Management Office at Wuxi Institute of Technology. His research interests include educational management, machine learning and artificial general intelligent control.

1 Introduction

1.1 Background and motivation for the study

The fast advancement of technology and the acceleration of globalisation have made the employment environment and career paths increasingly complicated and diverse (Li, 2024). Usually based on static career models and the knowledge of HR specialists, traditional career planning approaches fail to completely address the multifarious needs of individuals and the dynamic changes in the market. Particularly traditional methods ignore more thorough and varied elements including an individual's interests, psychological traits, and social networks in favour of single-dimensional assessments including qualifications, work experience, and other strict criteria (Zeng et al., 2018). This renders conventional career planning insufficient in handling the increasingly complicated career environment, especially in the fast-changing workplace, how to give individuals with up-to-date and customised career routes has become an urgent concern.

Professional planning has progressively taken front stage in recent years, not only in terms of the personal professional path but also directly influences the long-term evolution of career and career satisfaction. Usually carried inside a rigid framework, traditional career planning approaches are largely based on past job experience and market demand and find it challenging to dynamically adjust to the fast-changing needs of the workplace. More and more career planning research have started to investigate how to mix big data, machine learning (ML) and intelligent analytics to increase the accuracy and customisation of career path recommendations as artificial intelligence (AI) technology develops constantly.

In this regard, multimodal learning offers a fresh viewpoint and solution since it is a creative technical method (Baltrušaitis et al., 2018). More and more academics and professionals are now emphasising on how to implement multimodal learning in the domain of career planning. Researchers wish to create a more dynamic and realistic career path generating model by combining information from social networks, psychological data, educational background, work experience and other elements. Combining text data and social network data, for instance, can examine a person's career development potential and his/her influence in the social circle; combining sentiment analysis and interest data can provide insights into an individual's career motivation and job satisfaction, so offering more in line career planning recommendations based on actual needs. Furthermore, more and more businesses and educational institutions are using big data and AI to try to offer customised career development solutions for job seekers and employees given the popularity of intelligent recruitment systems and online learning platforms.

Although the use of multimodal learning in career planning is still in the exploratory stage, its strong data integration and analytical capacity open fresh opportunities for future career planning approaches. Particularly in the information explosion era of fast change of today, conventional static career planning techniques cannot fit the fast-changing needs of the professional market (George, 2023). By means of multimodal learning, it not only enable thorough consideration of the individual's interest, ability, experience, and other elements but also analyse industry trends and technological innovations in real-time, enabling individuals to dynamically change their career paths and fit to new market challenges.

This work attempts to investigate the applicability of multimodal learning in career planning and path generating based on this background. This work will specifically offer tailored and dynamic career path recommendations by building a career planning model based on multimodal data analysis and merging several data sources, including but not limited to educational background, work experience, social networks, emotional data, and market trends. Apart from addressing the lack of personalisation and accuracy in conventional career planning approaches, the approach will also introduce more sophisticated data-driven and intelligent analysis approaches to the field of career planning, so supporting the intelligent and personalised development of career planning.

1.2 Research objectives and questions

This work aims to present a new career planning and route generating model using multimodal learning technology to deliver people more tailored, accurate and flexible career development recommendations. The conventional single-modality-based career planning approaches cannot satisfy the increasingly complicated personal needs and different market trends given the ongoing changes in the workplace environment. Thus, the main goal of this work is to build a multimodal learning framework that integrates several data sources and considers several elements such personal characteristics, career interests, industry development, and market demand, so customising a personalised career path for every individual.

The main effort of this research concentrates on the following elements to reach this target:

1. Fusion and analysis of multimodal data: This work initially addresses the creation of a multimodal learning model capable of efficiently merging data from several modalities. These multidimensional data will enable the model to investigate personally the career growth possibilities of people and offer a more complete analytical viewpoint. The main challenge of this work is how to adequately combine and analyse these heterogeneous data, particularly how to exploit the inherent link between several data modalities.
2. Generation of personalised career paths: This work will investigate ways to create customised career routes based on everyone's unique traits, abilities, interests, experience, and other circumstances using fused multimodal data. By means of multimodal learning, we can deliver people more accurate and focused career planning recommendations. Realising this goal will improve the flexibility and long-term growth of personal careers as well as make career planning more pertinent to the demands of people.
3. Dynamic career planning model: Career path planning has to be somewhat dynamic and flexible given the ongoing changes in personal growth environment and market need. This work will provide a dynamically updated career planning model tracking changes in industry trends and individual career advancement in real-time and automatically updates suggested career routes to guarantee that advised solutions are aligned with market trends, therefore addressing this difficulty.

By means of these objectives, this study will be able to enable the intelligent growth of the career planning area and offer individuals more accurate and tailored solutions for their unique planning needs.

1.3 Research contributions

Design and implementation of a multimodal learning-based analytical model for career planning and path development constitutes the main contribution of this work. The model can customise personalised career trajectories for every person depending on their interests, aptitudes, background, and market need by combining several heterogeneous data sources. Here are the contributions:

1 Designed a multimodal analysis model: This work presents and uses a novel multimodal learning framework able to combine data from several modalities (text, social activities, industry trends, sentiment analysis, etc.). This incorporation of multimodal data significantly enhances the capacity of the model to fully evaluate unique traits. The methodology offers more effective and tailored support for career planning than conventional approaches, therefore more precisely capturing the potential and needs of people.

2 Personalised career path generation: Based on the designed multimodal analysis model, this work develops a personalised career path generating approach. Based on an individual's multidimensional traits (e.g., job experience, interest, personality, etc.), the approach can automatically create the best appropriate career path for him or her depending on real-time market demand. This approach improves the personalisation and accuracy of the recommendation outcomes by transcending the restrictions of single-modal data in conventional career planning.

3 Promoting the intelligent transformation of career planning methods: This work allows the intelligent transformation of career planning strategies by adding multimodal learning technologies. Unlike conventional static planning approaches, the model can dynamically shift career routes depending on different phases of individual development and market changes, therefore offering more flexible and forward-looking planning recommendations.

4 Provides a new methodological framework for the field of career planning: This paper advances the research process of intelligent and customised career planning by offering a creative technical framework based on the multimodal analysis model. Apart from offering fresh concepts for scholarly research, the model offers a technical foundation for the creation of career planning systems in upcoming pragmatic uses.

Finally, by means of a multimodal analysis model, this work transcends the constraints of conventional career planning strategies, enhances the personalisation, accuracy and dynamic adaptability of planning, and offers a fresh path and solution for research and pragmatic application in the field of career planning.

2 Relevant technologies

2.1 Multimodal learning

Learning and reasoning by combining input from several senses is the approach known as multimodal learning. Modality in information refers to several kinds of data involving often diverse properties and expressions: text, image, speech, video, etc.; these forms represent information. Conventional unimodal learning is often treated based on just one data type, therefore restricting understanding of challenging problems. On the other hand, multimodal learning allows information from several modalities to be combined, therefore offering a more complete and accurate answer for addressing useful challenges. By combining features from many modalities, multimodal learning aims to offset the details and diversity that single-modal learning cannot adequately represent (Huang et al., 2021).

Multimodal learning has been extensively used and investigated with the fast expansion of big data technologies and deep learning approaches. Particularly in the domains of computer vision, speech recognition and natural language processing (NLP), multimodal learning presents significant promise. In the task of speech recognition and sentiment analysis, the speaker's emotional tendency can be more precisely analysed by combining the speech and sentiment label data; in the application of image and text combination, the model not only understands the content of the image but also further improves the understanding and prediction ability of the image based on the textual information describing the image.

Usually using two major fusion algorithms, multimodal learning uses feature layer fusion and decision layer fusion.

Feature layer fusion is the process of directly splicing or weighted averaging the feature representations of several modalities therefore producing a unified representation in the feature space of several modalities (Zhang et al., 2020). This method can maximise the information provided by every modality and enable the model to learn the commonalities and variations among several modalities in the same feature space. This method has the benefit in immediately integrating feature-level information at the first stage, thereby supplying more data for next learning.

Conversely, decision level fusion combines the outcomes of every modality following individual separate processing of each modality. Usually, every modality produces an autonomous forecast and finally aggregates several outcomes using weighted average, voting, or another technique. This fusion strategy has the benefit in preserving the independence of every modality while avoiding learning challenges brought on by too dissimilar aspects of several modalities. But it might not be able to completely leverage the possible connections across modalities; hence, feature layer fusion would be more efficient in some applications.

The relevance between several modalities and the requirements of the job typically determines the efficacy of data fusion in multimodal learning applications. While the decision layer fusion method is appropriate for the scenario whereby the modalities are more independent and each modality has independent decision-making ability, the feature layer fusion method is usually suitable for the situation whereby the modalities are closely related, and the feature differences are not significant.

Usually, one needs to create realistic model architecture if one is to properly execute multimodal fusion. Particularly when picture and video input have to be handled, multimodal learning makes extensive use of architectures including deep neural networks (DNNs) and convolutional neural networks (CNNs). Recurrent neural networks (RNN) and their derivatives (e.g., LSTM, GRU) are also rather often employed to collect time series features for text and speech data (Tembhurne and Diwan, 2021). Particularly in text and picture fusion, where the self-attention mechanism of the transformer model can efficiently capture the correlation between several modalities, models based on the Transformer design have made major advancement in multimodal learning in recent years.

Suppose, mathematically, that two modal datasets X_1 and X_2 have respective features $f_1(X_1)$ and $f_2(X_2)$, respectively, and that these two data are fused by a function; then, the fused feature f_{fusion} may be stated by the following equation:

$$f_{fusion} = g\left(f_1\left(X_1\right), f_2\left(X_2\right)\right) \tag{1}$$

where common fusion techniques are weighted averaging and splicing and $g(\cdot)$ is a fusion function. For the splicing technique, for instance, the merged features might be stated as:

$$f_{fusion} = \left[f_1(X_1), f_2(X_2) \right] \tag{2}$$

Moreover, in multimodal learning, cooperative training is a fundamental idea particularly regarding challenging multimodal tasks. By means of combined training, data from several modalities can concurrently affect the learning process, therefore enhancing the task performance (Kaplan et al., 2021). Allow Z to be a joint representation; then, one may define the loss function L as follows:

$$L = \lambda_1 L_1(Z, X_1) + \lambda_2 L_2(Z, X_2) \tag{3}$$

where λ_1 and λ_2 are hyperparameters used to balance the contributions of several modes and L_1 and L_2 respectively indicate the loss functions connected with modes X_1 and X_2. Reducing this loss function helps the model to develop a more efficient joint representation, hence improving the multimodal task performance.

Not only may multimodal learning demonstrate great adaptability in real-time tasks, but it also helps to properly manage the heterogeneity among several modalities. More and more difficult tasks have started to adopt multimodal learning approaches, including autonomous driving, medical image analysis, and human-computer interaction, etc., which are actively investigating how to improve the accuracy and efficiency of tasks using multimodal learning as technology develops, especially regarding constant innovation in the field of deep learning.

2.2 Current status of career planning research

Key component of a person's career development and decision-making process, career planning has attracted a lot of interest recently. The workplace environment is always changing with the process of globalisation and fast technological development; conventional approaches of career planning cannot satisfy the needs of contemporary people and businesses. More and more study has thus started to concentrate on how to employ modern technology, particularly advanced methodologies including big data, AI and ML, to give individuals more personalised, precise and dynamic career planning options.

Data-driven career planning approaches are starting to be applied increasingly as ML develops. Usually depending on a lot of historical data, these techniques forecast individual career development potential and path choice by means of data mining and pattern recognition. By means of deep correlations between an individual's personal information (e.g., interests, abilities, personality, etc.) and the demands of the workplace, ML approaches enable us to provide more tailored career recommendations for every individual.

ML-based career planning approaches most often use supervised learning, unsupervised learning, and reinforcement learning (RL). Learning patterns using tagged training data help one to forecast the kinds of jobs or places for which a person would be suitable by means of supervised learning techniques (Toch et al., 2019). From an individual's past career data, e.g., educational qualifications, work experience, etc. Algorithms including support vector machine (SVM) and random forest (RF) are extensively applied to forecast future career pathways. Conversely, unsupervised learning can find possible career categories or trends from vast volumes of occupational data. Clustering techniques, for instance K-means, DBSCAN, etc., can categorise various occupational categories and hence offer appropriate career paths for job seekers. By means of unsupervised learning, possible trends and patterns about career growth can be identified from the vast volume of job market data to assist individuals in determining more appropriate paths in challenging career decisions.

As a very fresh technical approach, RL has also been progressively included in career planning research recently. Unlike conventional prediction models, RL may give people career planning recommendations by means of constant environmental interaction in continual adjustment and optimisation (Kokkodis and Ipeirotis, 2021). Designing a fair compensation mechanism allows RL models to enable people to dynamically change their career routes in real-time in the constantly changing workplace environment thereby optimising the rewards of long-term career progress.

NLP technology is a major focus of career planning research since, particularly in the application of text classification, sentiment analysis, and information extraction, it has made great development recently. NLP technology may extract useful information from enormous job ads, career descriptions, business cultures, resumes, etc., thereby provide reference for career planning by processing and analysing career-related text data (Roulin and Stronach, 2022). In this arena, NLP technology's implementation depends much on CV analysis. NLP techniques can find an individual's key abilities, experience background, and career interests by extracting and evaluating the language information on job seekers' resumes. Resume parsing methods based on word embedding (Word2Vec, GloVe) and deep learning models (e.g., BERT) can, for instance, extract the individual's skill labels, work experience, educational background, and other information from the resume and subsequently match the individual's appropriate career depending on the job requirements.

Big data technology's prominence has also helped career planning research to start using big data analysis techniques extensively. Processing and analysing big amounts of data from many sources will enable more accurate and real-time career planning models to be created. Career planning has also seen growing application for big data analytics techniques grounded on social network data (Chang, 2018). Deeper knowledge of an individual's career interests, social circles and development potential can be obtained by means of information analysis of their interactions on social media and career networks, thereby enabling better planning of an acceptable career route for them. Social network analysis, for instance, can highlight elements of an individual's influence, leadership, and workplace networking resources, therefore improving career growth prospects.

Research on career planning has made greater use of technical approaches including multimodal learning. By combining text, visuals, social media, and sentimental data, multimodal learning makes career planning more accurate and individualised than more conventional approaches. Using CV, social network, and online activity data, multimodal learning could evaluate an individual's professional potential and development path in a career recommendation system. Combining professional interest test results, social media expression of emotions, and interpersonal interaction style helps the model to better grasp an individual's interest propensity and offer tailored career development recommendations.

The real-time and dynamic character of career planning has grown to be a major concern given the continuous change of market demand and personal condition. More and more career planning systems are starting to incorporate real-time recommendation systems to help to address this issue. The system can update and modify the career planning program depending on the real-time changing profession market need and individual development by merging real-time data flow and personal feedback.

3 Methodology

3.1 Design of a multimodal learning model

Three basic components make up the multimodal learning model that this work proposes. Every module serves to solve problems at various levels and taken collectively they offer tailored career path recommendations.

3.1.1 Data preprocessing module

The module for data preparation seeks to give excellent data input for next learning assignments. To guarantee authenticity and consistency, the data has to be thus cleansed, normalised, and converted using suitable pre-processing methods. This work selected the following methods to handle data preprocessing choreography.

Text data in the data cleaning process might include noise, repeated information and deactivated words, which can influence the learning impact of the model. This work so selected the TF-IDF (textual denoising and feature extraction algorithm). By computing the ratio of the frequency of each word in the text to the frequency of the word in the whole corpus, TF-IDF assesses the relative value of every word, thereby eliminating common terms that less affect the analysis (Qaiser and Ali, 2018). Image data were also cleaned using Gaussian filtering, a technique that smooths the image and reduces noise, hence improving image quality. This work uses the voice activity detection (VAD) algorithm to exclude extraneous muted sections of audio data such that the legitimate speech signal is retained.

Using the KNN filling technique, this work addresses missing value processing. This method guarantees data integrity by means of the identification of K neighbours in the dataset most comparable to the missing values and subsequent inference of the missing values using their values. Multiple imputation is applied for modes with large missing rates to create several complete datasets by repeatedly filling the missing data, therefore lessening the effect of filling errors on the analysis outcomes.

Min-max normalisation and Z-score normalisation are applied in the stage of data normalisation and standardisation in this work (Kappal, 2019). The data is scaled to the interval of [0, 1] using min-max normalisation; therefore, removing the influence of various feature value ranges and guaranteeing that, in the model training, the contributions of all the features are somewhat balanced. There is a normalisation formula as follows:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{4}$$

where x is the original data; x_{norm} is the normalised data; min(x) and max(x) are respectively the minimum and maximum values of the data.

Z-score normalisation was applied in this work for the feature data that follows Gaussian distribution by converting the data into a standard normal distribution with mean 0 and standard deviation 1 with the formula:

$$x_{std} = \frac{x - \mu}{\sigma} \tag{5}$$

where x is the original data; μ is the data mean; σ is the data standard deviation; x_{std} is the normalised data.

Regarding feature extraction, for text, image, and audio data this work selects several techniques. This work employs the Word2Vec technique for text data since it maps words into dense vectors, therefore capturing the semantic links between words. CNN is used in this work for picture data to extract high-level images by means of hierarchical convolution and pooling processes. Appropriate for speech signal analysis, the mel frequency cepstrum coefficient (MFCC) method extracts spectral properties of audio signals for audio data.

This leads the data preparation module to guarantee the consistency of the multimodal data in terms of quality and format, therefore ensuring reliable input data for the next multimodal learning model and assures the accuracy and robustness of the final findings.

3.1.2 Feature fusion module

Effective merging of data features from many modalities to improve the performance of the model depends on the feature fusion module, which is quite important. First, this work implements a decision-level fusion approach. Under this approach, independent models first process and predict the data from every modality; then, the prediction results from every model are merged. With a weighted average, the fusion can dynamically change the weights of several modalities based on their respective contributions (Yang et al., 2020). One can articulate the fusion process by means of the following equation:

$$y_{final} = \sum_{i=1}^{n} w_i y_i \tag{6}$$

where y_{final} is the final prediction result; y_i is the prediction result of the i^{th} modality; w_i is the weight of the modality; n is the number of modalities engaged in the fusion. This approach guarantees that, in the end, the predicted outcomes of every modality are rather integrated.

This work also presents the self-attention method to improve the expressiveness of the fusion process. Learning the interrelationships between modalities, this process automatically assigns weights to several modalities. The self-attention mechanism helps the model to concentrate on the most important characteristics depending on the contextual information of the input data, therefore improving the quality of fusion. The self-attention process has a mathematical form as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{7}$$

where Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the normalisation factor. The self-attention mechanism can dynamically change the relevance of any modal characteristic to improve the model performance on challenging tasks.

In the process of feature fusion, this work additionally incorporates a multilayer perception (MLP) to further increase the fusion effect. By nonlinearly translating the feature vectors from each modality through a DNN, the MLP can capture the intricate correlations between the features of each modality (Sun et al., 2018). The features from every modality are nonlinearly altered through several hidden layers and then combined at the output layer after going via distinct input layers. This method strengthens the representational power of the model and provides for a better grasp of the underlying links between multimodal data.

In addition, this work also tackles the problem of dimensionality reduction of features, especially when the data dimensionality is high, the dimensionality reduction technique can significantly lower the computational complexity. Thus, this work uses principal component analysis (PCA) for dimensionality reduction of the fused features. From high-dimensional data, PCA can extract most representative features, lower data redundancy, and increase computing efficiency. The mathematical form of it is:

$$Z = XW \tag{8}$$

where X is the original feature matrix; W is the transforming matrix; Z is the feature matrix following dimensionality reduction. By means of dimensionality reduction, the model can lower the computational load while preserving the information quantity.

In summary, the feature fusion module in this study accomplishes effective integration of multimodal data by incorporating multiple techniques such as decision layer fusion, self-attention mechanism, MLP network, and PCA dimensionality reduction. These techniques not only increase the expressive capability of feature fusion but also, in handling challenging data, strengthen the model's durability and precision. By means of these creative fusion techniques, this study can supply people with more individualised and exact job paths.

3.1.3 Forecasting module

By feeding the fused features into the model, the aim of the prediction module of multimodal learning is to create tailored career path prediction results. This work proposes a simplified and effective deep learning framework capable of processing multimodal information and generating correct career path suggestions to increase the accuracy and real-time performance of the prediction.

First, for this work DNN is selected as the fundamental prediction model. DNN can automatically learn and discover probable links between features when working with complex data due to its powerful nonlinear modelling capabilities. In this step, the data following feature fusion is sent into many hidden layers of the DNN, each of which is nonlinearly changed by an activation function (e.g., ReLU) to boost the expressive capability of the model. By learning the deep correlations between features layer by layer, the model can offer personalised career path predictions at the output layer.

To enhance the stability and efficiency of the training process, this work introduces the batch normalisation technique in DNN. Batch normalisation can standardise the data distribution of each output layer, making the training of data more stable, expediting the convergence process, and lowering the model's dependence on initial parameters (Garbin et al., 2020). In addition, batch normalisation can help ease the gradient vanishing problem, thus enhancing the learning efficiency of the network.

In the choice of loss function, this work chooses the cross-entropy loss function. The cross-entropy loss function can efficiently assess the difference between the prediction results and the true labels, which is highly useful for classification tasks. In this study, the prediction aim is to identify viable career routes for individuals; hence, the application of cross-entropy loss can ensure that the model is tuned to best meet the goal. The formula for cross-entropy loss is as follows:

$$L = -\sum_{i=1}^{n} y_i \log\left(\hat{y}_i\right) \tag{9}$$

where n as the sample count, $\hat{y}_i$ is the model's prediction probability and y_i is the actual label. Reducing this loss function helps the model to progressively change the parameters to raise the career path prediction accuracy.

This work also uses the dropout strategy during model training and inference to enhance computing efficiency and stability. By randomly deleting a fraction of neurone connections, dropout helps the model to generalise and stops overfitting of the network. This helps the model to keep great prediction stability across noisy and complicated data.

This work also considers the model integration technique to raise the final prediction accuracy of the model. More especially, the RF algorithm was used as a supplemental model. Constructing several decision trees and voting on their results helps RF to efficiently increase model resilience. Combining DNN with RF helps to balance the advantages and disadvantages of several models, thereby enhancing the prediction accuracy.

Combining DNN, batch normalisation, cross-entropy loss function, dropout techniques and model integration approaches allows the prediction module in this work to effectively generate customised career path prediction. These simplified and effective algorithms enable the model to fully use multimodal data to offer more accurate and flexible career planning recommendations and to change with the demands of every person.

Algorithm 1 presents the three primary components' three-fold workflow. Every module uses reasonable actions to guarantee complete processing of the data, efficient fusion of features, and finally correct career path projections.

Algorithm 1 Pseudo-code for multi-modal learning model

Input: Multi-modal raw data ($X_{\_\text{text}}$, $X_{\_\text{image}}$, $X_{\_\text{social}}$), model parameters (W, b), learning rate (η), number of iterations (T)

Output: Optimised model ($W_{\_\text{opt}}$, $b_{\_\text{opt}}$) for career path recommendation

begin
  Initialise model parameters (W, b) randomly;
  Initialise data preprocessing module with appropriate transformation functions;
  Initialise feature fusion module with fusion strategy;
  Initialise prediction model;
  for t = 1 to T **do**
    Collect multi-modal raw data ($X_{\_\text{text}}$, $X_{\_\text{image}}$, $X_{\_\text{social}}$);

```
    Preprocess the raw data:
        X_text′ ← NLP_Processing(X_text); // Tokenisation, Embedding
        X_image′ ← CNN_FeatureExtraction(X_image); // CNN feature extraction
        X_social′ ← SocialData_Embedding(X_social); // Social network embedding
        X′ ← Normalise(X_text′, X_image′, X_social′); // Normalise the features
        X′_cleaned ← HandleMissingData(X′); // Handle missing or incomplete data
    end for
    Apply feature fusion:
        If early_fusion then
            X_fused ← Concatenate(X_text′, X_image′, X_social′); // Concatenate features
            X_fused′ ← FusionLayer(X_fused); // Pass through fully connected layers
        else if late_fusion then
            X_text_pred ← DNN(X_text′); // Predict for text modality
            X_image_pred ← CNN(X_image′); // Predict for image modality
            X_social_pred ← MLP(X_social′); // Predict for social modality
            X_fused ← Average(X_text_pred, X_image_pred, X_social_pred); // Combine predictions
            by averaging
        end if
    end for
    Perform prediction:
        Y_pred ← DNN(X_fused′); // Forward propagation in DNN
        L ← CrossEntropyLoss(Y_pred, Y_true); // Compute loss (cross-entropy)
        ∇W, ∇b ← Backpropagate(L, W, b); // Backpropagation to compute gradients
        W ← W – η * ∇W; // Update weights using gradient descent
        b ← b – η * ∇b; // Update biases using gradient descent
        Evaluate performance using validation data (accuracy, precision, etc.);
        Optionally apply regularisation (e.g., dropout) to avoid overfitting;
    end for
Return optimised model parameters (W_opt, b_opt) for career path recommendation;
```

3.2 Career path generation and personalised recommendation methods

By combining elements from several modalities, the career path generation and personalised recommendation approach under the multimodal learning framework is meant to deliver people accurate and personalised professional development recommendations. Apart from considering the fundamental traits of individuals, the multimodal learning method used in this work combines information from several data sources: textual data, social network data, sentiment analysis data, even external market demand and industry trend data. This combination of multi-dimensional data makes the advice for the career route more precise and tailored, so it is better to address the demands for professional growth of the individual.

Generation of a career path consists in several phases. First, the data pretreatment module cleans and standardises an individual's past data and current skill set as input features. The feature fusion module then generates a single feature representation by fusing data from many sources. These records not only show personal background information but also capture the career interest and development potential of the person in several environments, therefore offering multi-level assistance for later path generation.

Second, DNN under this framework models personalised recommendation methods and together with collaborative filtering-based recommendation algorithms learns similarities from an individual's prior career routes and social network activities to deliver reliable path predictions. Based on input individual traits and real-time updated market demand, the model can dynamically change career route recommendations (Ashrafi et al., 2023). If demand for a given sector rises significantly, for instance, the model may quickly adapt the path recommendation to enable people to identify a more promising development track in the evolving employment environment.

Personalised recommendations improve the flexibility and real-time character of the advice by including dynamic information, including changes in career preferences and updates in industry demand, so augmenting the basis of the recommendation from the static aspects of a person. By examining the personal emotional changes and social network interaction patterns, the approach maximises the suggestion outcomes even more. The model's output on the career route is ultimately a recommendation sequence that combines personal demands and market trends; so, people can adapt their career planning depending on this outcome.

Assuming that x_i is the feature vector of individual i, the formula representation of route generation yields a predicted output of a probability distribution reflecting the recommendation of career paths, which may be stated by the following formula:

$$\hat{y}_i = \sigma\left(W_2 \cdot \sigma\left(W_1 \cdot x_i + b_1\right) + b_2\right) \tag{10}$$

where W_1 and W_2 are the weight matrices of the network; σ is the activation function; b_1 and b_2 are the bias terms. By layer-by- layer mapping of the DNN, this formulation turns the multimodal feature vector xi into a tailored career path recommendation.

This work also presents a dynamic RL-based adjustment method to help to maximise the accuracy of path recommendations. This system lets the model maximise the recommendations in real-time based on changes in personal career development and market demand. By means of interactive learning between the intelligent body and the surroundings, RL constantly changes the method to get the dynamic updating of the career path advice (Da Silva and Costa, 2019). One can write the objective function of RL by means of the following equation:

$$L_{RL} = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r_t\right] \tag{11}$$

where r_t is the reward at moment t; L_{RL} is the total loss function of RL; γ is the discount factor; T is the maximum number of time steps for path recommendation. Optimising this loss function helps the model to modify the suggested courses depending on real-time market and individual information, therefore assuring that every person is always in a good position in the long run evolution of the employment market.

In the end, our work can give individuals tailored and reliable career path suggestions by combining multimodal data fusion, DNN modelling and RL mechanisms. This approach not only enables people to create more acceptable career plans but also increases the success rate and happiness of professional development, hence having great practical value and vast application possibilities.

4 Experimental design and result analysis

4.1 Data collection and pre-processing

The main data source for this study was the O*NET Database, which the US Bureau of Labour Statistics (BLS) provides and is an open database including details on all kind of jobs in the USA. For specifics, see Table 1.

Table 1 Specific information on the dataset

Dataset name	*O*NET database*
Dataset description	Provides detailed occupational information, including job tasks, skills, and requirements
Dataset size	Contains data on approximately 1,000 occupations, with multiple feature dimensions
Data structure	Includes occupation classification, job activities, skills, abilities, knowledge, task descriptions, etc.
Application areas	Career planning, employment recommendations, education and training, industry trend analysis
Update frequency	Data is updated periodically to reflect labour market and occupational demand changes

Covering a vast spectrum of jobs, including occupational tasks, skills need, work settings, and industry trends, the O*NET database is extensively used in career planning, skills analysis, and employment trend predictions. These facts can support tailored job path recommendations strongly.

Following data collecting, the raw data of the ONET database must undergo several pre-processing procedures to enable better fit to the requirements of multimodal learning systems. The processing of rich textual data and structured numerical data in the ONET database has to consider the features of several kinds of data. First, NLP methods are applied for textual data processing including procedures like word splitting, deactivation, stemming extraction, etc., so enabling the textual information to be merged with the numerical data in a suitable way. Standardisation and normalisation are then used for numerical data to ensure that all feature data have a consistent scale, therefore preventing the situation whereby some features dominate the overall learning process in model development (Park et al., 2019). Furthermore, needing supplementation and correction are missing numbers and outliers found in the O*NET database. Mean padding or interpolation allows one to treat missing values so guaranteeing the completeness and consistency of the data.

Feature engineering is quite vital in still another crucial stage of data preparation. The most discriminative and representative elements from the O*NET database can be obtained by use of feature selection and dimensionality reduction approaches. In multimodal learning systems, this procedure not only increases data availability but also lowers duplicate information and strengthens model training. Furthermore, the data from every modality must be aligned and standardised if successful fusion of data from several modalities is to be attained and they may be easily combined. These preprocessing activities enable the final model to produce more accurate and customised career planning guidance based on multi-dimensional individual characteristics and industry trends, therefore providing strong data support for the subsequent career path generating and tailored recommendation.

4.2 Experimental setup

This work planned a thorough experimental setup comprising model training, evaluation, hyperparameter setting and experimental environment to confirm the efficiency of multimodal learning in career path generating and personalised recommendation.

Rich data regarding career information, including education requirements, job tasks, skill needs, salary levels, career prospects and other variables, was included into this analysis using the O*NET database. The dataset was split in a training set, a validation set, and a test set of 70%, 15%, and 15%, respectively, so guaranteeing the dependability of the experimental outcomes. Model training uses the training set; hyper-parameter tuning uses the validation set; and the ultimate performance of the model is assessed from the test set during the experiment. Data division made use of random sampling guarantees consistent distribution of the dataset.

Using an Adam optimiser with a learning rate set to 0.001 and a batch size of 64, the experiments were optimised throughout the model training process. Using a dropout rate of thirty%, the dropout approach helped to prevent overfitting. Furthermore, employed as the optimisation goal and an early halting mechanism to prevent model overfitting was a cross-entropy loss function. Based on performance on the validation set, 50 cycles of training were executed and aborted overall. Accuracy, recall, and F1 values are essentially utilised as classification evaluation metrics to gauge the recommendation quality of the model during the review process.

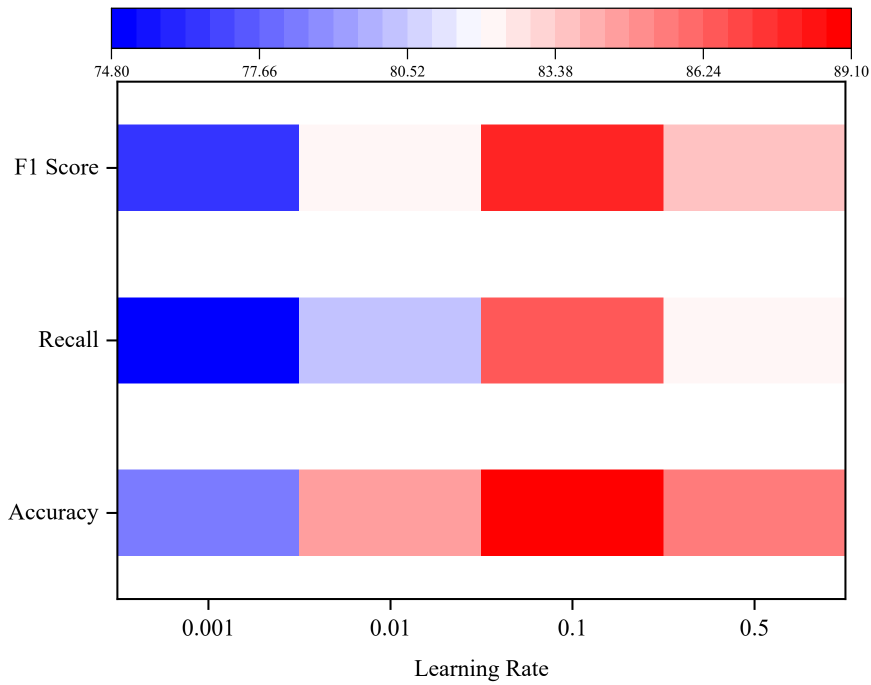
Every experiment is on NVIDIA A100 GPUs under the PyTorch deep learning model implementation paradigm. Every experiment was performed five times separately, five times independently to guarantee the dependability of the findings and hence minimise the impact of random elements.

4.3 Experimental results and analyses

This work carried out two experiments. Experiment 1 mostly aimed to enhance the performance of the multimodal learning model by turning into the important hyperparameter, i.e., the learning rate. Specifically, the aim is to experimentally verify the effects of various learning rate settings on the model in terms of metrics such accuracy, recall and F1 score, to ascertain the most suitable learning rate value to improve the prediction ability of the model in the career path recommendation task. By means of this approach, stronger and more efficient hyperparameter settings may be supplied for next model training to guarantee the optimal performance of the multimodal learning model in useful applications.

The sole hyperparameter that changes during the experimental procedure is the learning rate; its value is adjusted to several distinct values at numerous experimental phases. Other hyperparameters, such batch size and regularisation factors, were maintained fixed. With a batch size of 32 and a regularisation factor set to 0.01, one avoided overfitting the model. Several tests reveal how learning rate modification affects model performance. Figure 1 displays the experimental outcomes.

Figure 1 Results of parameter tuning experiments for multimodal learning models (see online version for colours)



The experimental findings show that with a learning rate of 0.1 the model achieves best in terms of accuracy, recall and F1 score. More especially, the F1 score is 87.8%, the recall rate is 86.7%, and the accuracy rate reaches 89.1%. This outcome shows that the model can better balance the bias and variance under a learning rate of 0.1, so producing more consistent and accurate prediction outcomes. Particularly in terms of recall and F1 score, a learning rate of 0.1 can reasonably raise the general performance of the model when compared with other learning rate values.

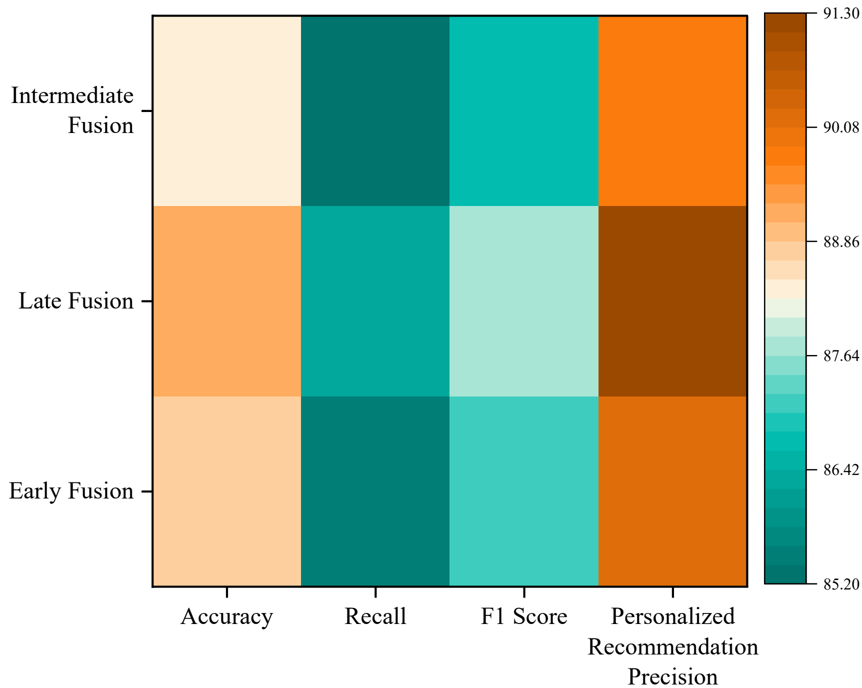
Furthermore, whilst the model performs better, a learning rate of 0.01 causes a decline in performance when compared to a learning rate of 0.1. Respectively, the accuracy, recall, and F1 scores were 84.5%, 80.2% and 82.3%. This implies that a reduced learning rate could lead to slower training and the model may fail to sufficiently change the parameters during the optimisation process, therefore producing some performance loss. Although the model's quicker training speed results in a learning rate of 0.5, the too high learning rate caused instability throughout the training process, thereby lowering the accuracy, recall, and F1 score relative to the learning rate of 0.1. The experimental data indicates generally that in this experiment a learning rate of 0.1 is the best choice.

By choosing several feature fusion techniques and evaluating their effectiveness in the career path generating activity, experiment 2 seeks the best suitable fusion strategy.

The model is trained and assessed for three typical feature fusion techniques: early fusion, late fusion, and intermediate fusion in Experiment 2 using the optimal learning rate (0.1) from experiment 1 retained. Every feature fusion technique is trained independently under the same model architecture; so, at last the effect of feature fusion strategies on multimodal learning models in personalised career path development is investigated by means of method performance comparison.

Figure 2 presents the results of experiment 2.

Figure 2 Results of different feature fusion methods on model performance (see online version for colours)



The late fusion technique shows best on all criteria based on the experimental results. Particularly, the F1 score is 87.7%, the recall rate is 86.4%, the accuracy rate is 89.1%, and the personalised suggestion precision is 91.3%. Although early fusion and mid fusion have quite poor performance, particularly in customised suggestion precision, late fusion is far better than other techniques. This outcome implies that late fusion can better integrate modal information and lower the information interference, so boosting the performance of the model in personalised recommendation and career path generating.

Experiment 2 results confirm the relevance of feature fusion techniques in multimodal learning models, particularly in challenging tasks where selecting a suitable fusion strategy can greatly enhance model performance. Strong adaptability and performance in career path suggestion are shown by the late fusion technique with high information independence and flexibility.

By means of two experiments, the findings of parameter tuning in experiment 1 indicate that the model works best in the indicators of accuracy, recall, and F1 score when the learning rate is 0.1, so offering appropriate baseline parameters for the next studies. Furthermore, by means of a comparison of several feature fusion techniques, experiment 2 revealed that the late fusion method performs well in all the assessment indices, particularly in the personal tailored suggestion precision, which has attained notable improvement. Together, the two studies confirm the effect of learning rate and feature fusion on the performance of the multimodal learning model, thereby offering an efficient optimisation approach for development of career paths and tailored recommendations.

5 Conclusions

This work presents a multimodal learning-based career planning and route building approach to provide customised and accurate career development advice employing many data sources. Breaking down the limitations of conventional career planning strategies, this paper integrates data from educational background, work experience, social behaviour, sentiment analysis, and industry development using a multimodal learning model. Experimental validation reveals that the approach enhances in practice job route suggestion accuracy and customisation. The learning rate has a major effect on model performance in the parameter tuning studies; hence, the results reveal that the model performs best when the learning rate is 0.1 and may provide tailored career advice.

Although it has limits, the multimodal learning approach proposed in this work offers certain career path suggestion outcomes. Although O*NET offers a lot of career information, it does not cover all sectors. Second, the computational complexity of the model requiring a long training period and significant resource utilisation calls for careful consideration of its real-time and scalability. Furthermore, limited in several domains, particularly in the establishment of cross-industry career paths, is the model under investigation. The low interpretability of the model raises questions regarding how to make its recommendation outputs clearer so people may understand its reasoning process.

Future studies can grow and improve in several spheres:

1 Interpretability and transparency of the model: Practical applications depend on model interpretability for user approval and confidence. By means of visualisation tools or rule-based models, we can investigate future ways to increase the transparency of multimodal learning models thereby enabling users to grasp the suggested career choices and their underlying reasoning. This could help to raise people's career happiness and capacity for making decisions.

2 Introducing the impact of changes in the external environment: The present paradigm mostly relies on past data for analysis, therefore neglecting the impact of the surroundings on career pathways. Future studies should consider modifying the external environment to guarantee that recommendations on career planning are more flexible and realistic (Tomlinson et al., 2018).

3 Enhancement of intelligent and personalised recommendation systems: Deep RL and other AI technologies' ongoing development will help career planning systems' intelligence to be even better (Shi et al., 2022). Simulating the interaction between an

individual and the surroundings allows the model to learn and adapt constantly to give consumers more flexible, varied, and real-time career path recommendations.

By means of these extensions, future research will be able to adapt to the fast-changing workplace environment, enhance the intelligence of career planning and route generating systems, and give individuals more accurate and tailored career development help.

Declarations

All authors declare that they have no conflicts of interest.

References

Ashrafi, S., Majidi, B., Akhtarkavan, E. and Hajiagha, S.H.R. (2023) 'Efficient resume-based re-education for career recommendation in rapidly evolving job markets', *IEEE Access*, Vol. 11, pp.124350–124367.

Baltrušaitis, T., Ahuja, C. and Morency, L-P. (2018) 'Multimodal machine learning: a survey and taxonomy', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 2, pp.423–443.

Chang, V. (2018) 'A proposed social network analysis platform for big data analytics', *Technological Forecasting and Social Change*, Vol. 130, pp.57–68.

Da Silva, F.L. and Costa, A.H.R. (2019) 'A survey on transfer learning for multi-agent reinforcement learning systems', *Journal of Artificial Intelligence Research*, Vol. 64, pp.645–703.

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.

George, A.S. (2023) 'The ephemeral career: how technological change necessitates flexible employment', *Partners Universal International Innovation Journal*, Vol. 1, No. 5, pp.47–62.

Huang, Y., Du, C., Xue, Z., Chen, X., Zhao, H. and Huang, L. (2021) 'What makes multi-modal learning better than single (provably)', *Advances in Neural Information Processing Systems*, Vol. 34, pp.10944–10956.

Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D. and Hancock, P.A. (2021) 'The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis', *Human Factors*, Vol. 63, No. 4, pp.706–726.

Kappal, S. (2019) 'Data normalization using median median absolute deviation MMAD based Z-score for robust predictions vs. min-max normalization', *London Journal of Research in Science: Natural and Formal*, Vol. 19, No. 4, pp.39–44.

Kokkodis, M. and Ipeirotis, P.G. (2021) 'Demand-aware career path recommendations: a reinforcement learning approach', *Management Science*, Vol. 67, No. 7, pp.4362–4383.

Li, L. (2024) 'Reskilling and upskilling the future-ready workforce for industry 4.0 and beyond', *Information Systems Frontiers*, Vol. 26, No. 5, pp.1697–1712.

Park, J.E., Park, S.Y., Kim, H.J. and Kim, H.S. (2019) 'Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives', *Korean Journal of Radiology*, Vol. 20, No. 7, pp.1124–1137.

Qaiser, S. and Ali, R. (2018) 'Text mining: use of TF-IDF to examine the relevance of words to documents', *International Journal of Computer Applications*, Vol. 181, No. 1, pp.25–29.

Roulin, N. and Stronach, R. (2022) 'LinkedIn-based assessments of applicant personality, cognitive ability, and likelihood of organizational citizenship behaviors: comparing self-, other-, and language-based automated ratings', *International Journal of Selection and Assessment*, Vol. 30, No. 4, pp.503–525.

Shi, Z., Wu, Z., Zhang, Z., Chen, Y. and Liu, X. (2022) 'Learning path planning algorithm based on career goals and artificial intelligence', *International Journal of Emerging Technologies in Learning (IJET)*, Vol. 17, No. 10, pp.256–272.

Sun, D., Wang, M. and Li, A. (2018) 'A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data', *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 16, No. 3, pp.841–850.

Tembhurne, J.V. and Diwan, T. (2021) 'Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks', *Multimedia Tools and Applications*, Vol. 80, No. 5, pp.6871–6910.

Toch, E., Lerner, B., Ben-Zion, E. and Ben-Gal, I. (2019) 'Analyzing large-scale human mobility data: a survey of machine learning methods and applications', *Knowledge and Information Systems*, Vol. 58, pp.501–523.

Tomlinson, J., Baird, M., Berg, P. and Cooper, R. (2018) 'Flexible careers across the life course: advancing theory, research and practice', *Human Relations*, Vol. 71, No. 1, pp.4–22.

Yang, Y., Cao, S., Huang, S. and Wan, W. (2020) 'Multimodal medical image fusion based on weighted local energy matching measurement and improved spatial frequency', *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, pp.1–16.

Zeng, W., Huang, F., Yu, L. and Chen, S. (2018) 'Towards a learning-oriented assessment to improve students' learning – a critical review of literature', *Educational Assessment, Evaluation and Accountability*, Vol. 30, pp.211–250.

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.