

A Hybrid Fuzzy–Neural Decision-Support Framework for Graduate Learning Outcome Attainment Evaluation in a Computer Science Program

Suendri^{1*}, Ratni Sirait², Ilka Zufria³

^{1,2,3} Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

ABSTRACT

Open Access

Type: Research Paper

Article:

Received: April 12, 2026

Revised: May 20, 2026

Accepted: August 13, 2026

Issued: October 30, 2026

© Suendri et al. (2026).



This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

*Correspondence Address:
suendri@uinsu.ac.id

A passing threshold of 70 assigns different attainment labels to students scoring 71.0 and 66.5, although both are close enough to warrant review. This study develops a fuzzy-neural decision-support framework for graduate learning outcome assessment in an undergraduate Computer Science program. Weighted course learning outcome scores determine the reference status. Fuzzy rules provide category memberships and a soft score, with cases within five points of 70 referred to a human reviewer. A single-layer perceptron produces a separate estimate from normalized course scores. Four anonymized records and one partial-profile scenario illustrate the calculations. The fuzzy results retain all reference classifications. Two of the four complete records are referred for review; including the additional scenario gives three referrals among five cases. The referral set is unchanged at half-widths of 5, 7, and 10. After one training pass, the perceptron misclassifies the partial-profile scenario even when the assumed missing score is supplied. The weighted calculation gives the reference status using either the available scores or the assumed complete profile. These results document the referral procedure and the limits of the neural prototype. The design keeps its estimate separate from the official status and leaves borderline decisions to a reviewer. Predictive performance and the consistency of human review require evaluation on larger datasets.

Keywords: *graduate learning outcomes; fuzzy logic; neural network; outcome-based education; educational assessment; hybrid artificial intelligence; decision support; human-in-the-loop*

1. INTRODUCTION

The evaluation of graduate learning outcome attainment (Capaian Pembelajaran Lulusan, CPL) plays a strategic role in ensuring graduate quality and institutional accountability in higher education [1], [2]. In Indonesian outcome-based curricula aligned with the Indonesian National Qualifications Framework (KKNI) and Outcome-Based Education (OBE), program-level learning outcomes (CPL) are operationalized through course-level learning outcomes (Capaian Pembelajaran Mata Kuliah, CPMK), and attainment is determined by aggregating weighted CPMK scores and comparing the result against a predefined threshold. This aggregation-and-threshold procedure is transparent and easy to compute, and it underpins the quality-assurance reporting expected under national accreditation [3]. The procedure nonetheless has a well-known weakness. A crisp threshold treats a score of 69.9 and a score of 70.0 as categorically different, even though the two are practically indistinguishable, and it collapses all failing scores into a single class regardless of how close a student is to the boundary. For a student whose aggregate sits just below the threshold, a binary verdict is difficult to defend pedagogically and conveys no information about how near the student is to attainment. As assessment data grow in volume, manual spreadsheet-based handling of these borderline cases also becomes inconsistent across lecturers, a limitation that has motivated fuzzy approaches to student assessment [4].

These limitations motivate an evaluation approach that is more transparent about the borderline region while preserving the official aggregation rule. Fuzzy logic is well suited to the first requirement: through graded membership, in linguistic categories it represents how strongly a score belongs to adjacent categories, and because it is rule-based, each decision remains inspectable [5], [6]. A neural network can serve a second, narrower requirement: when the complete CPMK profile or its contribution weights are not yet available, a learned model can produce an early, complementary estimate of likely attainment from the partial information at hand [7], [8]. Combining the two paradigms yields a decision-support instrument that keeps the interpretability of fuzzy

inference and adds a predictive component for incomplete-information settings [9].

We examine how a program can retain its weighted CPL calculation while giving reviewers more information about scores near the passing threshold. The proposed workflow specifies which output determines attainment, what the fuzzy layer reports, and when a case is referred for human review. A perceptron is included to illustrate a supplementary estimate from course scores. Section 2.6 defines these roles. The study evaluates the calculation procedure through worked examples; it does not test predictive superiority or establish generalizability. The contribution is the decision and referral procedure. Related work follows below, with the methodology, results, and conclusion presented in Sections 2–3 and the Conclusion section.

CPL and program learning outcome (PLO) reporting have been implemented in institutional assessment systems. Kushari and Septiadi [3] designed an assessment information system that links PLO to course outcomes and curricular activities so that program managers, lecturers, and students can monitor attainment throughout a semester. Yuniarti and colleagues [10] developed a learner assessment system for e-learning that reports activity performance and ability levels under an OBE approach, while Diyasa and colleagues [11] implemented an integrated evaluation and learning system built on an OBE ecosystem. Assessment tools and rubrics for such outcome-based reporting have also been consolidated by Sasipraba and colleagues [12]. These systems establish the practical need for transparent, system-supported CPL reporting, and they largely rely on deterministic weighted aggregation.

Fuzzy logic has been applied to educational assessment precisely because of its interpretability. Aziz and Hashem [13] used a Mamdani fuzzy inference system to assess student learning outcomes in the implementation of outcome-based education, and Ozdemir and Balbal [14] employed fuzzy logic to analyse the performance of engineering students, while comparable fuzzy grading schemes have been applied to best-student selection and online engineering courses [5], [15], highlighting the flexibility of fuzzy methods in handling the many parameters that influence

attainment. Yağcı [16] investigated student performance prediction using machine-learning algorithms, while Ahmad and colleagues [17] used an artificial neural network. Nachouki and colleagues [9] examined factors affecting student performance with a neuro-fuzzy model. These studies place prediction alongside fuzzy approaches to assessment. Two recent lines of work bear directly on the present study. Loresco and Pilueta [18] proposed an adaptive neuro-fuzzy framework with an individualized dashboard for monitoring student outcomes in outcome-based education, in which neural learning adapts the parameters of the fuzzy inference system itself. Naccache and Hleiss [19] showed, in preservice teacher education, that a fuzzy assessment layer reduces classification risk for candidates whose aggregate scores fall near a decision boundary, and the meta-analysis of Tuan and colleagues [20] reports that the measured effect of fuzzy assessment depends strongly on the configuration of the input and output variables, which makes complete parameter reporting a precondition for comparability. In parallel, human-in-the-loop designs for academic assessment have been formalized by Khan and colleagues [21], who address the policy gaps that arise when automated scoring is used without an explicit human-review stage, and by Fajardo-Ramos and colleagues [22], who argue that human-in-the-loop assessment requires an institutionally defined division of responsibility rather than an accumulation of tools. Reviews of explainable artificial intelligence in education [23], [24] converge on the same requirement for high-stakes decisions: the mechanism that produces a decision must be inspectable by the person accountable for it. Shabbir and colleagues [25] further show that course- and program-level outcome attainment is itself sensitive to institutional conditions such as teaching workload, which limits the interpretation of any attainment figure as a pure measure of student ability.

Existing outcome-based assessment systems commonly rely on deterministic weighted aggregation followed by threshold comparison, while fuzzy and neuro-fuzzy approaches have separately been used to represent uncertainty in scoring [13], [19], [20] and to predict student performance [9], [16], [17], [18]. For the program

considered here, the design requirement is to keep the weighted CPL result authoritative, describe scores near the threshold through fuzzy memberships, and refer borderline cases to a reviewer. Neural estimates from incomplete profiles are recorded as supplementary information. The novelty claimed here is therefore not the combination of fuzzy logic and neural networks, which is well established, but the ordering of the components and the allocation of authority among them: official weighted CPL calculation, fuzzy interpretation, borderline identification, neural supplementary estimate, human review. This distinguishes the design from ANFIS-type systems such as [18], in which neural learning directly adapts the parameters of the fuzzy inference system, so that the resulting membership functions are no longer fixed by institutional policy. The present study addresses this gap at the level of framework design and illustrative demonstration. The neural component is trained only on a small worked set and a single pass through the data, so no claim of predictive novelty or predictive superiority is made.

2. METHODS

2.1 Institutional Context

The study draws on the undergraduate Computer Science program of the Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, which has implemented a KKNI–OBE curriculum since 2021. The program defines five graduate profiles and eleven CPL derived from them, spanning the attitude, general-skill, specific-skill, and knowledge elements required by the national standard (SN-Dikti) at KKNI Level 6. Each CPL is operationalized by course-level CPMK embedded in the semester learning plan (RPS) of each course. The demonstration uses these CPMK-to-CPL mappings and the weights recorded in the RPS. Another program would need to supply its own mappings, weights, and attainment threshold before applying the procedure.

2.2 CPMK and CPL Data Structure

Each record associates a student with a course offering and contains the vector of CPMK scores for that course, the contribution weights of

each CPMK to the CPL it supports (defined in the RPS), the aggregate CPL score, and the resulting attainment status. All identifying data used for demonstration were anonymized. Table 1 lists the

eleven program CPL, and Table 2 maps them to the five graduate profiles (PL01–PL05); these institutional definitions make the target of evaluation explicit and auditable.

Table 1. Graduate learning outcomes (CPL) of the Computer Science program

No	CPL	CPL code	CPL description
1	CPL-01	CPL-S01	Demonstrates devotion to God Almighty and a religious attitude.
2	CPL-02	CPL-S02	Respects cultural, philosophical, religious, and belief diversity, and the original opinions or findings of others.
3	CPL-03	CPL-S03	Possesses the ability to collaborate, stays motivated, and takes responsibility for their work.
4	CPL-04	CPL-KU03	Makes appropriate decisions in solving problems within their field of expertise, based on the analysis of information and data.
5	CPL-05	CPL-KU04	Is able to define user or market needs regarding performance (analyzing, evaluating, and developing) of computer-based algorithms/methods.
6	CPL-06	CPL-KK05	Is able to build and process big data using programming languages, and to apply platform-based development concepts in mobile computing.
7	CPL-07	CPL-KK06	Is able to build wireless network systems and assemble intelligent-system devices; analyzes, designs, and implements computer peripherals and cloud computing.
8	CPL-08	CPL-KK07	Is able to build intelligent devices, integrate them into computer networks, and upload data through intelligent applications related to the Internet of Things (IoT).
9	CPL-09	CPL-P08	Understands the fundamental theory of computer architecture, including computer hardware and networks.
10	CPL-10	CPL-P09	Has adequate knowledge of how computer systems work and can apply various algorithms/methods to solve problems in an organization.
11	CPL-11	CPL-P10	Has the competence to analyze complex computing problems in order to identify project-management solutions in the field of informatics/computer science, considering transdisciplinary developments.

Table 2. Mapping of graduate learning outcomes (CPL) to graduate profiles (PL)

CPL code	PL01	PL02	PL03	PL04	PL05
CPL-1	✓	✓	✓	✓	✓
CPL-2	✓	✓	✓	✓	✓
CPL-3	✓	✓	✓	✓	✓
CPL-4	✓	✓	✓	✓	✓
CPL-5	✓	✓		✓	✓
CPL-6	✓				✓
CPL-7		✓	✓		✓
CPL-8			✓		✓
CPL-9		✓	✓		✓
CPL-10	✓				✓
CPL-11	✓	✓	✓		✓

2.3 Weighted CPL Calculation

For a given student i and a target outcome CPL k , the attainment score is the weighted average of the contributing CPMK scores:

$$S(i, k) = \sum_j (x(i, j) \cdot w(j \rightarrow k)) / \sum_j w(j \rightarrow k), \quad (1)$$
 where $x(i, j)$ is the score of student i on CPMK j , $w(j \rightarrow k)$ is the contribution weight of CPMK j to CPL k , and the sum runs over the CPMK that

contribute to CPL k . The institutional decision rule declares CPL k attained when $S(i, k) \geq \tau$, with the program threshold $\tau = 70$. This weighted average is the official reference decision that the hybrid framework supports rather than replaces.

Three aspects of Equation (1) are fixed by institutional policy rather than chosen for this study. First, the contribution weights $w(j \rightarrow k)$ are declared in the semester learning plan (RPS) of each course when the course is designed and are approved together with the curriculum document; they are not estimated from student data. Second, the within-course weights for each CPL sum to 1. The denominator in Equation (1) is therefore 1 in the complete worked example. The equation retains the denominator because a CPL may draw on several courses, and the contributing CPMK sets can differ or overlap across CPL. Third, the program's academic regulations set $\tau = 70$ on the 0–100 scale. This value was not selected through model fitting. When a score is unavailable, its CPMK is omitted from both sums in Equation (1). The result then summarizes only the assessed

components. Section 2.5 describes the separate neural estimate.

2.4 Fuzzy Inference Design

The fuzzy input is the weighted score $S(i, k)$. Table 3 assigns it membership in Not Attained, Fair, Good, and Excellent using trapezoidal or triangular functions. For $65 < S < 75$, Fair and Good both have positive membership. A score of 66.5, for example, has memberships of 0.85 and 0.15 in these two categories. The implementation uses Mamdani-type rules with singleton consequents. Each rule fires at its antecedent

membership degree, and Equation (2) combines the outputs into a soft score. The score, memberships, and active rules are retained in the assessment record [26], [27].

To make the fuzzy component reproducible, its parameters are reported in full. The four linguistic categories form a Ruspini partition of the input universe $[0, 100]$: the membership degrees of any score sum to unity and at most two categories are active simultaneously. Table 3 gives the type and the exact parameters of each membership function, in the conventional notation trapezoid(a, b, c, d) and triangle(a, b, c).

Table 3. Membership-function parameters of the fuzzy input variable S (universe $[0, 100]$)

Linguistic category	MF type	Parameters	Support	Core
Not Attained	Trapezoidal	(0, 0, 55, 65)	$[0, 65]$	$[0, 55]$
Fair	Triangular	(55, 65, 75)	$(55, 75)$	$\{65\}$
Good	Triangular	(65, 75, 85)	$(65, 85)$	$\{75\}$
Excellent	Trapezoidal	(75, 85, 100, 100)	$(75, 100]$	$[85, 100]$

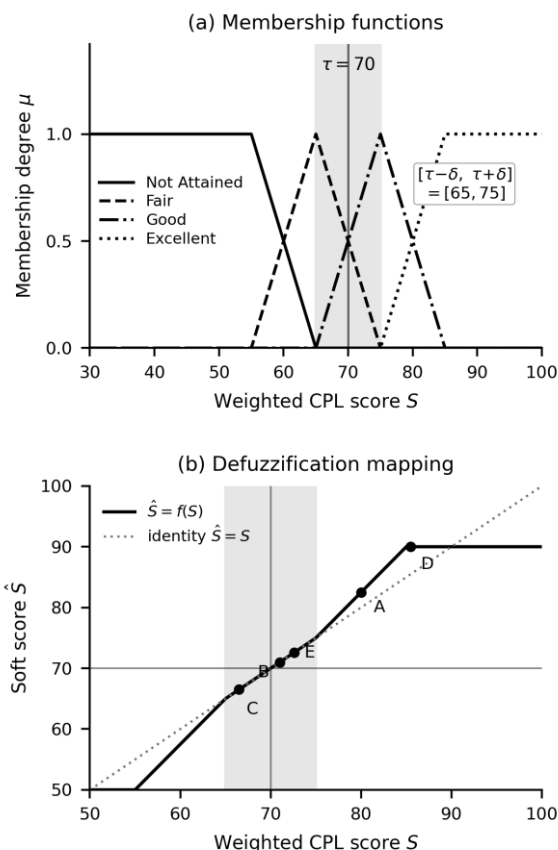


Figure 1. (a) Membership functions of the four linguistic categories over the input universe, with the threshold $\tau = 70$ and the borderline band $[65, 75]$ shaded; (b) the resulting defuzzification mapping $S \rightarrow \hat{S}$, with the five worked cases marked. Inside the borderline band the mapping coincides with the identity.

Figure 1(a) plots the resulting partition. The threshold $\tau = 70$ lies at the midpoint of the Fair–Good overlap region $[65, 75]$, the interval in which two categories are simultaneously active and in which a crisp threshold is therefore least defensible. Table 4 reports the complete rule base together with the location of each output singleton.

Table 4. Complete fuzzy rule base and output singleton locations

Rule	Antecedent (S is ...)	Consequent (attainment is ...)	Output singleton c_i
R1	Not Attained	Not Attained	50
R2	Fair	Fair	65
R3	Good	Good	75
R4	Excellent	Excellent	90

Each of the four rules fires at the membership degree of its antecedent. The consequents are fuzzy singletons located at c_i (Table 4), so defuzzification reduces to the weighted average

$$\hat{S} = \sum_i (\mu_i(S) \cdot c_i) / \sum_i \mu_i(S), \quad (2)$$

where $\mu_i(S)$ is the firing strength of rule R_i . Because the antecedent partition is Ruspini, the denominator equals 1 and Equation (2) is exact. Singleton consequents make the defuzzified value analytically reproducible from Table 3 and Table 4 alone. With constant singleton outputs, Equation (2) is equivalent to zero-order Takagi–Sugeno inference.

Each rule evaluates the aggregate S as its only antecedent. This keeps the number of rules equal to the number of linguistic categories, makes every evaluation auditable by inspecting at most two fired rules, and guarantees that the fuzzy layer cannot alter the ordering produced by the official weighted calculation. The cost is that the fuzzy layer cannot represent interaction between CPMK, for instance a uniformly mediocre profile against one that combines a high and a failing component at the same aggregate. Capturing such interaction would require a multi-antecedent rule base whose rules would have to be elicited from and validated by the teaching team rather than assumed.

2.5 Neural Network Prototype

The neural component is a feed-forward perceptron that takes the normalized CPMK score vector $\tilde{x} = x / 100$ as input and produces a binary attainment estimate for the target CPL, using a step activation during this prototype demonstration [7], [28]. Its purpose is narrow: it estimates attainment from the CPMK scores without using the contribution weights, which is the situation that arises when the weighting scheme for a CPL has not yet been fixed. When the weights are available, Equation (1) can itself be applied to the components already assessed, and Section 3.4 shows that in the example examined the weighted calculation is the more accurate of the two. The worked set is used only to illustrate error-driven updates to the weights and bias. No predictive validation is provided, and the output cannot override the official weighted status.

Table 5 lists the perceptron's initialization, update rule, training order, and missing-input convention used in Section 3.4.

Table 5. Complete configuration of the perceptron prototype

Item	Specification
Architecture	Single-layer perceptron; three inputs, one output, no hidden layer
Inputs	Normalized CPMK scores $\tilde{x}(i, j) = x(i, j) / 100 \in [0, 1]$; all three CPMK of the target course
Activation	Unit step: $y = 1$ if $z \geq 0$, otherwise $y = 0$, with $z = w \cdot \tilde{x} + b$
Initial weights	$w^0 = (0.10, 0.10, 0.10)$
Initial bias	$b^0 = -0.50$
Learning rate	$\eta = 0.10$
Update rule	$w \leftarrow w + \eta (d - y) \tilde{x}$; $b \leftarrow b + \eta (d - y)$
Reference label d	Official weighted status from Equation (1): 1 = attained, 0 = not attained
Training set	The four records of Table 6, presented in the order A, B, C, D

Item	Specification
Epochs	1 (fixed by design, for illustration)
Stopping criterion	None; training halts after the single pass
Missing CPMK handling	The unavailable input is substituted with 0 after normalization, so a missing component contributes nothing to z
Validation	None; no held-out set, therefore no accuracy, precision, recall, or AUC is reported

One pass over four records illustrates how the parameters are updated but provides no evidence of generalization: four points in a three-dimensional input space are easily separated, so reproducing them says nothing about unseen records. The prototype also replaces an unavailable normalized score with zero. With the positive weights obtained in this example, that substitution lowers the pre-activation z relative to supplying any positive score for the missing component, so the prototype tends to report not attained on incomplete profiles. The direction of this effect depends on the sign of the learned weights and is therefore not a general property of the design. The relative cost of a false attained and a false not attained estimate was not evaluated, so the tendency cannot be described as conservative in a useful sense. Missing-data handling and the consequences of both error types require evaluation before operational use.

2.6 Fusion Mechanism

The fusion rule uses the fuzzy output to decide whether a case requires human review. Let $\hat{S}$ denote the soft score, $\tau = 70$ the threshold, and $\delta = 5$ the borderline half-width. Cases satisfying $|\hat{S} - \tau| \leq \delta$ are referred; the reviewer receives the weighted score, the membership degrees, the soft score, and, when an incomplete profile is under consideration, the supplementary neural estimate. Cases satisfying $|\hat{S} - \tau| > \delta$ are not referred by this rule. In both cases the official weighted calculation remains the reference for attainment status; neither the fuzzy layer nor the neural component can change it, and final academic decisions remain with the responsible lecturers and program managers.

The half-width $\delta = 5$ was not tuned on data. It follows from two institutional considerations. First, the program's letter-grade bands on the 0–100 scale are 10 points wide, so $\delta = 5$ makes the borderline interval $[\tau - \delta, \tau + \delta] = [65, 75]$ exactly one grade band centred on the threshold: a case is

referred for human review whenever it could plausibly have received the adjacent letter grade. Second, this interval coincides exactly with the support overlap of the Fair and Good membership functions in Table 3, that is, with the region in which the fuzzy layer itself reports two simultaneously active categories. The band is therefore centred on the Fair–Good transition and includes its endpoints, at which only one category is active. Referral is determined by $|\hat{S} - \tau| \leq \delta$ rather than by the presence of two active categories, which also occurs in the 55–65 and 75–85 regions without triggering a referral. Because δ nevertheless determines how much work is delegated to human reviewers, its effect is examined empirically in Section 3.3.

2.7 Comparative Baselines

The proposed framework is compared against three baselines applied to the same records: conventional weighted thresholding, that is Equation (1) followed by the crisp rule $S \geq \tau$; a fuzzy-only configuration, in which the defuzzified soft score $\hat{S}$ replaces S in the crisp rule and neither the neural component nor the referral rule is used; and a neural-only configuration, in which the perceptron of Section 2.5 produces the attainment decision directly. The comparison is reported in Section 3.5. It is a structural comparison on five records and not a statistical benchmark; its purpose is to identify what each component contributes and what it cannot contribute, not to rank the approaches by accuracy.

2.8 Illustrative Evaluation Procedure

Four anonymized CPMK records from the Computer Science program were used to trace the calculation procedure. For each record, Equation (1) produced the weighted CPL score and reference status. The fuzzy rules then supplied membership degrees and a soft score for the referral check. The perceptron was trained on these records for one epoch, and its outputs were compared with the reference labels. Student E provided an additional scenario with one unavailable score. For that case, we calculated the partial weighted score and the neural estimate, then repeated the calculations after supplying an assumed value for the missing component. The comparison examines the outputs and referral decisions of the configurations. No

held-out predictive evaluation or external validation was performed.

3. RESULTS AND DISCUSSION

3.1 Worked Example: Weighted CPL Calculation

Table 6 uses four student records from Software Engineering (Rekayasa Perangkat Lunak). Three CPMK contribute to CPL-11 with weights of 0.40, 0.30, and 0.30. The table reports their scores, the aggregate from Equation (1), and the status at $\tau = 70$.

Table 6. CPMK scores and computed CPL-11 attainment for four students (weights 0.40 / 0.30 / 0.30)

Student	CPMK-1 (0.40)	CPMK-2 (0.30)	CPMK-3 (0.30)	CPL-11 score	Status ($\tau = 70$)
Student A	80	85	75	80.0	Attained
Student B	65	80	70	71.0	Attained
Student C	65	70	65	66.5	Not attained
Student D	90	80	85	85.5	Attained

For example, Student A obtains $(80 \cdot 0.40 + 85 \cdot 0.30 + 75 \cdot 0.30) / 1.00 = 80.0$, and Student C obtains $(65 \cdot 0.40 + 70 \cdot 0.30 + 65 \cdot 0.30) / 1.00 = 66.5$. Students A, B, and D meet the threshold for CPL-11; Student C does not. The binary labels give no indication of distance from the threshold. B is 1.0 point above it, and D is 15.5 points above it, yet both are labeled attained. C is 3.5 points below it. These distances remain available in the weighted scores. The fuzzy layer adds membership in adjacent categories for use during review.

3.2 Interpretation of the Hybrid Output

The illustrative results show that the proposed framework produces more informative outputs than a conventional binary threshold. In the conventional procedure, students are classified only as attained or not attained. In the fuzzy component, each score is accompanied by membership degrees that describe its position relative to adjacent categories, as summarized in Table 7.

Table 7. Fuzzy evaluation of the four students in the worked example

Student	CPL	$\mu(\text{Fair})$	$\mu(\text{Good})$	$\mu(\text{Excellent})$	Soft score	Status
Student A	80.0	0.00	0.50	0.50	82.5	Attained
Student B	71.0	0.40	0.60	0.00	71.0	Attained (borderline)
Student C	66.5	0.85	0.15	0.00	66.5	Not attained (borderline)
Student D	85.5	0.00	0.00	1.00	90.0	Attained

Students B and C fall within the review band, although their reference statuses differ. For B, $S = 71.0$ gives $\mu(\text{Fair}) = (75 - 71)/10 = 0.40$ and $\mu(\text{Good}) = (71 - 65)/10 = 0.60$. Rules R2 and R3 therefore produce $\hat{S} = 0.40 \times 65 + 0.60 \times 75 = 71.0$. For C, $S = 66.5$ gives memberships of 0.85 and 0.15, producing $\hat{S} = 0.85 \times 65 + 0.15 \times 75 = 66.5$. Their distances from 70 are 1.0 and 3.5 points, so

Table 8. Sensitivity of the human-review referral set to the borderline half-width δ ($\tau = 70$)

δ	Borderline band $[\tau - \delta, \tau + \delta]$	Referred for human review	Not referred by the borderline rule	Referral rate
3	[67, 73]	B, E	A, C, D	2/5 (40%)
5	[65, 75]	B, C, E	A, D	3/5 (60%)
7	[63, 77]	B, C, E	A, D	3/5 (60%)
10	[60, 80]	B, C, E	A, D	3/5 (60%)

At $\delta = 3$, the rule refers B and E. Increasing δ to 5 adds C; increasing it to 7 or 10 adds no further case. The ordered distances from the threshold are 1.0 (B), 2.6 (E), 3.5 (C), 12.5 (A), and 20.0 (D). Thus, the referral set remains B, C, and E for $3.5 \leq \delta < 12.5$. C enters at the lower endpoint, while A enters at 12.5. The interval is determined by the third- and fourth-smallest distances in this example. Its stability describes these five cases only. A program-level choice of δ would require the cohort distribution of distances and the number of reviews that can be handled.

both meet the referral condition at $\delta = 5$. A and D lie outside the band.

Only R2 and R3 fire for B and C. None of the four records has positive membership in Not Attained. Equation (2) returns $\hat{S} = S$ within the Fair–Good interval because the output singletons are located at the corresponding membership peaks of 65 and 75. Outside this interval, the soft score can differ from S : A moves from 80.0 to 82.5, and D from 85.5 to 90.0. These changes do not alter any of the four attainment labels. Figure 1(b) plots the mapping. The membership values and distance $|\hat{S} - 70|$ accompany the reference status for review.

3.3 Sensitivity of the Borderline Band to δ

We applied referral half-widths of 3, 5, 7, and 10 to the four complete records and to the assumed complete profile for Student E. For E, both the weighted score and soft score are 72.6. Table 8 lists the resulting referral sets.

3.4 Complementary Neural Estimate from a Partial Profile

The neural component serves as a complementary estimator when one or more CPMK scores are unavailable. To illustrate the mechanism, a single-perceptron prototype was initialized with small weights and a bias and updated over one pass through the worked set, adjusting its parameters whenever its prediction disagreed with the reference label. The point of the illustration is the update mechanism itself, not a performance figure; with a single pass over four records the prototype is not expected to converge. Table 9 traces one pass through the training records. Parameter updates occur for Students A, B, and C; Student D produces no update. The parameters after the pass are $w = (0.180, 0.195, 0.180)$ and $b = -0.40$.

Table 9. Trace of the single training epoch of the perceptron prototype

Step	Record	$z = w \cdot \tilde{x} + b$	y	d	d - y	w after step	b after step
init	n/a	n/a	n/a	n/a	n/a	(0.100, 0.100, 0.100)	-0.50
1	A	-0.2600	0	1	+1	(0.180, 0.185, 0.175)	-0.40
2	B	-0.0125	0	1	+1	(0.245, 0.265, 0.245)	-0.30
3	C	+0.2040	1	0	-1	(0.180, 0.195, 0.180)	-0.40
4	D	+0.0710	1	1	0	(0.180, 0.195, 0.180)	-0.40

Re-applying the resulting parameters to the training records themselves gives $z = +0.0448$ for Student A, -0.0010 for Student B, -0.0295 for Student C, and $+0.0710$ for Student D, so three of the four records are reproduced and Student B is misclassified. The single residual training error is the borderline case. That is the expected behaviour of an unconverged linear classifier, and it is also the reason the fusion rule of Section 2.6 does not permit the neural estimate to decide borderline cases.

Student E has recorded scores of 78 for CPMK-1 and 68 for CPMK-3; CPMK-2 is unavailable. Equation (1) gives a partial weighted score of $(78 \times 0.40 + 68 \times 0.30)/(0.40 + 0.30) = 73.7$, rounded to one decimal place. The soft score is also 73.7. Its distance of 3.7 points from 70 places the case inside the review band. Using the normalized vector $\tilde{x} = (0.78, 0, 0.68)$, the perceptron gives $z = -0.1372$ and predicts not attained.

To complete the scenario, we assume a score of 70 for CPMK-2. The weighted score becomes 72.6, which still meets the attainment threshold and the referral condition. The perceptron gives $z = -0.0007$ for $\tilde{x} = (0.78, 0.70, 0.68)$, so its prediction remains not attained. Supplying the missing value does not resolve this error. In this example, the partial weighted calculation agrees with the assumed complete-profile status, while the neural estimate does not. The framework retains the weighted status and refers the case at both stages. The example shows the handling of disagreement between outputs; it provides no evidence about the quality of the reviewer's eventual decision.

3.5 Comparison with Baseline Configurations

Table 10 applies the three baseline configurations defined in Section 2.7 and the proposed framework to the same five records, so that the contribution of each component can be located rather than assumed.

Table 10. Structural comparison of the proposed framework with three baseline configurations (five records)

Criterion	Conventional weighted threshold	Fuzzy-only	Neural-only (prototype)	Proposed framework
Decision authority	Crisp rule on S	Crisp rule on $\tilde{S}$	Perceptron output	Official weighted rule on S, retained
Agreement with official status	5/5 by construction	5/5	3/5 (B and E wrong)	5/5
Information near the threshold	None	Membership vector and soft score	None	Membership vector, soft score, distance to τ
Operates on an incomplete CPMK profile	Yes, over the assessed CPMK under Equation (1); not a prediction of the complete-profile status	Yes, from the partial aggregate	Yes, without using the contribution weights; incorrect in the example	Yes; partial weighted score, referral flag, and neural estimate reported together
Traceability of a single decision	Full	Full (at most two fired rules)	Low; weights are not interpretable	Full
Explicit human-review trigger	None	None	None	Yes, $\tau \pm \delta$
Parameters adapted from data	None	None	Weights and bias	Neural component only; fuzzy MFs remain policy-fixed

The table locates what each component contributes. Relative to conventional thresholding, the fuzzy layer adds no decision and changes no

outcome; it adds the distance of a case from the boundary and the two categories in which it simultaneously lies, which the crisp rule discards.

Relative to fuzzy inference, the neural component adds only the ability to produce an estimate without the contribution weights. The incomplete-profile row requires care: under the convention of Section 2.3 the weighted and fuzzy configurations also operate on a partial profile, and in the Student E scenario they agree with the eventual status while the perceptron does not. On the evidence available the neural component contributes no measurable benefit, and its place in the architecture rests on the narrower case in which the contribution weights are not yet defined.

Relative to an integrated ANFIS design such as [18], the proposed architecture gives up parameter learning inside the fuzzy layer. That is a deliberate trade-off. In ANFIS, the membership functions are fitted to the data and therefore become a function of the cohort, whereas here they encode the program's grading bands and remain fixed, auditable, and stable across cohorts, which is what allows the fuzzy output to be shown to a student as an account of that student's own assessment. ANFIS also requires a training set large enough to identify its premise and consequent parameters, a condition the single-course data used here does not meet. Keeping the two components separate has a further consequence an integrated model cannot offer: because the neural estimate never enters the defuzzification, an incorrect estimate cannot propagate into the interpretable output, which is precisely what happens with Students B and E.

The behavior of the framework when the components disagree is fully determined and is not left to a weighting heuristic. If the neural estimate contradicts the fuzzy result or the official weighted status, the official status stands; the disagreement is recorded and, when the case lies inside the referral band, is shown to the reviewer as an additional signal. The framework never resolves such a conflict autonomously.

Whether the human-in-the-loop stage improves accountability, consistency, or both cannot be settled by this demonstration. What it demonstrably provides is procedural: a case satisfying $|\hat{S} - \tau| \leq \delta$ cannot be closed by the system, so the record of a borderline decision always carries the identity of the person who made it, which is the requirement that the human-in-the-loop assessment literature [21], [22] and the

explainable-AI literature in education [23], [24] place on high-stakes automated decisions. Whether reviewers then decide consistently with one another is an empirical question about inter-rater agreement that this study does not address.

3.6 Discussion

The main contribution of the proposed framework lies in the transparency of the evaluation process rather than in predictive superiority. The fuzzy component provides additional information that is not available in a conventional threshold-based procedure. Instead of treating two nearby scores as completely different categories, the framework represents the degree to which each score belongs to adjacent attainment categories, which is particularly useful for students near the threshold.

The neural network component has a supporting role. It is particularly relevant when the complete CPMK profile or contribution-weight information is not yet available. In this condition, the estimate may be used as an early academic indicator. Nevertheless, the neural output should be verified once the complete assessment data become available, and it does not carry authority over the official weighted result.

The framework is therefore intended as a decision-support instrument rather than an autonomous assessment system. Final academic decisions remain under the responsibility of lecturers and study-program managers. This positioning also responds to the accountability concerns commonly raised about AI-based assessment, since no decision is delegated entirely to an unexplained model [29]. Compared with recent outcome-based assessment research, the framework occupies a deliberately narrow position. Information systems for outcome attainment [3], [10], [11] make the deterministic aggregation visible and auditable but leave the threshold decision itself unqualified. Fuzzy assessment work [13], [19], [20] qualifies the decision but typically replaces the official rule with the fuzzy one, which would require the institution to amend its assessment regulation. Neuro-fuzzy and neural approaches [9], [16], [17], [18] focus on prediction and are evaluated using predictive metrics, which presuppose a dataset that the present single-program setting does not yet provide. The

reported results concern the information available for review and the rule used to refer a case. They do not establish a more accurate measurement of attainment. The official calculation and its underlying scores remain unchanged. Shabbir and colleagues [25] associate attainment figures with institutional conditions such as teaching workload; the additional fuzzy outputs do not remove these influences.

3.7 Limitations

This study has several limitations. First, the framework was demonstrated within a single Computer Science program, and its membership functions reflect the local attainment policy. Second, the illustrative dataset was not intended to provide evidence of broad predictive generalization. Third, the neural component was evaluated at the prototype level and requires further testing using larger datasets and multiple academic cohorts. Finally, the reliability of the framework remains dependent on the quality of the original CPMK scores assigned by lecturers, since the framework improves the aggregation and decision stage but cannot correct bias introduced at the point of grading. The following limitations also apply. First, the membership functions of Table 3 were specified manually from the program's grading bands and were neither elicited from a panel of lecturers nor fitted to data, so their validity rests on the plausibility of those bands. Second, the borderline half-width $\delta = 5$ is justified institutionally in Section 2.6 and shown to be stable over a plateau in Section 3.3, but it has not been validated against the distribution of $|\hat{S} - \tau|$ in a real cohort. Third, no inter-rater or lecturer-consistency analysis was conducted, so it is not known whether the human-review stage yields consistent decisions; the framework guarantees that a human decides, not that humans agree. Fourth, no baseline accuracy comparison and no standard predictive metrics are reported for the neural component, because a four-record training set with no held-out data would make such metrics uninterpretable rather than merely weak. Fifth, no external validation across programs or institutions was performed, and the threshold, the grading bands, and the weight-normalization convention embedded in the design are specific to the program studied. Sixth, the incomplete-profile scenario does

not exercise the condition for which the neural component was designed, because the contribution weights were available throughout; on the evidence presented, the component adds no measurable benefit over Equation (1) applied to the components already assessed.

CONCLUSION

The worked examples illustrate a CPL assessment procedure in which the weighted score supplies the reference status and fuzzy memberships describe its position among attainment categories. The soft score determines whether the case enters the human-review band. The perceptron's estimate is reported separately and has no authority to change the weighted status. The study establishes that the framework reproduces the official weighted CPL decision exactly in every case examined, that it reports membership degrees and a soft score locating a case relative to the attainment threshold, that it identifies borderline cases by an explicit rule whose outcome is stable over a range of borderline widths, that it produces a preliminary estimate from an incomplete CPMK profile, and that it routes borderline cases to a human reviewer instead of closing them automatically. It does not establish improved prediction accuracy, improved assessment reliability, reduced lecturer workload, improved student outcomes, superiority over conventional, fuzzy-only, or ANFIS-type alternatives, or generalizability beyond the program studied. In the incomplete-profile scenario examined, the neural estimate disagreed with the eventual official status while Equation (1), applied to the components already assessed, did not. This is reported as a limitation of the prototype rather than as evidence for the component.

The framework should be viewed as a prototype decision-support approach rather than as a fully validated predictive system. The next evaluation needs CPMK records from several courses and cohorts. Their score distribution would support the choice of δ and an estimate of the review workload. A replacement for the neural prototype should be trained and evaluated on separate datasets, with predictive metrics reported. Conventional thresholding, fuzzy-only inference, and ANFIS should be evaluated on the same

records. Reviewer agreement must also be measured, since the current examples do not show whether lecturers reach consistent decisions. Evaluation in other study programs and institutions would then test how well the procedure transfers.

DECLARATIONS

Author Contributions: The first author: conceptualization, methodology, writing original draft, and supervision. The second author: investigation, data curation, and writing the original draft. The third author: formal analysis, literature review, and writing - review and editing.

Funding: The Ministry of Religious Affairs of the Republic of Indonesia funded this research through Bantuan Operasional Perguruan Tinggi Negeri (BOPTN), Fiscal Year 2025.

Conflict of Interest: The authors declare no conflict of interest.

Generative Artificial Intelligence Use Statement: ChatGPT (OpenAI) and Claude (Anthropic) were used for language editing and checking for linguistic errors during manuscript preparation. Their use was limited to these editorial tasks. The authors remain fully responsible for the originality, accuracy, integrity, citations, analysis, interpretation, and conclusions of the manuscript.

Ethics Approval and Informed Consent: Permission to use the student assessment records was obtained from the Computer Science Study Program at Universitas Islam Negeri Sumatera Utara, Indonesia. Student-identifying information was anonymized in the records used to demonstrate the framework.

Data Availability: The student assessment records supporting this study are held by the Computer Science Study Program at Universitas Islam Negeri Sumatera Utara, Indonesia. The data are not publicly available, as access is restricted by the study program.

ACKNOWLEDGEMENT

The Ministry of Religious Affairs of the Republic of Indonesia funded this research through Bantuan Operasional Perguruan Tinggi Negeri (BOPTN), Fiscal Year 2025. The leadership of Universitas Islam Negeri Sumatera Utara, the Directorate of Research and Community Service, and the Computer Science study program provided support and facilities for the work.

REFERENCES

- [1] M. Z. Zamir, M. I. Abid, M. R. Fazal, M. A. A. R. Qazi, and M. Kamran, "Switching to outcome-based education (OBE) system: A paradigm shift in engineering education," *IEEE Transactions on Education*, vol. 65, no. 4, 2022.
- [2] R. Mufanti, D. Carter, and N. England, "Outcomes-based education in Indonesian higher education: Reporting on the understanding, challenges and support available to teachers," *Social Sciences & Humanities Open*, vol. 9, 2024.
- [3] B. Kushari and L. Septiadi, "A learning outcome assessment information system to facilitate outcome-based education (OBE) implementation," *Jurnal Pendidikan Teknologi dan Kejuruan*, vol. 28, no. 2, pp. 238–250, 2022, doi: 10.21831/jptk.v28i2.42339.
- [4] A. Aziz, M. A. U. Golap, and M. M. A. Hashem, "Student's academic performance evaluation method using fuzzy logic system," in *Proc. 2019 1st Int. Conf. on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 2019, pp. 1–6, doi: 10.1109/ICASERT.2019.8934464.
- [5] A. Krouska, C. Troussas, and C. Sgouropoulou, "Fuzzy logic for refining the evaluation of learners' performance in online engineering education," *European Journal of Engineering and Technology Research*, vol. 4, no. 6, pp. 50–56, 2019.
- [6] N. U. Jan, S. Naqvi, and Q. Ali, "Using fuzzy logic for monitoring students' academic performance in higher education," *Engineering Proceedings*, vol. 46, no. 1, art. 21, 2023, doi: 10.3390/engproc2023046021.

- [7] E. T. Lau, L. Sun, and Q. Yang, "Modelling, prediction and classification of student academic performance using artificial neural networks," *SN Applied Sciences*, vol. 1, no. 9, art. 982, 2019, doi: 10.1007/s42452-019-0884-7.
- [8] Y. Baashar et al., "Toward predicting student's academic performance using artificial neural networks (ANNs)," *Applied Sciences*, vol. 12, no. 3, art. 1289, 2022, doi: 10.3390/app12031289.
- [9] M. Nachouki and M. Abou Naaj, "Analysis of the factors affecting student performance using a neuro-fuzzy approach," *Education Sciences*, vol. 13, no. 3, art. 313, 2023, doi: 10.3390/educsci13030313.
- [10] W. D. Yuniarti, S. Hartati, S. Priyatna, and H. D. Surjono, "Learner assessment system in e-learning with OBE approach: Activity performance, ability level and recommendation," *HighTech and Innovation Journal*, vol. 5, no. 3, pp. 572–602, 2024.
- [11] I. G. S. M. Diyasa, I. N. D. P. Putra, N. K. I. M. Mandenni, M. R. M. Ariefwan, and R. L. Aryananda, "Implementation of innovation in integrated evaluation and learning system using outcome-based education ecosystem," *International Journal of Entrepreneurship and Business Development*, vol. 6, 2023.
- [12] T. Sasipraba et al., "Assessment tools and rubrics for evaluating the capstone projects in outcome based education," *Procedia Computer Science*, vol. 172, pp. 296–301, 2020, doi: 10.1016/j.procs.2020.05.047.
- [13] A. Aziz and M. M. A. Hashem, "Fuzzy logic-based assessment of students learning outcome in implementing outcome-based education," in *Proc. Int. Conf. on Big Data, IoT, and Machine Learning*, 2022.
- [14] A. Ozdemir and K. F. Balbal, "Fuzzy logic based performance analysis of educational mobile game for engineering students," *Computer Applications in Engineering Education*, vol. 28, no. 6, pp. 1536–1548, 2020.
- [15] T. A. Ajoin, S. S. Gran, A. Kanyan, and S. F. Lajim, "An enhanced systematic student performance evaluation based on fuzzy logic approach for selection of best student award," *Asian Journal of University Education*, vol. 16, no. 4, pp. 10–20, 2021.
- [16] M. Yağcı, "Educational data mining: Prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, art. 11, 2022, doi: 10.1186/s40561-022-00192-z.
- [17] N. Ahmad, N. Hassan, H. Jaafar, and N. I. M. Enzai, "Students' performance prediction using artificial neural network," *IOP Conference Series: Materials Science and Engineering*, vol. 1176, art. 012020, 2021, doi: 10.1088/1757-899X/1176/1/012020.
- [18] P. J. Loresco and N. Pilueta, "An adaptive neuro-fuzzy framework for monitoring student outcomes with individualized dashboard in outcome-based education," in *Proc. TENCON 2024 – 2024 IEEE Region 10 Conference (TENCON)*, 2024, pp. 1286–1289, doi: 10.1109/TENCON61640.2024.10903009.
- [19] H. Naccache and R. Hleiss, "Reducing classification risk in developmental performance assessment using fuzzy logic: evidence from preservice mathematics teacher education," *Frontiers in Education*, vol. 11, art. 1840093, 2026, doi: 10.3389/feduc.2026.1840093.
- [20] T. A. Tuan, N. D. Tho, N. H. Nghia, and D. T. T. Loan, "A systematic review and meta-analysis of fuzzy logic for students' performance assessment," *International Journal of Engineering Trends and Technology*, vol. 73, no. 9, pp. 155–174, 2025, doi: 10.14445/22315381/IJETT-V73I9P114.
- [21] W. Khan et al., "Auto-assessment of assessment: a human-in-the-loop AI framework addressing policy gaps in academic assessment," *PLOS ONE*, vol. 21, no. 4, art. e0346815, 2026, doi: 10.1371/journal.pone.0346815.
- [22] D. C. Fajardo-Ramos, A. Chiappe, and J. Mella-Norambuena, "Human-in-the-loop assessment with AI: implications for teacher education in Ibero-American universities," *Frontiers in Education*, vol. 10, art. 1710992, 2025, doi: 10.3389/feduc.2025.1710992.

- [23] G. Türkmen, “The review of studies on explainable artificial intelligence in educational research,” *Journal of Educational Computing Research*, vol. 63, no. 2, pp. 277–310, 2025, doi: 10.1177/07356331241310915.
- [24] S. Gunasekara and M. Saarela, “Explainable AI in education: techniques and qualitative assessment,” *Applied Sciences*, vol. 15, no. 3, art. 1239, 2025, doi: 10.3390/app15031239.
- [25] A. Shabbir, T. Rehman, M. Salman, S. Rizvi, and N. Saaludin, “Unmasking the impact of teaching workload on learning outcome attainment in outcome-based engineering education: a multi-level analysis of CLO and PLO achievement,” *Education Sciences*, vol. 16, no. 9, art. 1473, 2026, doi: 10.3390/educsci16091473.
- [26] P. Lasunon, “Triangular fuzzy grading method and its software,” in *Proc. 2019 5th Int. Conf. on Engineering, Applied Sciences and Technology (ICEAST)*, 2019.
- [27] N. A. Namli and O. Senkal, “Using the fuzzy logic in assessing the programming performance of students,” *International Journal of Assessment Tools in Education*, pp. 701–712, 2018.
- [28] G. Zheng et al., “SAPPNet: Students’ academic performance prediction during COVID-19 using neural network,” *Scientific Reports*, vol. 14, art. 24605, 2024, doi: 10.1038/s41598-024-75242-2.
- [29] H. Khosravi et al., “Explainable artificial intelligence in education,” *Computers and Education: Artificial Intelligence*, vol. 3, art. 100074, 2022, doi: 10.1016/j.caeai.2022.100074.