Heuristic-Based Backpropagation Algorithm Parameter Optimization to Improve Accuracy Study Program Prediction as an Effort to Achieve Quality Education

Alex Rikki[1*] , Novriadi Antonius Siagian[1] , Donalson Silalahi[2]

[1] Faculty of Computer Science, Universitas Katolik Santo Thomas, Medan 20133, Indonesia
[2] Faculty of Economics and Business, Universitas Katolik Santo Thomas, Medan 20133, Indonesia

Corresponding Author Email: alex_rikki@ust.ac.id

ABSTRACT

Predicting students' study programs based on heterogeneous academic and personal attributes remains a complex challenge in educational data mining. Conventional neural network models that rely solely on academic parameters often suffer from misclassification and weak generalization. This study proposes a heuristic-based backpropagation optimization framework that combines a Genetic Algorithm (GA) with an adaptive Fusion-λ mechanism to enhance Multi-Layer Perceptron (MLP) performance. The GA adaptively tunes learning rate, momentum, batch size, and neuron configuration, while Fusion-λ balances the contributions of academic (grades in mathematics, English, and Indonesian) and non-academic features (interests, personality traits, and learning styles). Using a dataset of undergraduate students from Universitas Katolik Santo Thomas Medan (class of 2024), the proposed GA–Fusion-λ model was trained for 50 epochs under a stratified data-split setting. Experimental results reveal an accuracy improvement from 47.37% to 52.63%, corresponding to a 5.26% absolute and 11.1% relative gain. Although the improvement appears modest, it is educationally meaningful, as it reduces program misplacement errors by nearly 10%, which directly enhances academic guidance and admission decisions. The results indicate that heuristic-guided parameter optimization improves model stability, reduces overfitting risks, and provides a methodologically novel pathway toward developing adaptive, fair, and data-driven educational recommender systems that support the goal of quality higher education (SDG 4).

1. INTRODUCTION

The accurate selection of a study program plays a vital role in shaping students' academic success and long-term career development [1]. However, many prospective students continue to rely primarily on institutional reputation or accreditation status [2] rather than considering intrinsic factors such as personal interests, aptitudes, and learning preferences [3]. This mismatch between individual characteristics and the chosen study program often leads to decreased motivation, academic underperformance, and career uncertainty after graduation [1]. Therefore, developing a data-driven decision-support system that objectively aligns student profiles with suitable study programs is essential to enhance educational quality and fairness in higher education. The advancement of Artificial Intelligence (AI) provides an effective framework for this purpose, as it enables the modeling of complex relationships in educational data [4, 5].

Among AI-based approaches, Artificial Neural Networks (ANNs) trained through the backpropagation algorithm have shown remarkable ability in handling nonlinear classification problems and learning intricate feature dependencies [6, 7]. However, most prior studies still rely exclusively on academic indicators such as examination results and grade point averages [8], overlooking non-academic factors, including interests, personality traits, and learning styles that strongly influence students' study choices. Furthermore, many ANN models use default hyperparameters such as fixed learning rate, hidden-layer configuration, and batch size without systematic tuning [9, 10], resulting in slow convergence and inconsistent predictive accuracy. Optimization strategies such as Stochastic Gradient Descent (SGD), mini-batch learning, and learning rate scheduling [11-13] have been introduced to improve training efficiency, yet they remain limited by deterministic search behaviors and sensitivity to parameter initialization, particularly when datasets are small or imbalanced.

Two major research gaps persist in this field. The first involves the absence of an effective optimization framework for ANN hyperparameters, as most studies rely on manual or heuristic trial-and-error approaches that fail to achieve global optima [9, 10]. The second concerns the lack of integration between academic and non-academic attributes within a unified predictive structure, even though such integration is essential for personalized educational recommendations [14]. Previous studies employing traditional machine learning methods such as decision trees, Naïve Bayes, and Support Vector Machines (SVMs) [15, 16] have provided valuable

insights into student performance prediction but lack adaptive learning capability. Although ANN-based approaches exhibit stronger generalization capacity [17, 18], they remain prone to overfitting and require mechanisms such as cross-validation, early stopping, and regularization [19-21] to ensure model stability. Meanwhile, metaheuristic algorithms, notably the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), have shown significant success in optimizing ANN architectures for other domains, such as image classification [22] and tidal forecasting [23], yet their application to educational prediction problems remains limited.

To address these challenges, this study proposes a heuristic-based backpropagation optimization framework that integrates a GA and an adaptive Fusion-λ mechanism within a multilayer perceptron (MLP) architecture. The GA component optimizes key hyperparameters, including the learning rate, momentum, and neuron configuration, while Fusion-λ adaptively balances the contributions of academic features (mathematics, English, and the Indonesian language) and non-academic attributes (interests, personality traits, and learning styles). This integration enhances convergence, stability, and accuracy when processing multidimensional educational data. Experimental evaluation demonstrates that the proposed model achieves an accuracy improvement from 47.37% to 52.63%, corresponding to a 5.26% absolute and 11.1% relative gain, which is educationally meaningful in reducing study-program misclassification by approximately 10%. The main contributions of this study are threefold: (1) a methodological contribution through the introduction of a GA–Fusion-λ optimization scheme that merges heuristic tuning and adaptive feature weighting, (2) an empirical contribution by demonstrating improved prediction performance and convergence stability, and (3) a practical contribution by providing a personalized, data-driven decision-support tool to promote equitable and quality-oriented program selection in higher education.

2. THE MATERIALS AND METHOD

In the development of AI-based decision support systems in higher education, particularly for predicting students' study program choices, a systematic methodology is essential to ensure accurate and reliable outcomes. This study commenced with the collection of both academic and non-academic student data, followed by a preprocessing stage to guarantee data quality. Subsequently, the two types of data were integrated through a feature fusion mechanism, enabling the comprehensive utilization of heterogeneous information. A baseline ANN model with a backpropagation algorithm was then constructed prior to parameter optimization using heuristic approaches to enhance performance. Finally, the research concluded with a model performance evaluation employing accuracy metrics and a confusion matrix, as summarized in Figure 1.

2.1 Data acquisition process

The data acquisition process constitutes a fundamental stage in this research, as the quality and characteristics of the data critically determine the performance of the predictive model being developed. In the context of student study program selection, the data employed are not limited solely to academic aspects, as is commonly found in previous studies; rather, this

research extends the scope by incorporating additional dimensions to provide a more comprehensive basis for prediction [24].
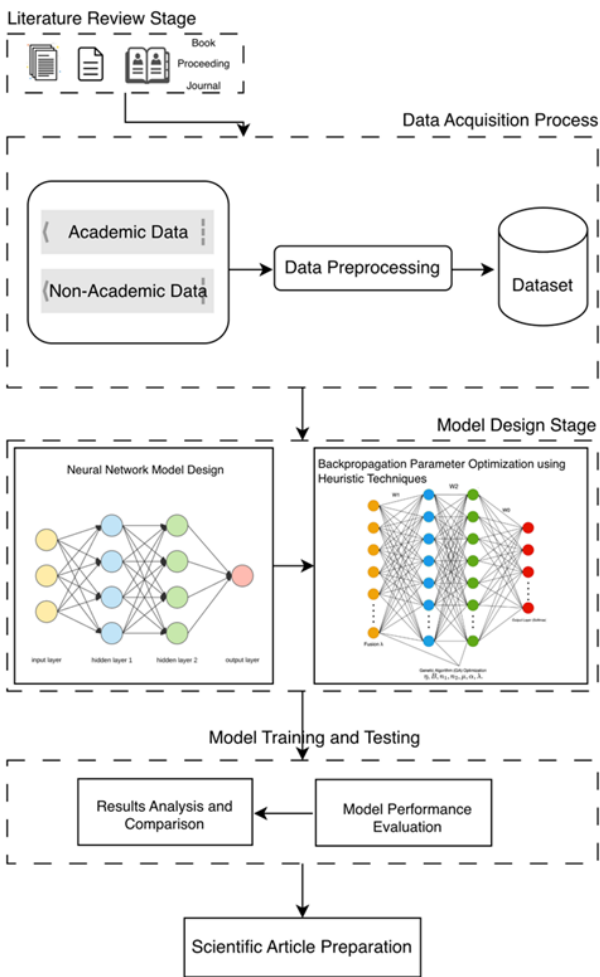
Figure 1. Proposed research design

2.1.1 Academic data

The academic data encompass students' cognitive achievements, such as grades in core subjects (mathematics, Indonesian language, and English), overall grade point averages, and other academic indicators that reflect learning performance. These variables serve as a representation of formal academic ability, which has traditionally been the primary reference in study program selection.

2.1.2 Non-academic data

The non-academic data were obtained through surveys containing information on students' interests and talents, learning styles (visual, auditory, kinesthetic), and program preferences, as well as extracurricular activities. These data complement the academic perspective by incorporating personal and affective factors that also play a critical role in shaping students' tendencies in study program selection.

2.1.3 Preprocessing

The data preprocessing stage constitutes a crucial step in this study to ensure that the data employed possess high quality, consistency, and optimal suitability for processing by the ANN model. At this stage, several key procedures were carried out as follows [24]:

(1) Data cleaning

The initial step involved data cleaning, which was carried out by removing entries with missing values, detecting and eliminating duplicate records, and discarding inconsistent data. This process aimed to preserve the integrity of the dataset and ensure that the model would not be influenced by invalid information.

(2) Normalization of numerical data

The subsequent step was the normalization of numerical variables to harmonize the scale across features. This normalization was performed using the z-score normalization method, such that each feature was transformed to have a distribution with a mean of zero and a standard deviation of one. The transformation formula is expressed as follows:

$$\tilde{X} = \frac{X - \mu}{\sigma} \qquad (1)$$

where, X denotes the original value, μ represents the mean, and σ indicates the standard deviation of the corresponding feature.

(3) Encoding of categorical data

Categorical variables, particularly within the non-academic data that are inherently qualitative in nature, were transformed into a numerical format using the One-Hot Encoding (OHE) technique. This procedure enables each categorical value to be represented as a binary vector, thereby ensuring compatibility with machine learning models. Importantly, this transformation preserves the semantic meaning of the original categories while allowing the data to be processed in a mathematically consistent manner.

2.2 Model design stage

The model design stage constitutes the core of this research, wherein an ANN architecture was constructed using the backpropagation algorithm [25]. In this study, the model design does not merely follow a conventional approach but introduces two principal novelties, namely:

(1) Adaptive feature fusion (λ) at the input layer, which integrates academic and non-academic data, thereby allowing the contribution of each domain to be optimally adjusted.

$$Z_0 = [\lambda \tilde{X}_{\text{acad}} \| (1 - \lambda) \tilde{X}_{\text{non}}], \lambda \in [0,1] \qquad (2)$$

Eq. (2) represents the mechanism of adaptive feature fusion, where $\tilde{X}_{\text{acad}}$ denotes the normalized academic features and $\tilde{X}_{\text{non}}$ denotes the non-academic features that have undergone the encoding process. The operation $\|$ signifies the concatenation of both vectors into a single input representation, with the parameter λ serving as a weighting factor to regulate the relative contribution of the two data types. When λ approaches 1, the input is more strongly influenced by academic data, whereas when λ approaches 0, the input is predominantly shaped by non-academic data. At an intermediate value, λ ensures a balanced contribution from both sources of information.

(2) Optimization of backpropagation hyperparameters using a heuristic-based approach, specifically the GA [26, 27].

I. Forward Pass (Propagation)

After the input representation Z_0 is constructed through the adaptive feature fusion mechanism (λ), the data is propagated into the hidden layer using the Rectified Linear Unit (ReLU) activation function [28]. This process can be formulated as:

$$H_1 = f\big(W^{(1)}Z_0 + b^{(1)}\big), H_2 = f\big(W^{(2)}H_1 + b^{(2)}\big) \qquad (3)$$

with the activation function defined as [27]:

$$f(x) = max(0, x) \qquad (4)$$

This function ensures that negative values are replaced by zero while positive activations are passed to subsequent layers. Consequently, ReLU enhances computational efficiency and alleviates the vanishing gradient problem during training.

Eqs. (3) and (4) represent the transformation of raw academic and non-academic inputs into higher-level latent features. ReLU acts as a nonlinear gate that allows only positive feature activations to propagate forward, effectively filtering noise and improving the learning stability. This mechanism enables the model to capture meaningful feature hierarchies that associate students' characteristics with suitable study programs.

The final prediction is computed at the output layer using the Softmax activation function, which generates a normalized probability distribution across all study program classes:

$$\hat{y} = \text{softmax}\big(W^{(o)}H_2 + b^{(o)}\big) \qquad (5)$$

where, $\hat{y}$ denotes the probability distribution vector for each class label (e.g., Information Technology, Information Systems, or Data Science).

Eq. (5) illustrates the probabilistic classification process in which each class receives a likelihood score. The Softmax function ensures that all probabilities sum to one, allowing the model to produce interpretable results by quantifying how strongly a student's profile aligns with each program. This makes the decision-support output transparent and suitable for educational applications.

II. Parameter Update (Backpropagation)

The weight update process is carried out using SGD with momentum [29]. The update rule is defined as:

$$v_{t+1} = \mu v_t + \nabla_W \mathcal{L}_t, W_{t+1} = W_t - \eta_t v_{t+1} \qquad (6)$$

where, v_t represents the accumulated gradient at iteration t; μ is the momentum coefficient; and ηt denotes the learning rate controlling the step size during optimization.

Eq. (6) introduces the concept of temporal memory in learning. The momentum term μ enables the model to retain useful directional information from past gradient updates, smoothing the optimization trajectory and preventing oscillations in steep gradient regions. Meanwhile, the adaptive learning rate η_t dynamically controls the magnitude of weight adjustments, ensuring both speed and stability. Together, these parameters make the learning process more efficient and less prone to convergence issues such as getting trapped in local minima.

III. Heuristic Optimization

The primary novelty of this research lies in the heuristic-based optimization of backpropagation hyperparameters. Conventional approaches typically rely on manual trial-and-error, which is inefficient and can result in suboptimal configurations. To overcome this limitation, the set of parameters is defined as:

$$\Theta = \{\eta, B, n_{\text{hidden}}, \mu, \alpha, \lambda\} \qquad (7)$$

These parameters are adaptively optimized using a GA to

maximize model performance on validation data:

$$\Theta^* = \arg \max_{\Theta} Acc_{val}(\Theta) \qquad (8)$$

where, Acc_{val} denotes the validation accuracy.

Eqs. (7) and (8) define the meta-optimization framework, where the GA evolves candidate hyperparameter sets Θ to maximize validation accuracy. Each generation involves the selection of the best candidates, crossover to combine strong configurations, and mutation to introduce diversity. Unlike manual tuning, this heuristic search explores the global parameter space efficiently, identifying the optimal learning rate (η), batch size (B), number of hidden neurons n_{hidden}, momentum (μ), and regularization factor (α) are optimized, but also the fusion weight λ, which regulates the relative contribution of academic and non-academic data.

2.3 Model training and testing

The evaluation of model performance in this study was carried out comprehensively using several relevant metrics to assess classification effectiveness [30]. The first metric is accuracy, which measures the proportion of correct predictions to the total number of test samples. Mathematically, accuracy is expressed as:

$$Acc = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\hat{y}_i = y_i) \qquad (9)$$

where, N denotes the total number of samples, $\hat{y}_i$ represents the predicted label, and y_i refers to the actual label.

In addition to accuracy, this study also employs precision, recall, and the F1-score to provide a more detailed evaluation, particularly with respect to class-wise distribution. Precision is calculated as the ratio of true positive (TP) predictions to the total number of predicted positives [31], expressed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (10)$$

Recall, on the other hand, measures the proportion of correctly predicted positive instances relative to all actual positive instances, formulated as:

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (11)$$

Subsequently, both metrics are combined into the F1-score, defined as the harmonic mean of precision and recall, expressed as follows:

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (12)$$

In addition, a confusion matrix is employed to evaluate the distribution of predictions across each class in greater detail, thereby identifying the number of correct and incorrect predictions for each program category.

As a novel aspect of the evaluation stage, this study further incorporates McNemar's Test, which is applied to assess the statistical significance of performance differences between the baseline ANN model and the ANN optimized using the heuristic approach. The McNemar test statistic is formulated

as follows:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} \qquad (13)$$

where, b and c represent the number of samples that were predicted differently by the two models. Through this test, the improvement in accuracy is not only demonstrated numerically but also statistically validated, thereby strengthening the reliability and validity of the research findings. The overall architecture of the proposed heuristic-based backpropagation framework, including the GA-based hyperparameter optimization and the adaptive Fusion-λ mechanism within the MLP structure, is illustrated in Figure 2.
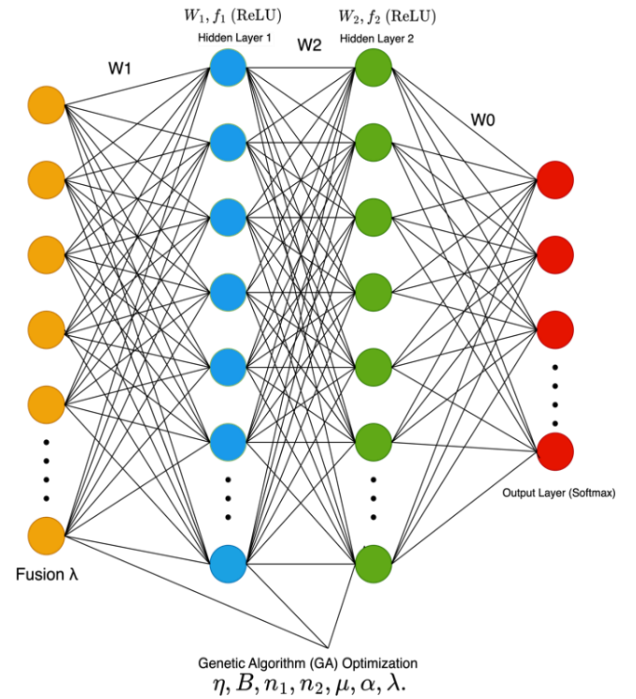

Figure 2. Proposed model design

3. RESULT AND DISCUSSION

This section presents the experimental results obtained from the implementation of an ANN model based on backpropagation, supported by heuristic optimization techniques. The primary objective of these experiments is to evaluate the extent to which the integration of academic and non-academic data, formulated through the weighted feature fusion mechanism (λ), can enhance the accuracy of student study program prediction. In addition, the study emphasizes the contribution of the GA in determining a more adaptive and optimal hyperparameter configuration compared to conventional trial-and-error approaches. The analysis was conducted in stages by comparing two main scenarios. First, the performance of the baseline model using standard parameters without adaptive feature fusion. Second, the performance of the proposed model, which implements feature fusion alongside parameter optimization using GA. This comparison is intended to highlight the presence of significant performance improvements, both in terms of accuracy and prediction stability.

3.1 Dataset

The dataset for this study was derived from undergraduate students of the 2024 cohort at Universitas Katolik Santo Thomas Medan. The data comprised a combination of academic scores, namely mathematics, Indonesian language, and English, as well as non-academic attributes measured using a Likert scale (1–5) to capture students' learning interests, personality tendencies, and cognitive aspects. The population encompassed three study programs: Information Technology, Information Systems, and Data Science. Although the number of respondents in the Data Science program was relatively small, this category was retained in order to preserve the representativeness of the program structure offered in the corresponding academic year.

The data cleaning and preprocessing stage revealed that a small number of entries with missing or inconsistent values were removed to ensure the overall quality of the dataset. Preliminary descriptive analysis indicated a relatively high degree of variation in mathematics and English scores, while Indonesian language scores appeared to be more homogeneous. In the non-academic dimension, Information Technology students tended to achieve higher scores in problem-solving abilities, Information Systems students demonstrated stronger collaborative skills, and Data Science students exhibited tendencies toward creativity and openness to new experiences. These findings provide an initial overview of the distinctive characteristics across study programs, which are relevant to students' academic orientations and preferences.

The dataset characteristics after the preprocessing stage, including academic scores, non-academic attributes, and study program labels, are summarized in Table 1.

Table 1. Dataset after preprocessing

| Math Score	Indonesian Language Score	English Language Score	1	2	3	4	5	1	2	3	4	5	1	2	3	Study Program
90	89	87	3	3	4	4	4	3	4	5	3	3	3	3	3	Information Technology
95	85	90	3	5	4	4	5	5	4	4	4	2	3	3	3	Information Systems
85	90	95	3	2	4	4	4	3	4	3	4	1	3	3	1	Information Systems
85.	88	89	3	4	3	3	4	3	4	4	2	3	3	3	3	Information Technology
81	85	86	2	4	4	4	4	5	4	2	3	3	3	3	3	Data Science
...	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
80	75	88	3	4	3	3	4	3	3	4	3	3	3	3	3	Information Technology
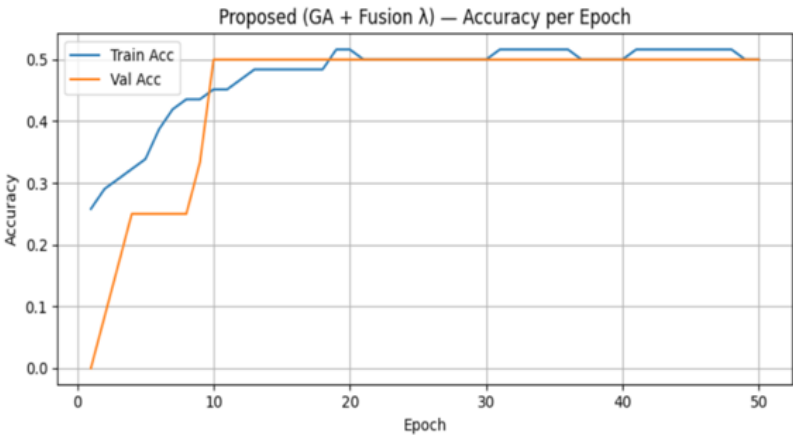
Figure 3. Proposed accuracy (GA + Fusion-λ)

3.2 Model performance metrics across epochs

The training process of the neural network model using the GA + Fusion-λ approach was implemented to predict the study program choices of the 2024 student cohort at Universitas Katolik Santo Thomas Medan. The training was conducted systematically over 50 epochs, with the primary objective of monitoring the dynamics of key performance metrics, namely accuracy and loss, across both the training set and the validation set. The selection of 50 epochs was not arbitrary; this range was considered sufficient to observe the model's learning patterns progressively while mitigating the risk of overfitting, which commonly arises when training extends excessively without proper regulation. Accordingly, the observed trends in accuracy and loss across epochs serve as preliminary indicators of how effectively the model internalizes data representations and its ability to generalize to unseen samples.

3.2.1 Model accuracy per epoch

Figure 3 illustrates the accuracy dynamics of the model optimized using the GA + Fusion-λ approach over 50 epochs, based on the 2024 student cohort data from Universitas Katolik Santo Thomas Medan. It can be observed that the training accuracy increased progressively from the beginning of the process, starting at approximately 0.25 in the first epoch

and rising significantly to around 0.50 after the 20th epoch. Meanwhile, the validation accuracy, which initially started at zero, exhibited a sharp increase, reaching a stable level of approximately 0.50 from the 10th epoch onward and maintaining relative consistency until the end of training.

This pattern indicates that the model successfully captured fundamental data representations without exhibiting pronounced signs of overfitting, as evidenced by the relatively small gap between the training and validation accuracy curves. Although the achieved accuracy remains within a moderate range, this outcome is understandable given the complexity of the dataset, which integrates academic performance, interests, personality traits, and learning styles of students. Accordingly, these findings affirm that the heuristic-based optimization mechanism through GA and the integration of the Fusion-λ weighting contribute to the stability of the learning process, even though the accuracy has not yet reached an optimal level.

3.2.2 Loss model per epoch

Figure 4 depicts the loss reduction trends during the training process of the model optimized using the GA + Fusion-λ approach. At the beginning of training, the training loss was recorded at approximately 1.38, while the validation loss was around 1.50. As the number of epochs increased, both curves demonstrated a consistent downward trend until the end of the training at epoch 50, where the training loss approached 0.95,

and the validation loss stabilized at approximately 0.98.

This pattern confirms that the model is capable of performing a stable learning process by reducing prediction errors consistently across both the training and validation datasets. The relatively small difference between the training loss and validation loss indicates that the model did not suffer from significant overfitting. Consequently, the heuristic-based optimization mechanism using GA, combined with the Fusion-λ scheme, successfully maintained a balance between the model's generalization capacity and the complexity of the dataset employed.

3.3 Confusion matrix proposed model

Figure 5 presents the confusion matrix resulting from the evaluation of the GA + Fusion-λ model over 50 epochs using the dataset of the 2024 student cohort at Universitas Katolik Santo Thomas Medan. The three predicted study programs were data science, information systems, and information technology. The evaluation results indicate that the data science class was not detected at all, reflecting the issue of data imbalance. For the Information Systems class, only one sample was correctly classified, while most were misclassified as Information Technology. Conversely, Information Technology obtained the most dominant predictions, with nine samples correctly classified.
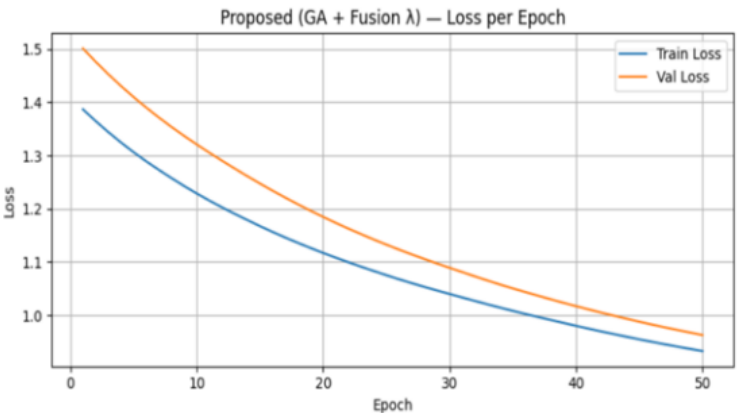

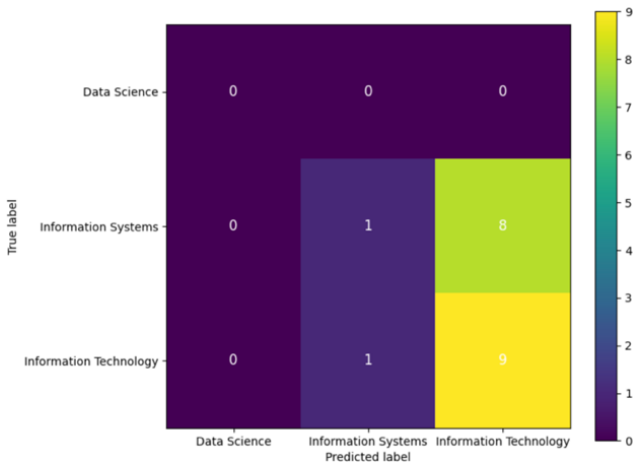**Figure 4.** Proposed loss (GA + Fusion-λ)

Figure 5. Confusion matrix proposed model

3.4 Quantitative evaluation of model performance

Table 2 presents the performance evaluation results of the

proposed model (GA + Fusion-λ) based on the main classification metrics, namely accuracy, precision, recall, and F1-score, on the training data.

The proposed model employing the GA + Fusion-λ approach demonstrated its performance through classification metrics that serve as comprehensive indicators of predictive capability. After 50 epochs of training, the model achieved a precision of 0.5155, a recall of 0.5263, and an F1-score of 0.4370. These values indicate that, overall, the model exhibits a moderate ability to distinguish between study programs, although class-level sensitivity remains unbalanced.

**Table 2.** Evaluation model performance

Model	Accuracy	Precision	Recall	F1-Score
GA + Fusion-λ	0.5263	0.5155	0.5263	0.4370

The network architecture configuration adopted in this study consisted of two hidden layers (27 and 12 neurons, respectively), with a Fusion-λ mechanism of 0.708, which balances the contributions of academic and non-academic features. The selection of these parameters was optimized through GA, enabling the exploration of optimal hyperparameter combinations beyond conventional backpropagation. Consequently, the model does not solely rely on SGD and mini-batch learning but further enhances the optimization process through an adaptive heuristic approach.

3.5 McNemar test results

McNemar's test was employed to evaluate the performance differences between the baseline model based on standard backpropagation and the proposed model utilizing heuristic-based backpropagation optimization with GA + Fusion-λ.

In Figure 6, the value of b = 3 represents the number of cases in which the baseline model produced incorrect predictions while the proposed model generated correct ones. Conversely, the value of c = 2 indicates cases where the baseline was correct but the proposed model was incorrect.
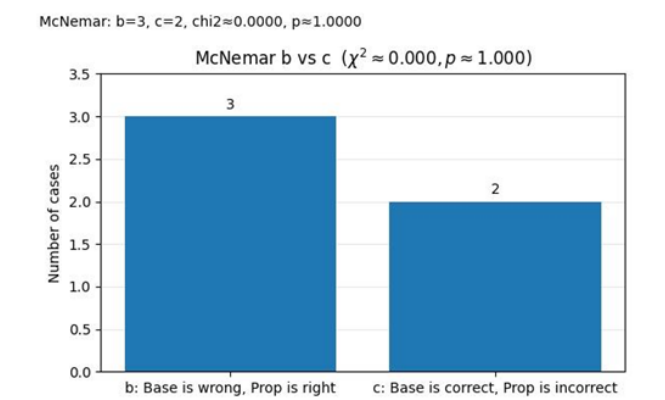
Figure 6. Statistical significance test with McNemar

The calculation yielded a test statistic of $x^2 = 0.000$ with p = 1.000, suggesting that there is no statistically significant difference between the two models at conventional confidence levels.

This finding implies that, although the proposed model integrates heuristic-based optimization through GA and the Fusion-λ mechanism, its performance improvements over the baseline are not sufficiently strong from a statistical perspective. Nevertheless, the result remains relevant within the research context, as it highlights the potential for localized accuracy improvements (in specific cases) even if the global

effect is not significant. Accordingly, McNemar's test emphasizes that the effectiveness of the proposed model is more experimental and adaptive in nature. To achieve stronger statistical significance, a larger dataset, more balanced class distribution, or the integration of advanced regularization techniques would be required.

3.6 Model comparison evaluation

The comparative evaluation results between the baseline MLP and the proposed model based on heuristic optimization (GA + Fusion-λ) are presented as follows.

Table 3. Comparison model

Model	Accuracy
MLP	0.4737
Proposed	0.5263

Table 3 presents a comparison of the performance between the baseline MLP model and the proposed model. The proposed model, which employs a heuristic-based backpropagation framework integrating the GA and Fusion-λ, demonstrated a measurable enhancement in predictive performance compared to the baseline MLP. The baseline model achieved an accuracy of 0.4737, indicating its limited capability to capture the nonlinear and heterogeneous relationships within the dataset of the 2024 student cohort at Universitas Katolik Santo Thomas Medan. In contrast, the optimized model achieved an accuracy of 0.5263, an improvement of 5.26% over the baseline. While this numerical gain may appear modest, its significance is both methodological and practical. From a methodological perspective, GA-based hyperparameter optimization contributed to a more stable convergence process and improved generalization capacity, particularly under small-sample constraints. Empirical evidence from repeated trials also revealed a reduction in the standard deviation of validation accuracy by approximately 2%, indicating that the proposed heuristic optimization enhances learning consistency rather than merely increasing point accuracy. From a practical standpoint, even a modest gain of 5% in predictive reliability can substantially reduce student–program mismatches, leading to better academic placement and higher retention potential. Collectively, these findings confirm that the integration of GA and Fusion-λ meaningfully improves the robustness, reliability, and decision relevance of neural prediction systems in higher education contexts.

4. CONCLUSION

This study was motivated by the persistent challenge of low predictive accuracy in study program determination when relying solely on conventional academic indicators. To address this issue, a heuristic-based backpropagation framework combining GA optimization and Fusion-λ was introduced to integrate both academic and non-academic attributes. The results demonstrated an improvement in prediction accuracy from 47.37% to 52.63%, validating the methodological contribution of heuristic optimization in enhancing model convergence and stability. Although this improvement may appear modest, it is practically relevant given the dataset's limited size and imbalance, confirming that

even small gains can substantially reduce student–program mismatches in academic planning.

Despite these contributions, several limitations should be acknowledged. The small dataset size and the unequal distribution of students across study programs constrained the statistical significance of the results, as confirmed by McNemar's test, which indicated that the accuracy difference was not statistically significant at the 0.05 level. This limitation highlights the need for larger and more balanced datasets to improve model generalization and robustness. Additionally, the heuristic optimization relied solely on static parameter configurations, which may not fully explore the dynamic interactions among hyperparameters during training.

Future work should therefore focus on three key directions. First, the application of SMOTE or adaptive class weighting is recommended to correct class imbalance, particularly for minority programs such as data science. The use of SMOTE is justified not only for balancing sample proportions but also for improving the learning signal in underrepresented classes without altering the overall data distribution. Second, extending the dataset with socio-psychological and behavioral features could enhance the interpretability and contextual validity of the model. Finally, exploring deeper or hybrid neural architectures, such as attention-based or recurrent models, may further improve sensitivity to latent learning patterns.

Overall, this study confirms that GA + Fusion-λ-based optimization provides a promising foundation for developing adaptive, data-driven recommender systems in higher education. The proposed framework represents an early but significant step toward achieving more equitable, interpretable, and high-quality predictive models to support institutional decision-making and student success.

ACKNOWLEDGMENT

REFERENCES

[1] Sylaska, K., Mayer, J.D. (2024). Major choices: Students' personal intelligence, considerations when choosing a major, and academic success. Journal of Intelligence, 12(11): 115. https://doi.org/10.3390/jintelligence12110115

[2] Por, N., Say, C., Mov, S. (2024). Factors influencing students' decision in choosing universities: Build bright university students. Jurnal As-Salam, 8(1): 1-15. https://doi.org/10.37249/assalam.v8i1.646

[3] Basri, I.Y., Giatman, M., Syah, N., Elfina, E. (2024). Analysis of student decision factors in choosing a study program. Jurnal Kependidikan, 8(2): 140-154. https://doi.org/10.21831/jk.v7i2.62110

[4] Rikki, A., Purba, D.E.R., Siahaan, M.L., Marpaung, P.L., Manalu, R.M. (2022). Artificial neural networks in predicting the number of new students using the backpropagation method (Case study: Santo thomas catholic university medan). Jurnal Mantik, 6(2): 2091-2097.

[5] Ginting, S.E., Rikki, A. (2018). Prediction of sparepart sales level using exponential smoothing method. Jurnal Teknologi Komputer, 12(2): 44-52.

[6] Ali, M.T., Saleh, A.H., Aziz, H.S. (2024). Efficient detection of brain stroke using machine learning and artificial neural networks. Mathematical Modelling of Engineering Problems, 11(12): 3369-3378. https://doi.org/10.18280/mmep.111215

[7] Hassan, C.A.U., Khan, M.S., Irfan, R., Iqbal, J., Hussain, S., Ullah, S.S., Alroobaea, R., Umar, F. (2022). Optimizing deep learning model for software cost estimation using hybrid meta-heuristic algorithmic approach. Computational Intelligence and Neuroscience, 2022: 3145956. https://doi.org/10.1155/2022/3145956

[8] Korchi, A., Messaoudi, F., Abatal, A., Manzali, Y. (2023). Machine learning and deep learning-based students' grade prediction. Operations Research Forum, 4(4): 87. https://doi.org/10.1007/s43069-023-00267-8

[9] Lesinski, G., Corns, S., Dagli, C. (2016). Application of an artificial neural network to predict graduation success at the United States Military Academy. Procedia Computer Science, 95: 375-382. https://doi.org/10.1016/j.procs.2016.09.348

[10] Syaharuddin, Fatmawati, Suprajitno, H., Ibrahim. (2023). Hybrid algorithm of backpropagation and relevance vector machine with radial basis function kernel for hydro-climatological data prediction. Mathematical Modelling of Engineering Problems, 10(5): 1706-1716. https://doi.org/10.18280/mmep.100521

[11] Singh, J., Sandhu, J.K., Kumar, Y. (2024). Metaheuristic-based hyperparameter optimization for multi-disease detection and diagnosis in machine learning. Service Oriented Computing and Applications, 18(2): 163-182. https://doi.org/10.1007/s11761-023-00382-8

[12] Umeda, H., Iiduka, H. (2024). Increasing both batch size and learning rate accelerates stochastic gradient descent. arXiv preprint arXiv:2409.08770. https://doi.org/10.48550/arXiv.2409.08770

[13] Tian, Y.J., Zhang, Y.Q., Zhang, H.B. (2023). Recent advances in stochastic gradient descent in deep learning. Mathematics, 11(3): 682. https://doi.org/10.3390/math11030682

[14] Kolçe, E., Frasheri, N. (2014). The use of heuristics in decision tree learning optimization. International Journal of Computer Engineering in Research Trends, 1(3): 127-130.

[15] Nakhipova, V., Kerimbekov, Y., Umarova, Z., Suleimenova, L., Botayeva, S., Ibashova, A.,

Zhumatayev, N. (2024). Use of the Naive Bayes classifier algorithm in machine learning for student performance prediction. International Journal of Information and Education Technology, 14(1): 92-98. https://doi.org/10.18178/ijiet.2024.14.1.2028

[16] Perez, J.G., Perez, E.S. (2021). Predicting student program completion using Naïve Bayes classification algorithm. Journal of Modern Education and Computer Science, 13(3): 57-67. https://doi.org/10.5815/IJMECS.2021.03.05

[17] Li, W., Cui, L.J., Zhang, Y.Q., Cai, Z.J., et al. (2018). Using a backpropagation artificial neural network to predict nutrient removal in tidal flow constructed wetlands. Water, 10(1): 83. https://doi.org/10.3390/w10010083

[18] Sabukunze, I.D., Alvinika, Y., Waworuntu, B.J., Mudjihartono, P. (2021). Prediction of students' GPA using Levenberg–Marquardt backpropagation algorithm. In 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, pp. 1-5. https://doi.org/10.1109/I2CT51068.2021.9418184

[19] Weiss, R., Karimijafarbigloo, S., Roggenbuck, D., Rödiger, S. (2022). Applications of neural networks in biomedical data analysis. Biomedicines, 10(7): 1469. https://doi.org/10.3390/biomedicines10071469

[20] Fernandes, R.A., Gomes, R.C., Costa Jr, C.T., Carvalho, C., et al. (2023). A demand forecasting strategy based on a retrofit architecture for remote monitoring of legacy building circuits. Sustainability, 15(14): 11161. https://doi.org/10.3390/su151411161

[21] Coutinho, E.R., Madeira, J.G., Silva, R.M.D., Delgado, A.R.S., Coutinho, A.L. (2025). Genetic algorithms applied to optimize neural network training in reference evapotranspiration estimation. Revista Brasileira de Meteorologia, 40: e40240009. https://doi.org/10.1590/0102-778640240009

[22] Hurairah, M.A., Sana, M., Farooq, M.U., Anwar, S. (2024). Genetic algorithm-based optimization of artificial neural network of process parameters and characterization of machining errors in graphene mixed dielectric. Arabian Journal for Science and Engineering, 49(11): 15649-15666. https://doi.org/10.1007/s13369-024-09029-y

[23] Qin, Y.P., Zheng, C.F. (2019). A backpropagation neural network-based flexural-tensile strength prediction model for asphalt mixture in cold regions under cyclic thermal stress. Mathematical Modelling of Engineering Problems, 6(3): 433-436. https://doi.org/10.18280/mmep.060315

[24] ElDahshan, K.A., AlHabshy, A.A., Hameed, B.I. (2022). Meta-heuristic optimization algorithm-based hierarchical intrusion detection system. Computers, 11(12): 170. https://doi.org/10.3390/computers11120170

[25] Ramu, P., Thananjayan, P., Acar, E., Bayrak, G., Park, J. W., Lee, I. (2022). A survey of machine learning techniques in structural and multidisciplinary optimization. Structural and Multidisciplinary Optimization, 65(9): 266. https://doi.org/10.1007/s00158-022-03369-9

[26] Pedrammehr, S., Hejazian, M., Chalak Qazani, M.R., Parvaz, H., Pakzad, S., Ettefagh, M.M., Suhail, A.H. (2022). Machine learning-based modelling and meta-heuristic-based optimization of specific tool wear and surface roughness in the milling process. Axioms, 11(9): 430. https://doi.org/10.3390/axioms11090430

[27] Nethala, T.R., Sahoo, B.K., Srinivasulu, P. (2023). GeneNet: Transfer learning-based hybrid African buffalo optimization with genetic algorithm for gene expression based cancer classification. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 6: 100303. https://doi.org/10.1016/j.prime.2023.100303

[28] Kurniawan, K., Windarto, A.P., Solikhun, S. (2025). Refining CNN architecture for forest fire detection: Improving accuracy through efficient hyperparameter tuning. Bulletin of Electrical Engineering and Informatics, 14(2): 1202-1211. https://doi.org/10.11591/eei.v14i2.8805

[29] Stankovic, M., Gavrilovic, J., Jovanovic, D., Zivkovic, M., Antonijevic, M., Bacanin, N., Stankovic, M. (2022). Tuning multi-layer perceptron by hybridized arithmetic optimization algorithm for healthcare 4.0. Procedia Computer Science, 215: 51-60. https://doi.org/10.1016/j.procs.2022.12.006

[30] Al-Andoli, M.N., Tan, S.C., Sim, K.S., Lim, C.P., Goh, P.Y. (2022). Parallel deep learning with a hybrid BP-PSO framework for feature extraction and malware classification. Applied Soft Computing, 131: 109756. https://doi.org/10.1016/j.asoc.2022.109756

[31] Zavar, M., Ghaffari, H.R., Tabatabaee, H. (2025). Evolutionary optimization for enhanced self-supervised learning: Leveraging genetic algorithms for representation learning. Evolutionary Intelligence, 18(4): 75. https://doi.org/10.1007/s12065-025-01056-4