WJCMS

Wasit Journal of Computer and Mathematics Science

Journal Homepage: https://wjcm.uowasit.edu.iq/index.php/WJCM
e-ISSN: 2788-5879 p-ISSN: 2788-5887

Educational Data Mining for Predicting Academic Attrition in Iraqi Universities: A Hybrid KDD-Based Approach using Weighted Euclidean Distance and Neural Architectures

Mayada Jabbar Kelain [1,*], Thura J. Mohammed[2]

[1] Department of Computer Sciences, Faculty of Tourism Sciences/Mustansiriy University, Baghdad, Iraq.

[2] Department of Computer Sciences, School of Computer Sciences, Universiti Sains Malaysia, George Town, Pulau Pinang 11800, Malaysia.

*Corresponding Author:Mayada Jabbar Kelain

DOI: https:// doi.org/10.31185/wjcms.492
Received 18 February 2026; Accepted 25 March 2026; Available online 31 March 2026

ABSTRACT: Academic attrition in the Iraqi higher educational system is a significant problem with socio-economic consequences. The purpose of this study is to address the problem through the development of a high-fidelity predictive model using existing educational data sets. The approach is to develop a hybrid model based on the Knowledge Discovery in Databases (KDD) process and Artificial Neural Networks (ANN), with a newly proposed Weighted Euclidean Distance metric. The approach was validated using a dataset of 5,000 Iraqi university students and texture analysis, which was inspired by image mining techniques. The analysis was performed to find the optimal weight for socioeconomic and academic features. The KDD + ANN model demonstrated considerable ability to grasp the intricacies and nonlinearity of behavioral patterns of students at the university level. Therefore, it achieved remarkable accuracy of 94.21%, precision of 93.15%, and F1-score of 93.32%. Moreover, feature importance analysis showed that Major Subject GPA had the highest weight of 0.48, followed by attendance in the first six weeks with a weight of 0.32. Moreover, it is important to point out the accuracy with which the weighted metric has been identified as an essential latent risk factor in the Iraqi scenario as well. Conclusion: The weighted proximity metric in neural architecture is an effective approach in dealing with the complexities of educational data sets, thereby reducing the chances of incorrect classification. Recommendations: This strategic early warning system needs to be incorporated in the administration of universities in Iraq to bring about a shift in the approach from generalized to more targeted advisories.

Keywords: Academic Attrition Prediction, KDD, Weighted Euclidean Distance, ANN.

1. INTRODUCTION

One of the most critical challenges facing institutions of higher education, both in the developed and in the developing world, is academic attrition or student drop out, which has serious academic, social and economic implications [1]. In the case of developing countries, especially Iraq, the situation is more challenging since it reflects the underutilization of national investment in human capital as well as the diminished future employment opportunities of students [1]. In general, in the field of higher education, universities have depended on post-hoc analyses of attrition statistics. In most cases, these analyses do not provide the opportunity for effective instructional intervention as the analyses do not identify students who are potentially at-risk before the students drop out. Therefore, this research is motivated by the need to move from reactive types of analyses to more proactive types of systems that provide early-warning predictive analyses to identify students before they officially leave the university [2], [3].

*Corresponding author: Mayadajabbar@uomustansiriyah.edu.iq

The rapid advancement of information technology within the last decade has transformed higher education resulting in the collection of unprecedented numbers of data pertaining to students. Educational Data Mining (EDM) and Knowledge Discovery in Databases (KDD) have both become the most important methods for examining educational data and discovering behavioral and learning patterns of students [4]. Most of the recent literature have applied machine learning methods to predict student attrition; however, most of the models are context neutral and use conventional indicators of academic performance [5].

The above illustrates the generic models existing in literature and that the absence of any models contextualizing the Iraqi educational system is what most significantly differentiates this work from others. Most models focus on academic data and neglect to include the localized socio-economic, geospatial, and temporal variables that exist in the Iraqi educational system. To realize this, the present study recommends designing a localized, hybrid mehtod, in which the multidimensional academic pathway of a student is treated as an "image" or "texture." To accomplish this goal, the model uses an image mining texture matrix along with an Artificial Neural Network (ANN) in which a Weighted Euclidean Distance is used for the first time. In this model, customized mathematical weights are used for certain variables in the context, such as the Distance to the campus and the attendance in the first semester, which are very important in the Iraqi context but are usually ignored in most international models. The present study aims to achieve this by advancing the objectives and overcoming the limitations of the existing systems in the following ways:

1) Development of a Hybrid KDD ANN Framework: This study proposes and implements a predictive system that utilizes ANN in combination with the Weighted Euclidean Distance metric, specifically developed to fit the unique constraints faced by the Iraqi higher education system.
2) Application of Texture-Based Feature Extraction: This study proposes a unique methodology in utilizing a texture-based representation of educational data, so that ANN can learn to identify complex, non-linear patterns linked to dropout risk
3) Identification of Context-Specific Risk Factors: The model is able to identify the risk factors successfully, proving, through analysis, that Major Subject GPA ($w_i = 0.48$) and attendance in the first six weeks ($w_i = 0.32$) are the most critical factors in dropout risk for this demographic group.

The proposed system also demonstrated superior predictive performance, which was significantly higher than that of the baseline model and achieved state-of-the-art performance metrics. The proposed system achieved an Accuracy of 94.21%, a precision of 93.15%, and an F1-Score of 93.32%.

2. LITERATURE REVIEW AND RELATED WORKS

The last few years have seen major advancements in using EDM and KDD to model student attrition. Earlier models used demographic and academic data on students to develop time and event driven models in a regression framework to model student engagement [5]. These methods were applicable to the simpler educational data sets, but as data sets became more complex, researchers moved toward more sophisticated machine learning and deep learning methods to recognize non-linear behaviors.

The recent body of work (2023 - 2025) focuses on the use of hybrid models to improve accuracy [3]. Investigated the tradeoff between accuracy and explainability in ensemble machine learning models for academic dropout prediction. While good accuracy was attained, the model could not address context specific, latent variables. In a similar way, [1] tapped into EDM techniques to model dropouts of master's students, although, in spite of noting the relevance of academic background, socioeconomic factors were not considered in the model.

Additionally, a recent systematic review was conducted that illustrated the growing dependency on clustering methods for the recognition of high-risk student profiles in the field of EDM [6]. Mentioned that although there have been advancements, a multitude of predictive models have been designed for general, Western educational systems and frequently do not mathematically consider unobvious, localized factors, such as commuting distance, or particular types of financial assistance.

To bridge these gaps in the literature, the present study proposes a localized methodology where student attributes are considered a "texture matrix" and are combined with a bespoke ANN. The distinguishing feature of our approach is the use of the Weighted Euclidean Distance as a means of giving mathematical focus to contextually pertinent factors, in this case, the Iraqi educational context. Table 1 shows the comparison between our study and others.

Table 1: Comparison between the proposed framework and recent related works

Reference	Year	Applied Methodology / Algorithm	Accuracy	Precision	F1-Score	Key Limitations Addressed by Our Model
Fei & Yeung	2015	Temporal Models (RNN, LSTM)	~86.0%	-	-	Lacked contextual socio-economic features.
Alhamad & Singh	2024	Standard EDM (Decision Trees, SVM)	89.40%	88.10%	88.50%	Ignored geospatial variables (e.g., Distance to campus).
Zanellati et al.	2024	Ensemble Machine Learning	91.20%	90.00%	90.50%	Used generic unweighted metrics for all variables.
Lu et al.	2025	Clustering-based Prediction Review	N/A	N/A	N/A	Highlighted the need for localized hybrid frameworks.
Proposed Work	**Current**	**Hybrid KDD + ANN + Weighted Euclidean**	**94.21%**	**93.15%**	**93.32%**	**Successfully quantified latent geospatial and local factors.**

(Note for formatting: Please ensure Table 1 is centered and formatted according to the journal's guidelines).

3. METHODOLOGY: THE PROPOSED KDD FRAMEWORK FOR STUDENT ATTRITION

The present study follows a very rigorous knowledge discovery in databases approach as suggested by Han & Kamber [7], closely following the image mining framework as suggested by Chouhan & Tiwari [8]. The methodological approach used in the present study consists of a unique set of five stages.

3.1. Dataset Description and Feature Selection

The empirical verification of the proposed methodology is carried out using a real-world dataset containing the academic and demographic information of 5,000 undergraduate students. The dataset is directly acquired by the authors from a highly established public university located in Iraq and contains the enrollment records of the students over four consecutive years. Due to the non-disclosure agreements and data privacy policies based on ethical considerations, the authors are not allowed to disclose the dataset as it is highly personal and not publicly available.
The structural features of the dataset are as follows:
The dataset represents the multidimensionality of student profiles in terms of three feature domains:
- Academic Metrics: high school graduation score, first-year major-specific GPA, mid-term examination results, and attendance in the initial six weeks of school.
- Socio-demographics: gender, age at admission, and social class of the students.
- Geo-economic Variables: Distance from home location to university, financial aid status, and employment status of the students.

The data points identified as relevant were in accordance with the literature and the particularities of the Iraqi higher education context. The dataset was subjected to thorough preprocessing, which involved normalization and handling of missing values in order to maintain the integrity of the predictive model before training.

3.2. DATA PREPROCESSING (CLEANING AND NOISE REMOVAL)

Another important factor in the prediction modeling process is the quality of the data. In this stage of the process, the inherent noise in the Iraqi institutional data set was handled through the use of the mean imputation method for handling missing academic scores. Furthermore, min-max normalization was carried out on all the features, which were GPA and Attendance, and were normalized between 0 and 1.

3.3. TRANSFORMATION AND MATHEMATICAL MODELING

This study centers on the essential component of the innovative phase, where the Weighted Euclidean Distance method is used. Unlike the conventional data mining approach, where all the features are considered to be of equal importance, the proposed model uses different weights for the features based on the statistical correlation between the features and the dropout history.
The Distance Formula:
To calculate the proximity of a current student X to a known "at-risk pattern" Y, we utilize the following weighted metric:
The feature weighting and similar measurement between student profiles are computed using the Weighted Euclidean Distance, as defined in Equation (1):

$$d(x,y) = \sqrt{\left\{\backslash sum^{2}_{\{i=1\}^{\{n\}w}_{i}(x_i - y_i)}\right\}} \qquad (1)$$

Where:
- d(x,y): Distance representing the risk level (smaller distance = higher risk).
- w_i: The assigned weight for feature i (e.g., w for GPA is higher than w for sex).
- x_i, y_i: Normalized values of the features for the current student and reference pattern, respectively.

3.4. DATA MINING (THE ALGORITHMIC IMPLEMENTATION)

Consistent with the techniques reviewed in the attached research (Neural Networks, Clustering, and Association), we implemented a hybrid algorithmic approach:
- ANN: A multilayer perceptron (MLP) was designed, and a backpropagation mechanism was incorporated into the ANN. The ANN was used to recognize non-linear patterns associated with the behavior of the students, similar to the pattern recognition process used in digital images.
- K-Means Clustering Algorithm: The student population was segmented into distinct clusters, i.e., Low, Medium, and High Risk, using Euclidean distance metrics.
- Decision Trees Algorithm (C4.5): IF-THEN rules are generated to explain the logic used by the Iraqi administrators to classify a student as at risk of dropping out, in order to improve the interpretability of the classification by the Iraqi administrators.

3.5. EVALUATION AND INTERPRETATION

The final step involved the validation of the model using the Confusion Matrix, and the measurement of the metrics of accuracy, precision, recall, and the F1-score.
This step of validation not only aims at the statistical validation of the patterns generated, but also at the applicability and usefulness of the patterns generated for the Iraqi academic counseling units.

3.6. ALGORITHM CONFIGURATION AND TECHNICAL PARAMETERS

In order to guarantee the reproducibility of the predictive model, while at the same time seeking to attain the highest possible convergence rate, an ANN, using an MLP topology, was implemented, owing to the intricacy involved in the behavior of the Iraqi students.
- Network Topology: The architecture of the network has 12 input neurons representing the chosen academic and socioeconomic factors, two hidden layers, and finally, the output layer. The Hidden Layer 1 has 24 neurons, and the Hidden Layer 2 has 12 neurons. The architecture is tapered, meaning there is a reduction in the number of neurons as we go from the first hidden layer to the second hidden layer, enabling the effective abstraction of features related to student risk.
- Activation Functions: In the hidden layers, the activation function used is the Rectified Linear Unit (ReLU) activation function, which helps in avoiding the vanishing gradient problem. In the output layer, the Sigmoid activation function is used to produce a probability value between 0 and 1 for the attrition prediction. For the final risk classification, the ANN utilizes the Sigmoid activation function to map the output probabilities, as expressed in Equation (2).

$$f(z) = 1 \cdot \left(1 + e^{\{-z\}}\right) \qquad (2)$$

- Optimization and Learning: An Adam optimizer with a learning rate of 0.001 is used. The selected loss function is the Binary Cross-Entropy function, which is best suited for classification-based EDM. This allows the model to differentiate between Noise and Patterns.

3.7. ALGORITHMIC PROCEDURE & PSEUDOCODE

To further clarify the operational mechanism of the proposed model, the following pseudocode is provided in order to illustrate the integration of KDD with the Weighted Euclidean Distance metric, along with the training of the neural network.

Algorithm 1: Predictive KDD Framework for Student Attrition

As suggested in Table 1, the hybrid model ANN+KDD has been found to be significantly better than traditional classifiers, and the precision-recall ratio ranging from 93.15% to 92.50% clearly shows the reliability of the model in reducing false alarms and dropout rates for students (see Algorithm 1).

Algorithm 1: Predictive KDD Framework for Student Attrition.

Line	**Procedural Logic & Implementation (Python Style)**
1	.Phase 1: Data Integration & Cleaning
2	Raw_Data = Load_Iraqi_Uni_Dataset(5000)
3	Cleaned_Data = Preprocess(Raw_Data) . Imputing missing values
4	.Phase 2 :3 & Weighted Feature Engineering & Distance Logic
5	Weights = {'GPA_Major': 0.48, 'Attendance': 0.32, 'Dist': 0.12, 'Edu': 0.08}
6	for students in Cleaned_Data:
7	Dist_Score = sqrt(sum(Weights[i] * (student[i] - Risk_Pattern[i])**2))
8	student.add_feature('Weighted_Proximity', Dist_Score)
9	.Phase 4: ANN Training (Backpropagation)
10	Model = NeuralNetwork(Layers=[12, 24, 12, 1], Activation='ReLU')
11	Model.Train(Cleaned_Data, Optimizer='Adam', Loss='BinaryCrossEntropy')
Line	**Procedural Logic & Implementation (Python Style)**
1	.Phase 1: Data Integration & Cleaning
2	Raw_Data = Load_Iraqi_Uni_Dataset(5000)
3	Cleaned_Data = Preprocess(Raw_Data) . Imputing missing values
4	.Phase 2 :3 & Weighted Feature Engineering & Distance Logic
5	Weights = {'GPA_Major': 0.48, 'Attendance': 0.32, 'Dist': 0.12, 'Edu': 0.08}
6	for students in Cleaned_Data:
7	Dist_Score = sqrt(sum(Weights[i] * (student[i] - Risk_Pattern[i])**2))
8	student.add_feature('Weighted_Proximity', Dist_Score)

Explanation of the Code: The algorithm proposed is a major departure from traditional EDM methods, especially with regard to the integration of a multiphase optimization framework. Data Integrity (Phase 1): Data integrity is said to be compromised due to technical instability in Iraq. The function of preprocess is not just filtering data but also involves imputing means, thus ensuring data integrity even without attendance logs.

Weighted Advantage (Phases 2 and 3): The core of the analysis is Equation 12. Unlike the conventional Euclidean approach, the weighting scheme was incorporated so that the principal subject GPA, w = 0.48, dominates the proximity score. This is comparable to the texture feature extraction in image mining, where particular pixel values related to the behavior of the students are given higher weights in the calculation of the object of interest, i.e., the risk of dropping out.

Architectural Depth (Phase 4): The model's ability to identify non-linear relationships between weighted Distance and socio-economic variables is demonstrated by the four-layer structure of the proposed ANN architecture. The use of the rectified linear activation function helps to ensure computational efficiency and rapid convergence when training the dataset with 5,000 records.

4. EXPERIMENTAL RESULTS AND STATISTICAL ANALYSIS

In this section, the empirical results obtained by employing the KDD process for the Iraqi university dataset are presented. The evaluation of the performance of the proposed hybrid model is carried out by comparing its performance with other machine learning algorithms.

4.1. COMPARATIVE PERFORMANCE METRICS

To assess the effectiveness of the proposed knowledge discovery and data mining KDD-based methodology, a set of performance metrics has been utilized, including Accuracy, Precision, Recall, and F1 Score. As shown in Table 2, there is a significant advantage of the Neural Network approach when integrated with Weighted Euclidean Distance.

Table 2: Comparative Performance of Mining Algorithms.

	Rank	Algorithm	F1-Score (%)	Precision (%)	(%) Accuracy
	1	Proposed Hybrid Model (ANN + KDD)	93.32	93.15	94.21
	2	Support Vector Machine (SVM)	88.73	89.42	91.84
	3	Decision Tree (C4.5)	86.76	87.02	88.50
	4	Naive Bayes	79.82	80.44	82.11

Table 2 Commentary: The results show the effectiveness of the ANN + KDD framework. An accuracy of 94.21% proves the effectiveness of the framework in incorporating non-linear behavioral textures using weighted distances as opposed to linear classifiers like SVM.4.2. Feature Importance Analysis.
Through the computation of Gini importance during decision tree induction, the most dominant factors that affect the results within the Iraqi educational system are identified. Contrary to the results that have been observed within the Western world, where financial debt has been identified as the most dominant factor, the results indicate that the GPA of major subjects and early engagement logs (within the first six weeks) are the most dominant factors, see Table 3.

Table 3: Weighted Influence of Academic and Social Features.

Weight (wi)	Feature Category & Specific Attribute Name
0.48	Academic :Major Subjects Cumulative GPA
0.32	Engagement :First 6-Weeks Attendance Rate
0.12	Socio-Economic :Residential Distance to Campus
0.08	Demographic :Parental Educational Background
Weight (wi)	Feature Category & Specific Attribute Name

Table 3 Commentary: The weighting scheme shows that academic performance is the primary factor, and the weight of 0.12 allocated to "Residential Distance" is such that it takes into account the contextual factors relevant to Iraqi universities, where logistics also play an important role in student retention.

4.4. CASE STUDY: FACULTY OF ENGINEERING

To exemplify the applicability of the model, a case study was conducted at the Faculty of Engineering in one of the most prestigious Iraqi universities. The faculty had a highly varied demographic profile and was under significant academic pressure, see Table 4.

Table 4: Distribution of Engineering Students across Risk Clusters | Risk.

Cluster ID	**Mid-term GPA**	**(%) Attendance**	**Count**	**Risk Category**	**Recommended Action**
1	4.0 / 1.8	%42	112	Critical	Urgent Intervention
2	4.0 / 2.4	%76	345	Vulnerable	Academic Counseling
3	4.0 / 3.5	%92	1043	Stable	Standard Monitoring

The practicability of the KDD Framework is illustrated through the case study as presented in Table 4. The categorization of the students into three unique groups allows the administration to move from generalized academic advisories into specific intervention strategies. The 112 students classified as "Critical" (Cluster ID 1) can be targeted through the weighted factors.

5. DISCUSSION: INTERPRETATION OF FINDINGS

The empirical results support the claim that the "Behavioral Texture" of student dropouts in Iraq is identifiable. The precision achieved by the ANN is 94.21%, thereby validating the claim that the non-linear relationships between the socioeconomic factors of the students and their academic performance are best modeled by deep learning algorithms.

One of the most prominent results in the Iraqi case study is the Distance-Performance Correlation, in which the probability of dropping out is greater in students who live 30 or more kilometers away from the campus, a phenomenon that our Weighted Euclidean Distance metric is able to accurately identify. This suggests that the geographical constraints are a latent pattern in the Iraqi data. Moreover, from the contrast analysis, based on the Image Mining GLCM methodology, it is evident that students with high attendance and low academic achievement tend to be misclassified by the traditional systems. On the contrary, in the proposed model, students with low academic achievement, despite high attendance, are correctly classified as in need of Special Pedagogical Support, rather than being misclassified as being in Risk of Dropout.

5.1. COMPARATIVE DISCUSSION WITH PREVIOUS LITERATURE

The performance of our proposed model (94.21% accuracy) presents a significant advancement compared to the existing literature reviewed in Section 2.

1) Comparative Analysis with Global Models: Romero and Ventura [9], and Baker [10] stated that the implementation of the EDM technique resulted in the achievement of an accuracy ranging from 85% to 88% by the models. However, the current model achieves better performance, enhancing the overall accuracy by 6% compared to the previous models. The improvement in the overall performance of the current model is achieved through the use of Weighted Euclidean Distance, as proposed by Chouhan and Tiwari [7]. The use of the Distance reduces the percentage of false positives, a common problem in general models.

2) Comparison with Regional Studies: In Al-Radaideh et al [11], the authors used Decision Trees in Jordan, obtaining an accuracy rate of 82%. The proposed methodology is original because it is based on a hybrid approach, using ANN and KDD, thus being able to consider non-linear relationships, which Decision Trees cannot adequately address. Moreover, by including Socio-Economic Weights, as proposed by Al-Hassani [12], the proposed model is more flexible and able to easily adjust to ambient noise, typical of the Iraqi environment, as opposed to the more academically oriented approach followed by El-Halees in Gaza, which could be more easily affected by external factors [13].

3) Algorithmic Advantage: The basic rationale for the efficacy of our results is based upon the idea of "Texture Logic." That is, the previous models (as discussed in Jain and Srivastava) [14] consider the student data as a group of discrete elements. In contrast, the present methodology considers the student trajectory as a continuum of pixels. Weighted Distance works as a carefully designed filter, removing redundant data and guiding the ANN towards the relevant parameters of student dropout [15].

6. Managerial Implications

The outcome of this research provides a number of tangible advantages as well as practical frameworks for policy makers and managers of higher education institutions, particularly on developing countries.

• Primary Intervention Models: The proposed hybrid KDD-ANN early warning system model will allow university managers to shift their approach from passive observation to active intervention. Within the first six weeks of the semester, a counselor is able to determine if a student is at risk and prior to the student dropping out, and therefore provide educational guidance or psycho-social counseling to the student.

• Resource Optimization: University advising centers often have limited staff. This predictive model allows university staff to prioritize students with high-risk flags for advising instead of offering advisory services to issues across the student body.

• Changes in Infrastructure and Policy: The model prioritizing "geographical distance to campus" offers empirical evidence for university administrators to reassess campus infrastructure. This may warrant university transportation services or low-cost on-campus housing to address the commuting dropout problem.

7. Limitations

It is acknowledged that although the proposed hybrid model has incredibly robust prediction properties, there are some limitations that must be taken into account. First and foremost, the proposed dataset is based on the data from a single major university in Iraq; as such, the weights assigned to the features (e.g., the weight assigned to the effect of geographical Distance) may not be the same when applied to universities in different geographical and cultural settings. Secondly, although the proposed model is based on quantified academic and socio-economic parameters, it does not take into account any unrecorded and unexpected life events that may suddenly force the student to quit university. Finally, the conversion of multi-dimensional educational parameters into a "texture-like matrix" is computationally intensive and may necessitate the upgrade of the infrastructure to enable the proposed system to be applied across the country's university network in real-time.

8. CONCLUSION AND FUTURE WORK

The problem addressed in this paper is the problem of academic attrition in the Iraqi higher education system, and the proposed solution is the development of a localized predictive model that utilizes the KDD process and the ANN approach, as well as the Weighted Euclidean Distance metric, to incorporate complex and non-linear variables that affect the problem of academic attrition in the Iraqi context. The proposed approach has achieved promising results, as it has achieved an accuracy level of 94.21%, as well as a precision level of 93.15% and an F1-score level of 93.32%, and it has identified the latent variables that affect the problem of academic attrition, such as the students' Major Subject GPA and the distance from the campus, as the most influential factors in the problem of academic attrition. Future Research Directions: Future research needs to expand the data scope by including more universities and doing a thorough analysis of the applicability of the model at the national level. Further, the inclusion of natural language processing is also proposed as a means to enhance the model and gain a deeper understanding of the psychological aspects related to student dropout through qualitative data, such as student feedback and counseling reports. The realization of the model as a real-time cloud-based early warning system for academic advisors is a significant milestone in the application of the model.

Funding

None

ACKNOWLEDGEMENT

Sincere appreciation is extended to Mustansiriyah University and Universiti Sains Malaysia for their support, and to all colleagues and reviewers for their guidance in the successful completion of this study on predicting Iraqi academic attrition.

Declaration of using AI

Artificial intelligence tools were used only for language editing and grammar correction, and did not contribute to the research design, methodology, analysis, or results of the study.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] I. A. Alhamad and H. P. Singh, "Predicting dropout at master level using educational data mining: A case of public health students in Saudi Arabia," *Amazonia Investiga*, vol. 13, no. 74, pp. 264-275, 2024.

[2] Y. Lu, S. Yeom, J. Maktoubian, M. M. Rahman, and S.-H. Kim, "Improve Student Risk Prediction with Clustering Techniques: A Systematic Review in Education Data Mining," *Education Sciences*, vol. 15, no. 12, p. 1695, 2025.

[3] A. Zanellati, S. P. Zingaro, and M. Gabbrielli, "Balancing performance and explainability in academic dropout prediction," *Interactive Learning Environments*, 2024.

[4] L. Waller and T. Van Duzer, "Computational Microscopy," CITRIS, 2017

[5] M. Fei and D. Y. Yeung, "Temporal Models for Predicting Student Dropout in Massive Open Online Courses," in *Proc. 2015 IEEE Int. Conf. on Data Mining Workshop (ICDMW)*, 2015, pp. 256-263, doi: 10.1109/ICDMW.2015.174.

[6] H. Abbas and M. Salim, "Digital Transformation in Iraqi Higher Education: Challenges of Data Integration," *Journal of Baghdad University for Science*, vol. 15, no. 4, pp. 210-225, 2018.

[7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.

[8] P. Chouhan and M. Tiwari, "Image Retrieval Using Data Mining and Image Processing Techniques," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 3, no. 12, pp. 102-107, 2015.

[9] C. Romero and S. Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 6, p. e1355, 2020.

[10] R. S. Baker and P. S. Inventado, "Educational Data Mining (EDM) and Learning Analytics (LA)," in *Learning Analytics*, Springer, 2014, pp. 61-75.

[11] Q. A. Al-Radaideh, E. M. Al-Shawakfa, and M. I. Al-Najjar, "Mining Student Data Using Decision Trees: A Case Study from Arab Universities," *International Journal of Computer Science Issues*, vol. 17, no. 2, pp. 45-55, 2020.

[12] H. A. Al-Hassani, *Impact of Socio-Economic Instability on Academic Persistence in Iraqi Universities*. Baghdad University Press, 2021.

[13] A. El-Halees, "Mining Educational Data to Improve Student Performance: A Case Study," *International Journal of Information and Communication Technology Education*, vol. 2, no. 1, 2009.

[14] N. Jain and V. Srivastava, "Data Mining Techniques: A Survey," *International Journal of Research in Engineering and Technology*, vol. 2, no. 11, 2013.

[15] M. Vahdat, S. G. Ghalebandi, and S. Geertshuis, "Using Neural Networks to Predict Student Success: A Systematic Review," *Journal of Educational Computing Research*, vol. 59, no. 1, 2021.